

Contents

- 1. Capacity additions 3
- 2. Energy procurement strategies 8
- 3. Financing trends 13
- 4. Leasing drives confidence in neoclouds as depreciation remains a risk 20
- 5. AI lab fundamentals 23
- About us 28

## Data Center Market Indicators 1H 2026: Capex To Jump Again

The digital infrastructure boom has continued apace over the last six months despite jitters in equity markets and widespread “bubble” fears. Data center developers are ramping up spending plans and procuring more energy than ever while private investors are pouring money into AI companies. Novel debt structures are being utilized to deploy capital.

### 59GW

Live IT capacity tracked by BNEF as of 3Q 2025

### \$3.3 trillion

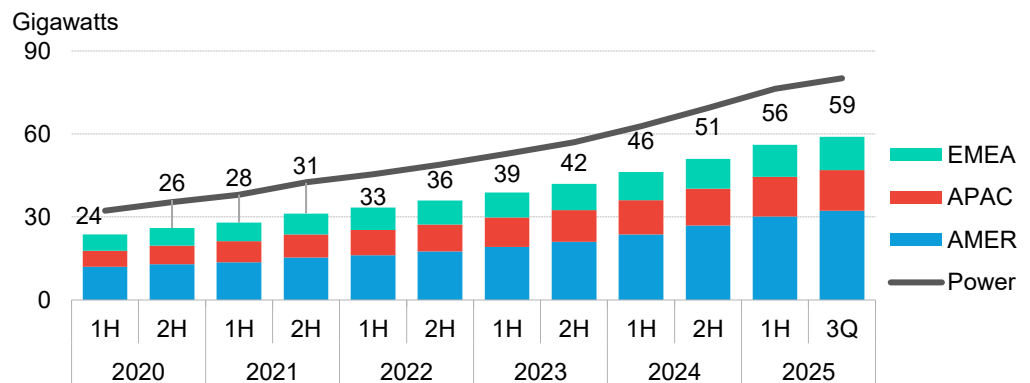
Expected data center capex through 2029

### \$160 billion

Raised by AI labs in first two months of 2026

- **IT capacity:** Connections of new live IT (server power draw) capacity continued to grow in 3Q 2025, reaching 2.8GW, driven by activity in the US. The least positive indicator tracked in this report is that 3Q 2025 construction starts were flat compared to a four-quarter average of 3.8GW. BNEF’s conversations with industry and capex projections from the hyperscalers suggest this figure is likely to be revised upwards as more projects are tracked over time.
- **Energy procurement:** The data center industry accounted for half the global corporate clean power purchase agreements (PPA) signed in 2025. Over 80% of this activity was in the US. Meanwhile, BNEF has now tracked 114GW of announced capacity for on-site gas generation at data centers as developers seek to accelerate speed to power. In the US, this gas fleet is equivalent to around half the entire data center IT capacity pipeline.
- **Financing:** The strongest indicator of continued positive momentum in the AI boom is the capex projections for FY2027 of the largest publicly listed data center operators, which have increased by more than 50% to \$806 billion since the last iteration of this report six months ago. These companies are the key source of capital driving infrastructure buildout.
- **AI labs:** Equity funding appetite for the leading AI startups remains sky high with 2026 financing totals already exceeding 2025. While business risk for scaling is high, both Anthropic and OpenAI are reporting gross margins of 30-40%, suggesting some business model based on today’s technology could ultimately turn a profit.

Figure 1: Global data center IT capacity and power load



Source: BloombergNEF, DC Byte. Note: 3Q 2025 data remains provisional. IT capacity refers to the power draw of IT equipment. Power load refers to total power draw, calculated using a Power Usage Effectiveness (PUE) of 1.36.

Table 1: Data center market indicators

Section	Subsection	Why this matters	Importance	What happened this period? Color indicates <b>positive</b> , neutral, or <b>negative</b> momentum. Darker shading indicates stronger change in direction.
<b>Capacity additions</b>	<b>Live IT capacity</b>	Additions indicate follow through on investment		<b>Additions in 3Q 2025 were up 10% on the rolling four-quarter average.</b> Global capacity is up 10.2GW year-on-year.
	<b>Capacity under construction</b>	New starts indicate capacity that is likely to come online in 1-3 years		<b>Construction started on 3.8GW of IT capacity in 3Q 2025, flat on the rolling four-quarter average.</b>
<b>Energy procurement strategies</b>	<b>Clean PPA volume</b>	Clean energy purchases are a signal for future energy demand		<b>Clean power procurement from data centers reached a new record in 2025 – at 28.3GW – and shifted towards higher capacity factor, and more expensive, technologies such as nuclear.</b> Activity was driven by US deals.
	<b>On-site gas generation</b>	The scale of on-site gas generation indicates how urgently developers are trying to energize data centers.		<b>The last iteration of this report tracked around 9GW of on-site data center gas capacity across five sites. BNEF has since identified an announced pipeline of 114GW of capacity across 115 projects.</b>
<b>Infrastructure finance</b>	<b>Data center developer capex</b>	Changes show shift in momentum and cash flows show potential for growth.		Data center operators have revised <b>2027 capex estimates up by 56%</b> since the last iteration of this report, despite some equity market jitters.
	<b>Project financing</b>	Scale and terms of new investments illustrate investor sentiment in the sector		BNEF has <b>not tracked any new special purpose vehicles</b> to finance large giga-scale campuses, <b>but revenue guarantees</b> by the likes of Google and Nvidia <b>did reportedly enable the financing</b> of some neocloud developments.
<b>Neoclouds</b>		The performance of these AI-specialized data center operators acts as a bellwether for AI demand		The market caps of the public <b>neoclouds</b> have been volatile throughout the year but the group of firms, both public and private. <b>BNEF tracked nine leasing contracts with hyperscalers</b> , worth up to more than <b>\$100 billion</b> highlighting demand for compute.
<b>AI lab fundamentals</b>	<b>Access to capital and cash burn rates</b>	AI products are driving a significant amount of new data center demand. The ability of these labs to access capital is a driver of demand for AI services		Appetite for the equity financing of frontier AI labs has remained high as do cash burn projections. <b>Equity funding in 2026 has already exceeded that of 2025.</b> While this would be true solely thanks to OpenAI's recent \$110 billion round, Anthropic also raised \$30 billion and xAI raised \$20 billion.
	<b>Monetization strategies and user adoption</b>	Demand for AI compute could drop if AI labs are unable to make money from AI services		Recent reporting indicates both <b>OpenAI and Anthropic have gross margins of 30-40% on inference services.</b> This suggests while there is still significant risk associated with scaling generative AI businesses, there is a <b>reasonable pathway to selling services profitably based on existing technology.</b>

Source: BloombergNEF. Note: Darker colors indicate the metric is more important.

### 1. Capacity additions

BNEF tracked 2.8GW of new IT capacity connected in 3Q of 2025, taking the global data center fleet to 59GW of IT capacity. This is a 10GW increase compared to 3Q 2024, with the sector growing 21% year-on-year. Nearly a quarter of the annual additions were connected across just 10 projects, primarily serving AI workloads. Nine of these projects were in the US, with one in Canada. BNEF expects the concentration of capacity into a handful of projects to continue, as more AI-focused megaprojects break ground.

BNEF has also revised historical numbers upwards since our last publication of this data, as more details about tracked projects has been made available. This included a 1.3GW upward revision to 1H 2025 additions, with 5.1GW now identified as having come online. This makes it the highest-growth half year BNEF has tracked, beating 2H 2024 by 276MW. Most changes were from projects identified as under construction which have now been revised as energized.

#### 1.1. Regional trends

The Americas region continued to lead on quarter-on-quarter capacity additions in 3Q 2025. It has connected the most capacity in all but eight periods since 1991. BNEF tracked 2GW of new IT capacity connected in the region over the three-month period, 72% of the 2.8GW seen globally. This is likely to be revised in the next data update – due to reporting delays in Asia-Pacific and parts of Europe, BNEF expects the true volume of additions in these regions to be slightly higher. Additions for 1H 2025 have been updated, with a total of 5.1GW of new connections tracked. This makes it the largest half on record for new connections. In absolute terms, the biggest shift was in Americas, with an additional 477MW of live capacity identified. In relative terms, the largest movement was in Asia-Pacific, with an additional 490MW of connections identified, taking the regional total to 990MW. New connections in 2H are generally greater than those in 1H, due to the dominance of the US in global numbers and the seasonality of construction in the Northern hemisphere. As a result, BNEF expects the 2H numbers for 2025 to break this record.

Figure 2: Data center live IT capacity, change on previous half

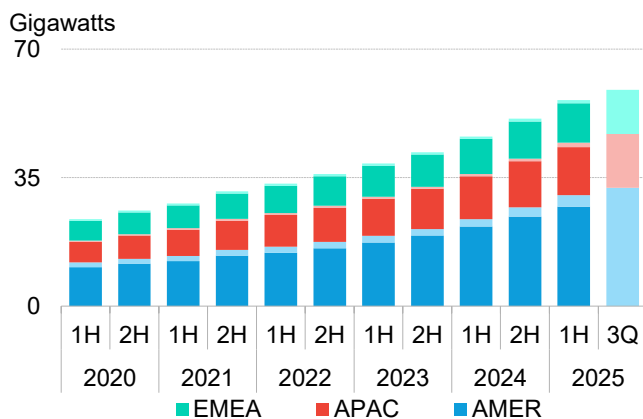
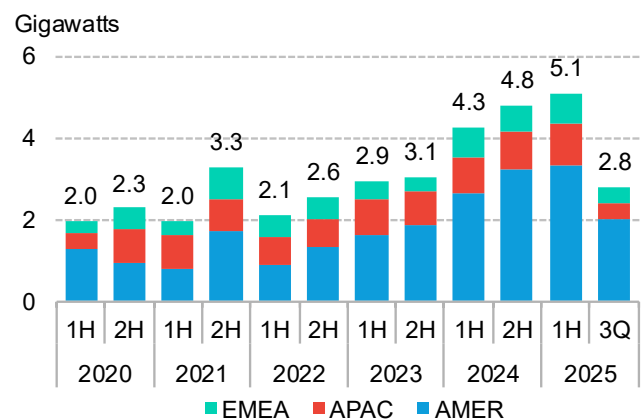


Figure 3: Data center new IT capacity, change on previous half



Source: BloombergNEF, DC Byte. Note: 3Q 2025 data remains provisional. Hashed boxes represent change in data since September 2025. EMEA refers to Europe, Middle East and Africa. APAC refers to Asia Pacific. AMER refers to the Americas.

### 1.2. Construction starts remain strong

In 3Q of 2025, over 3.8GW of new data center IT capacity entered its construction phase, which is up 58% on the quarterly average so far this decade. However, it is a marginal drop off compared to the previous quarter, slipping 16%. This could potentially indicate a slowdown in activity but BNEF believes that this is due to a delay in reporting of new starts. There has been no indication of a slowdown based on conversations BNEF has had with industry stakeholders. Any revisions to this figure will, however, be a key indicator to watch in the coming quarter.

Improved visibility on projects begun in 1H 2025 reveals that 8.5GW of new IT capacity broke ground in the six-month period, revised upward from 6.1GW reported in the previous iteration of this note.

The US hosts 15.9GW of current data center construction.

As has been the case since 2022, the Americas dominated new construction, with 78% of new IT capacity globally. This is highly concentrated in the US, with project starts being driven by demand for AI training and inference infrastructure within the market. Five sites broke ground on at least 100MW of new capacity, all of which are in the US (Table 3).

A total of 23.1GW of capacity is tracked as under construction globally, split across 831 sites. The Americas region contains 17GW of this capacity, across 311 locations. EMEA and APAC trail significantly. APAC has marginally more ongoing construction, with 3.2GW across 283 sites versus EMEA's 2.9GW across 258. The US alone hosts 15.9GW of ongoing construction.

Figure 4: Data center IT capacity, by start of construction date

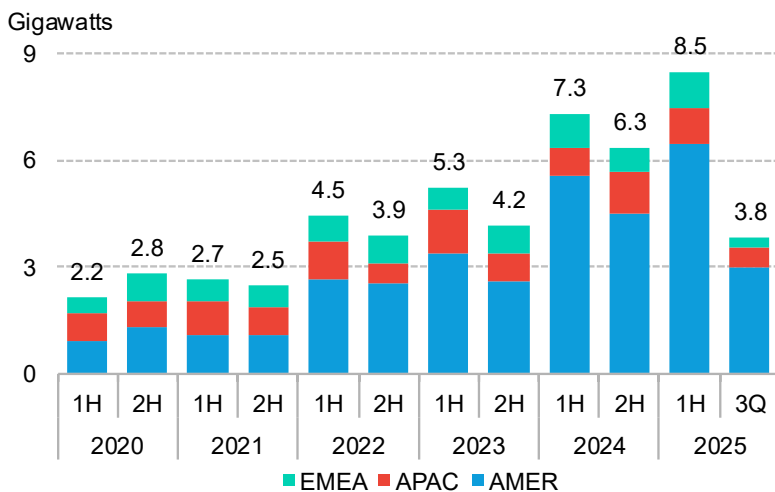
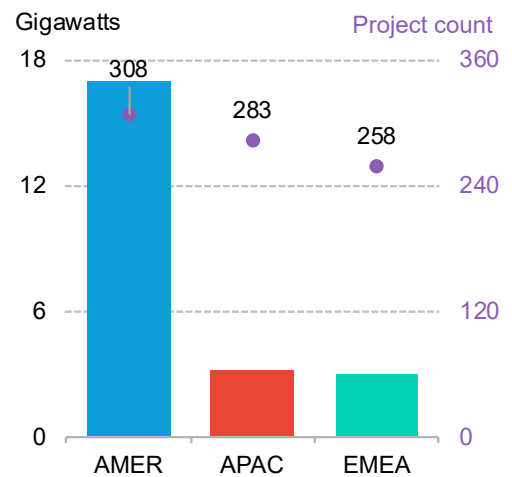


Figure 5: Total data center capacity under construction, and number of projects



Source: BloombergNEF, DC Byte. Note: EMEA refers to Europe, Middle East and Africa. APAC refers to Asia Pacific. AMER refers to the Americas. 3Q 2025 data remains provisional.

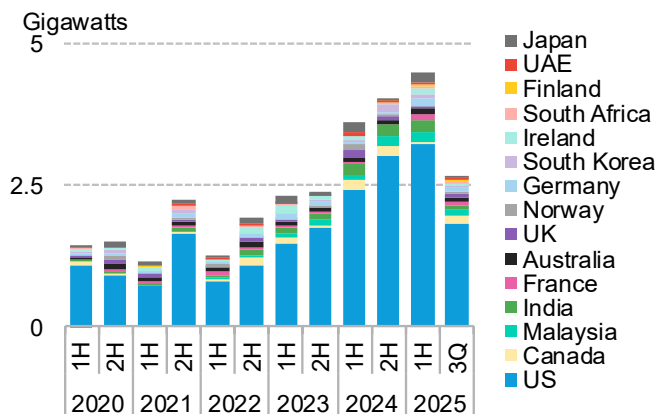
### 1.3. US accounts for 74% of additions in 3Q 2025

The US hosts 1.8GW of the 2.8GW IT capacity connected in 3Q 2025 (Figure 6), increasing its overall share of global IT capacity above 50%. This was largely driven by the connection of a few mega-projects, with over half of new US capacity contained in the largest five projects, all of which are serving AI workloads (Table 2). The US is likely to continue to add the most capacity in the coming years, given that it contains 67% of data center capacity currently under construction.

Outside the US, Canada added the most new capacity (Figure 7), although this was skewed by a 90MW connection at an AWS facility in Calgary. This made up two thirds of total new capacity in the market. Malaysia climbed to third, underscoring its importance as a growth market. 91MW of new capacity was connected, across six sites. BNEF tracked two sites breaking ground on new capacity, including a Google project in the Sengalor region, with 160MW planned. This is slightly north of most Malaysian data center capacity, which is concentrated in Johor due to its proximity to Singapore.

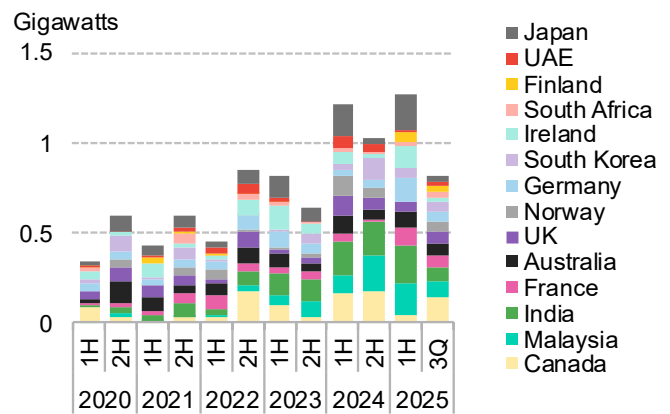
France came in at fifth for new capacity additions, and recently scored well in BNEF’s regional data center development ranking. It has structural advantages for AI facilities within Europe, which is likely to serve as a tailwind for growth. For more, see *Data Center Development: Regional Drivers and Hurdles* ([web](#) | [terminal](#)).

**Figure 6: Data center live IT capacity additions for 15 largest markets**



Source: DC Byte, BloombergNEF. Note: 3Q 2025 data remains provisional.

**Figure 7: Data center live IT capacity additions for 14 largest markets, excluding US**



Source: BloombergNEF, DC Byte. Note: 3Q 2025 data remains provisional.

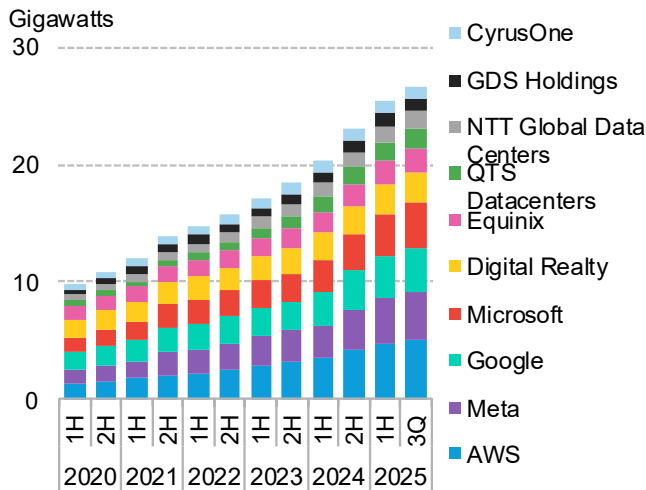
### 1.4. AWS adds most capacity in 3Q 2025

AWS connected the most new capacity in 3Q 2025 – at 377MW – having also led in 1H. The dominance of the US in data center deployments is evident in the list of top additions by company – all companies in the top 12 are American and operate primarily, or solely, in the US.

Google had the slowest quarter of the hyperscalers in terms of additions, connecting 52MW of new capacity across two sites. This was the first energization for both facilities – one in Arizona and one in London, UK. It also did not break ground on any hyperscale facilities or those that are more than 100MW in size, beginning 50MW of construction across two sites, including a 30MW expansion to a site in North Carolina dating back to 2007, and a greenfield project in Malaysia.

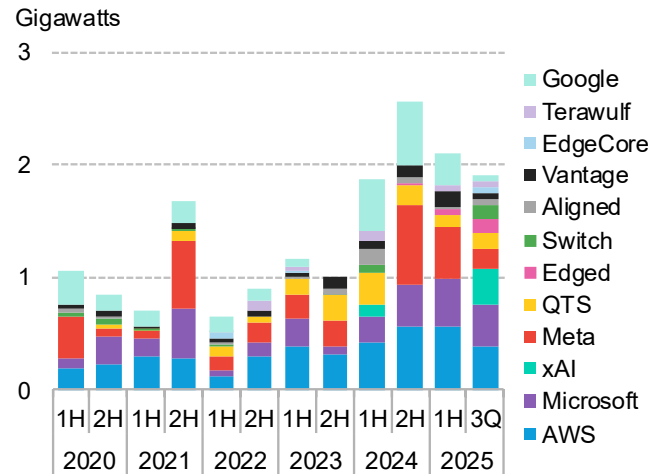
xAI had the most additions of a privately held company, with the Elon Musk-backed AI lab connecting 318MW of new capacity, across its Colossus 1 and Colossus 2 sites in Tennessee.

Figure 8: Live IT capacity of largest operators



Source: BloombergNEF, DC Byte. Note: 3Q 2025 data remains provisional. AWS refers to Amazon Web Services.

Figure 9: Largest IT capacity added by operators in 3Q 2025



Source: BloombergNEF, DC Byte. Note: 3Q 2025 data remains provisional. AWS refers to Amazon Web Services. QTS refers to QTS Datacenters. Vantage refers to Vantage Data Centers.

### 1.5. Project trends

The largest new IT capacity addition at a single site was the “Razor5” project in Indiana of AWS, which added 232MW. Ultimately planned to be nearly a gigawatt, a further 532MW is under construction. It has been developed rapidly – first publicly announced in April 2024, it secured permits in August and broke ground in September of the same year. It is part of a wider 2.4GW development plan in Indiana, supported by 2.6GW of natural gas and a 400MW, four-hour battery.

The second-largest new addition was xAI’s Colossus 2, which added 212MW. This was phase one of project development. Already the fastest project to hit 200MW that BNEF has tracked, we understand that it has since come fully online. In January 2026, it hit 1.4GW of total power draw (approximately 1.1GW of IT capacity). It is one of the prime examples of on-site gas generation being used as bridging generation while the site hopes to connect to the grid. A mix of gas turbines and Tesla megapack batteries are being deployed while a grid connection from Tennessee Valley Authority is upgraded. For more on this trend see *Section 2.2*.

Table 2: Data center IT capacity additions at selected sites, 3Q 2025

Operator	Facility	State, Country	New IT capacity in 3Q 2025 (MW)	Planned total IT capacity (MW)	Project purpose
AWS	Razor5	Indiana, US	232	928	AI / Cloud
xAI	Colossus 2	Tennessee, US	212	1112	AI
Meta	Eagle Mountain Data Center	Utah, US	113	450	AI
xAI	Colossus 1	Tennessee, US	106	212	AI
Edged	Atlanta	Georgia, US	100	180	AI

Operator	Facility	State, Country	New IT capacity in 3Q 2025 (MW)	Planned total IT capacity (MW)	Project purpose
AWS	South East Calgary	Alberta, Canada	90	90	Public cloud
Meta	Project Peregrine	Idaho, US	90	180	AI
QTS Datacenters	Phoenix 2	Arizona, US	84	280	Traditional colocation
Switch	The Keep Campus	Georgia, US	70	150	Traditional colocation
Aligned	PHX-04	Arizona, US	60	60	Traditional colocation

Source: BloombergNEF, DC Byte.

The largest additions in 3Q 2025 were dominated by AI facilities. This marks a shift from the facilities brought online in 1H 2025, discussed in the previous iteration of this note. Then, the largest additions were primarily traditional cloud and colocation projects. The increase in AI-driven projects appearing in 3Q is a result of development timelines. Assuming a construction timeline of 18-24 months, we are entering the phase when the first infrastructure decisions driven by AI development are coming online. Further, the huge amounts of capital flowing into the sector are driving expedited timelines, as has been seen with Colossus 1 and 2. Colossus 2 started with the acquisition of a former warehouse on March 7, 2025. It was operational at 1.4GW by January 2026.

Major construction starts were also dominated by AI facilities (Table 3). Terawulf's Lake Mariner is a prime example of this – the upstate New York campus is explicitly focused on AI workloads. Around half of its capacity is being leased to FluidStack, in a deal backstopped by Google (see section 3.3).

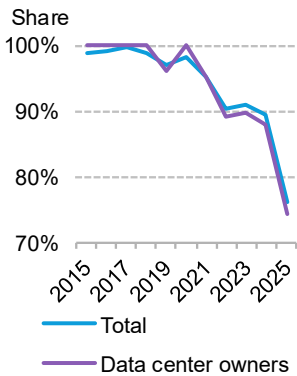
Table 3: Construction starts at selected US sites in 3Q 2025

Operator	Facility	State	New IT capacity under construction in 3Q 2025 (MW)	Planned total IT capacity (MW)	Project purpose
Terawulf	Lake Mariner	New York	459	752	Colocation, AI and HPC
Switch	The Keep 2.0 – Atlanta North	Georgia	300	2,000	No evidence
Meta	Richland Parish (Hyperion)	Louisiana	250	5,000	AI training
CyrusOne	DFW10	Texas	144	338	No evidence
T5 Data Centers	T5@Atlanta	Georgia	132	248	Traditional colocation

Source: BloombergNEF, DC Byte, company press releases. Note: Project purpose is a mixture of public reporting and BNEF analyst view. "No evidence" refers to sites where there is no conclusive evidence about what workloads it will serve. HPC refers to high performance computing. AI refers to artificial intelligence.

## 2. Energy procurement strategies

**Figure 10: PPA capacity that is exclusively wind or solar**



Source: BloombergNEF.

Energy procurement by data center owners and developers continued to accelerate in 2025. Clean power procurement reached new heights driven largely by growth from Meta and Alphabet. Meanwhile, the pipeline of on-site gas plants colocated with data centers has rapidly expanded.

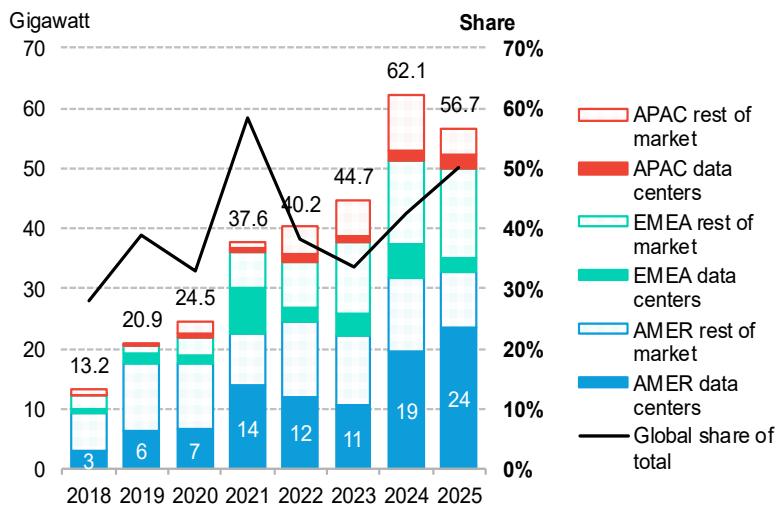
### 2.1. Clean power procurement

The world's largest data center operators<sup>1</sup> reached a peak in clean power procurement in 2025 with 28.3GW in agreements announced. This peak was almost exclusively driven by a rise of PPAs in the US, which accounted for 82% of all PPAs that were signed by data center operators. APAC also reached a new peak, almost doubling its previous best year. EMEA meanwhile saw the smallest volume of PPAs signed by data center operators since 2022 (Figure 11).

PPAs signed by the world's largest data center operators accounted for about half of all clean PPAs in 2025, which was the second highest share of any year since the market exceeded 5GW in annual purchases. The only year to exceed 2025 was 2021, when Microsoft and Amazon ramped up procurement to reach over 18GW in a single year. Amazon has maintained similar levels of procurement since, while Microsoft saw an immediate drop off in 2022.

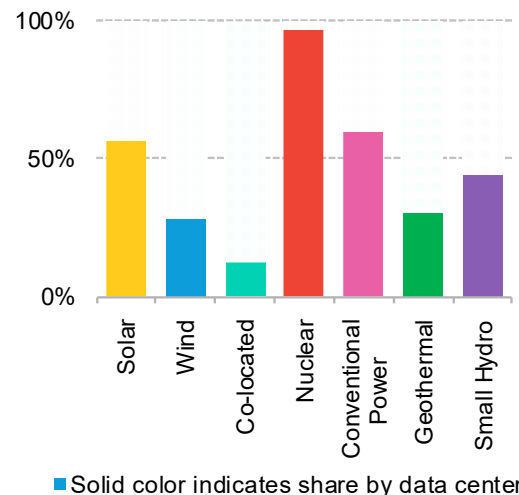
The data center industry is moving away from exclusively solar and wind PPAs exactly in lockstep with the wider market (Figure 10), where around 25% of capacity now combines renewable energy generation profiles like solar and wind, that are paired with storage, or some other source of clean power. This does not mean that data center operators do not have unique procurement profiles. The most distinctive characteristic is their propensity for nuclear. Some 97% of nuclear PPA capacity signed in 2025 came from data center operators (Figure 12).

**Figure 11: Clean PPA capacity signed by 20 largest data center operators vs. rest of the market**



Source: BloombergNEF

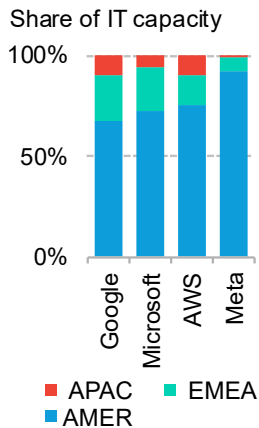
**Figure 12: Share of total 2025 PPA capacity signed by data center operators, by sector**



Source: BloombergNEF. Note: Conventional power refers largely to hydropower.

<sup>1</sup> This includes the 20 largest companies as measured by BNEF's Live IT Capacity Database ([web](#) | [terminal](#)). PPAs for these companies are tracked by applying [these filters](#) to BNEF's PPA Database. Throughout Section 1.1, this is the list of companies that is being referred to when speaking about data center owners.

**Figure 13: Distribution of data center live IT capacity**



Source: BloombergNEF, DCByte.

### Meta accelerates large clean power program in the US

Meta saw the biggest increase in clean energy purchasing in 2025 among the hyperscalers – more than doubling. It took the top position for the first time, narrowly edging out Amazon with 10.24GW in clean PPA capacity (Figure 14). Meta is the most US-focused of the hyperscalers with its power purchases, which is aligned with the company’s more US-centered data center fleet (Figure 13).

While 80% of this capacity was from solar PPAs, Meta was one of several hyperscalers that started to scale investment in nuclear in 2025. Its largest PPA was a 20-year contract to procure 1,121MW in power from Constellation’s Clinton nuclear plant in Illinois. In 2026, it has continued on this front, signing a 2.2GW PPA with Vistra that will easily bring its spend on nuclear into the billions. For more see *Meta’s Nuclear Buy Is Not Near Its Data Centers* ([web](#) | [terminal](#)).

### Amazon maintains consistent procurement and goes big on nuclear

In 2025, Amazon maintained a consistent level of clean power procurement, as it has for the last five years, with 10.22GW signed. While the volume of gigawatts signed was stable, Amazon’s clean energy spend likely increased substantially in 2025 as 35% of the power it procured was for nuclear energy, all of which came from three deals. Nuclear’s higher capacity factor and price compared to solar and wind mean its 2025 contracts will amount to significantly higher clean energy spend than previous years.

Amazon is the most regionally diverse clean energy buyer of the hyperscalers, with around a third of its purchases taking place outside of the US. Other large markets included Australia, India, Germany and Finland. Amazon’s geographical diversification has helped it be the biggest company by PPA carbon impact. For more on this trend see *Carbon Savings Redefine Impact of Corporate Energy Deals* ([web](#) | [terminal](#)).

### Alphabet reaches new heights and invests in emerging forms of generation

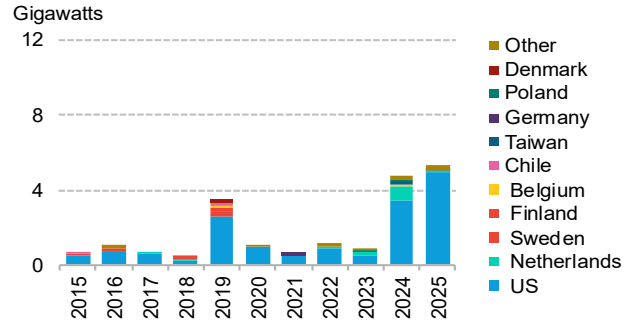
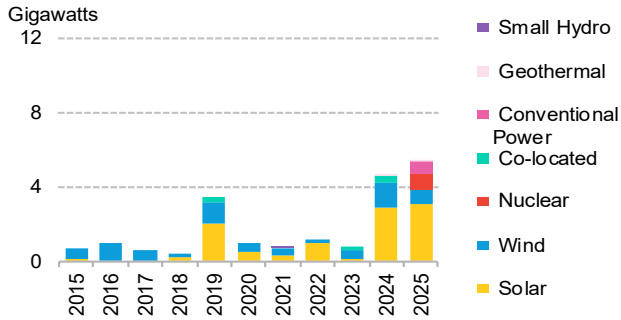
Alphabet set a new record for clean PPAs in 2025 in gigawatt terms, with 5.4GW in contracts signed, while simultaneously diversifying the sources of power across several kinds of baseload generation. Almost a third of Alphabet’s contracts came from non-intermittent clean energy such as nuclear, hydropower and geothermal. Some of the firm’s investment in firm generation was from novel kinds of energy generation such as Kairos Power’s advanced fission reactor and Commonwealth Fusion System’s fusion reactor. Both projects are unlikely to deliver any kind of power in the near term.

### Microsoft’s unusual disclosure strategy leads to varying levels of procurement

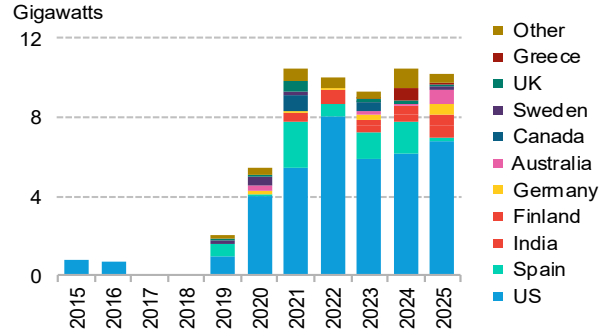
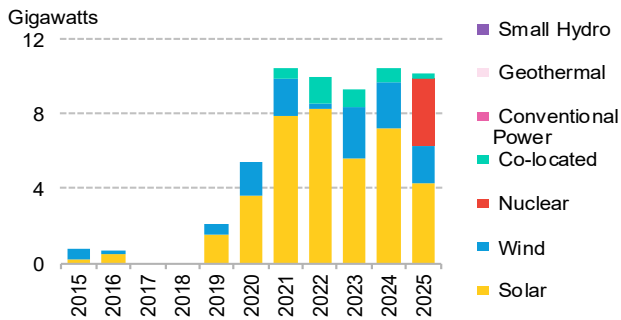
Microsoft’s procurement in 2024 exceeded that of Alphabet and Meta but this year, its procurement announcements dropped – as Meta and Alphabet’s accelerated – due to its policy of announcing PPAs when they are commissioned rather than when they are signed. Despite this, the company said it reached a key sustainability milestone by consuming 100% renewable energy on an annual basis in 2025. The technology giant achieved the goal five years ahead of its original target of 2030. While Microsoft was also the first of the hyperscalers to procure nuclear energy via a PPA – its 850MW deal to restart Three Mile Island – it was the only one not to procure any nuclear capacity in 2025.

Figure 14: Clean PPA activity by hyperscalers

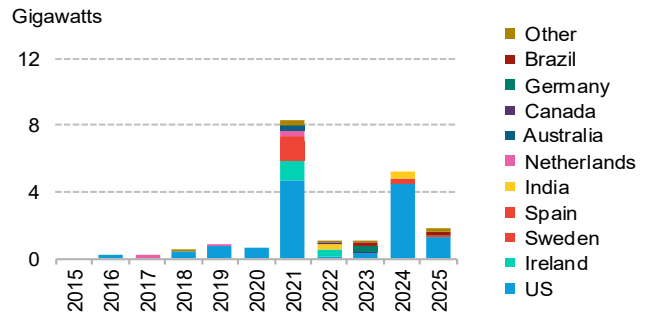
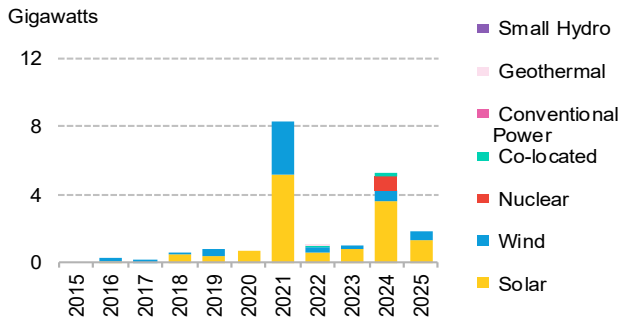
### Alphabet



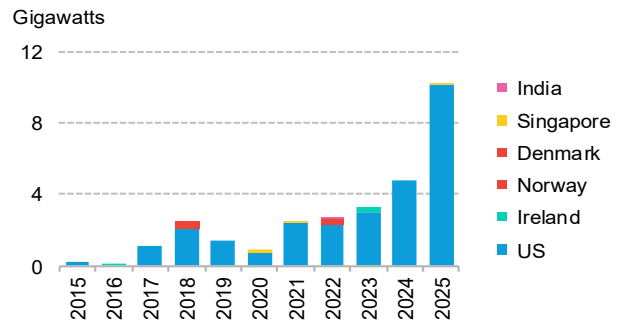
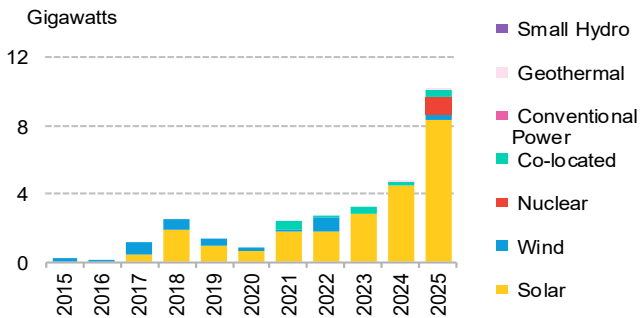
### Amazon



### Microsoft



### Meta

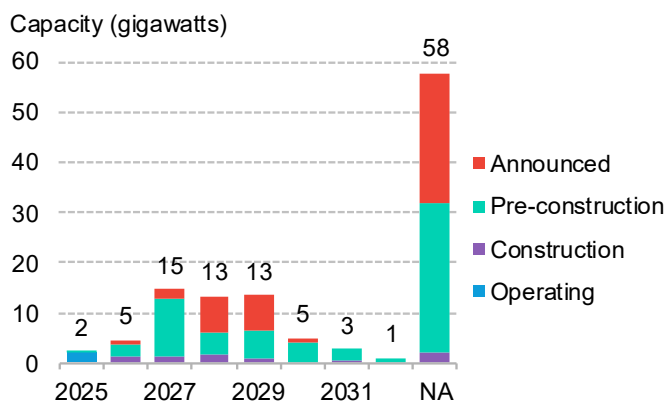


Source: BloombergNEF. Note: PPA is Power Purchase Agreement.

## 2.2. On-site gas generation announcements at data centers in the US now exceed 100GW

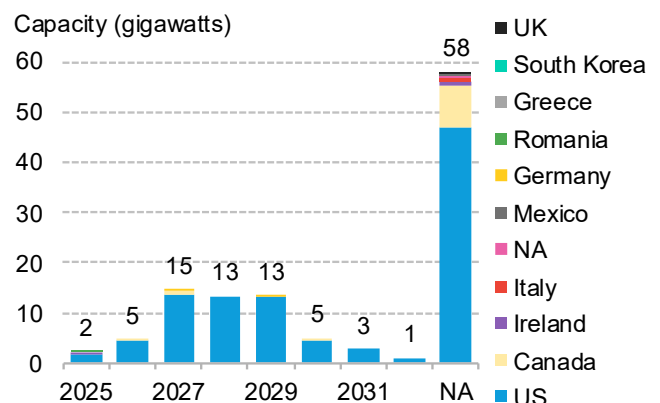
Since the last iteration of this report in 2025, it has become increasingly clear that on-site gas generation will contribute a significant share of electricity to the data center fleet, particularly in the US. That report tracked around 9GW of on-site gas capacity across five sites. BNEF has now identified 114GW of gas-fired capacity across 115 projects under development or already online today to specifically serve data center loads on-site (Figure 13).

**Figure 15: On-site gas generation capacity announcements for data centers, by status and proposed start date**



Source: BloombergNEF, Global Energy Monitor. Note: NA indicates undisclosed commissioning date.

**Figure 16: On-site gas generation capacity announcements for data centers, by market and proposed start date**



Source: BloombergNEF, Global Energy Monitor. Note: NA indicates undisclosed commissioning date.

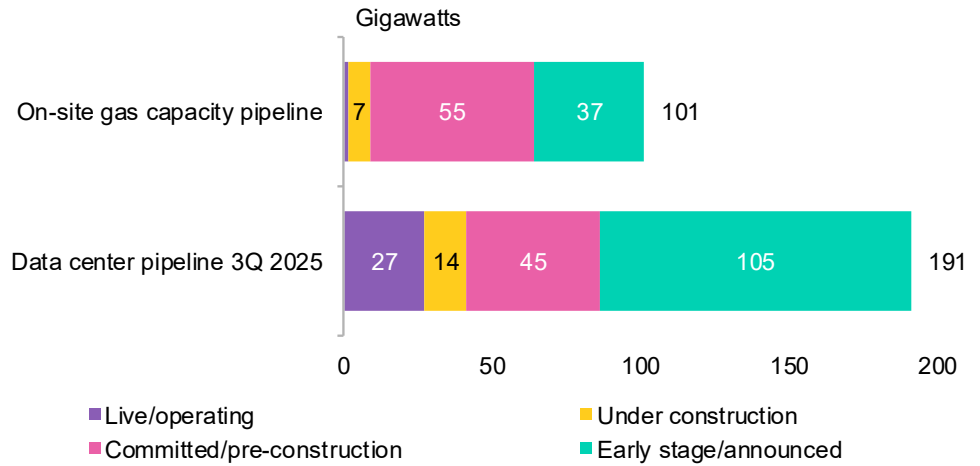
Nearly 90% of announced on-site gas generation capacity comes from US projects, totaling about 100GW. Limited activity was also tracked in a handful of non-US markets such as Canada, Ireland, Italy and Mexico (Figure 16). Within the US, the leading state for such projects is Texas. Some 38GW of announced capacity was tracked within Texas. Wyoming, Utah and New Mexico account for smaller shares of capacity.

The announced pipeline for on-site gas generation in the US is equivalent to 52% of BNEF's total data center pipeline in the market as of 3Q 2025, highlighting the key role of on-site gas in the digital infrastructure boom (Figure 17). This share drops to 28% when only considering gas projects that have a specific timeline associated with their construction. It is important to note that as per our most recent forecast BNEF estimates, 106GW in data center load will come online by 2035 based on historic project attrition rates<sup>2</sup>. Much of this gas capacity is therefore a bet on a more bullish scenario for data center buildout.

For more see: [Data Centers Eye 114GW of On-Site Gas Amid Grid Delays \(web | terminal\)](#).

<sup>2</sup> Due to lags in data reporting, this 106GW outlook is based on project pipeline as of 1Q 2025 and is expected to be revised upwards in an upcoming update to BNEF's upcoming data center forecast.

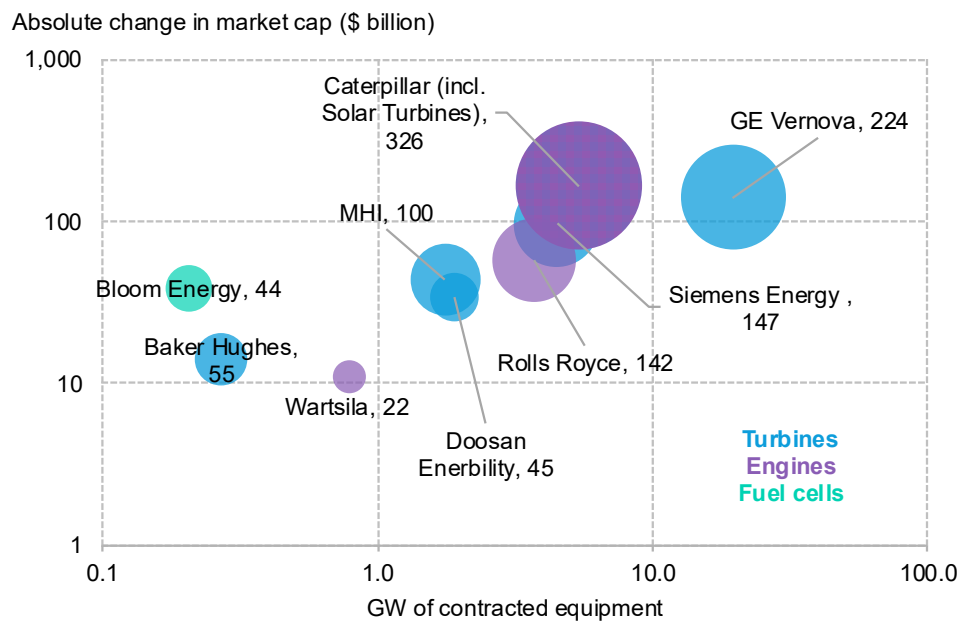
Figure 17: US data center IT capacity versus onsite gas capacity, by stage



Source: DC Byte, BloombergNEF.

All major manufacturers of gas turbine and engine equipment have seen big bumps in their market capitalizations this year, as investors are betting a supply crunch for generation equipment will boost margins as demand grows. Caterpillar has seen the biggest change in absolute market cap over the last 12 months, thanks to exposure in both engines and turbines (through its subsidiary Solar Turbines). It is closely followed by GE Vernova and Siemens Energy, which have contracted the first and third most gas capacity for on-site data centers respectively (Figure 18).

Figure 18: Change in market capitalization over last 12 months versus gigawatts of contracted equipment for on-site power generation at data centers



Source: BloombergNEF. Note: Bubble size indicates market cap as of March 13, 2026 and is denoted with a text label in billions. 12-month period covers March 13, 2025 to March 13, 2026.

### 3. Financing trends

Financial markets have shown significant volatility in the past few months, with fears of an “AI bubble” weighing on stocks in November and December, and a sell-off happening in February as investors worried AI would be wildly successful at replacing existing software businesses. Despite these competing narratives, and some nervous commentary, large volumes of capital continues to flow into digital infrastructure. Expectations of hyperscalers’ capex have climbed significantly since the last iteration of this note in September. Debt offerings to finance this buildout are frequently oversubscribed, including Google’s headline-grabbing century bond. The mood is not universally bullish, however, with some hyperscaler share prices sliding after higher-than-expected capex plans were announced in their earnings calls. Oracle has lost around half its market cap since September and there are long-term, open questions around asset values.

#### 3.1. Data center capex outlooks jump by over 50%

Capex projections for data center companies were revised up by a significant margin in the most recent round of earnings calls. The most significant jump came from Google, which announced plans to double its capex year on year to between \$175 and \$185 billion.

Since the last iteration of this note, BNEF has extended its tracking to all public companies that rank among the 20 largest data center operators (Figure 19). This gives a more developed picture of the overall investment landscape, and also highlights the huge share of capex tied to the hyperscalers. Amazon, Google, Meta and Microsoft together represented 80% of capex in the group analyzed, a share that is expected to remain broadly flat through 2028, according to consensus estimates of analysts.

Figure 19: Capital expenditure of largest publicly owned data center operators

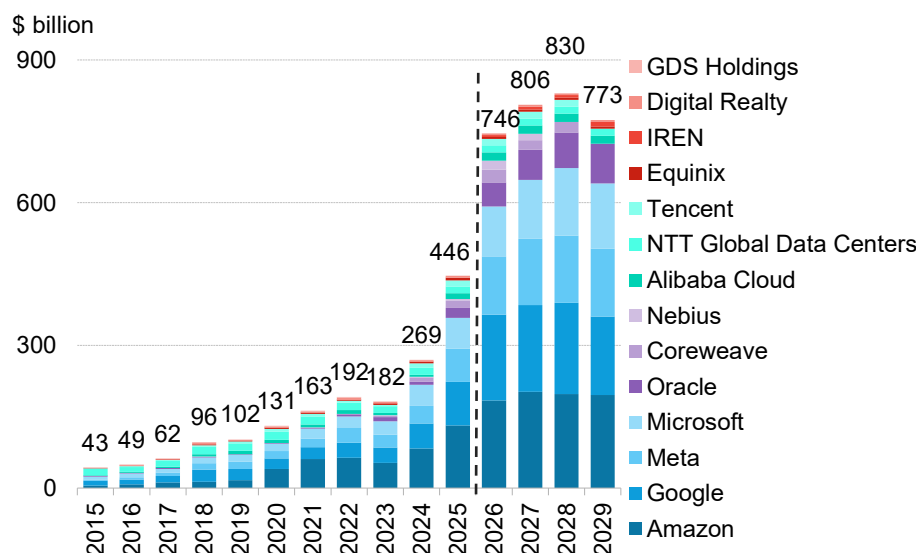
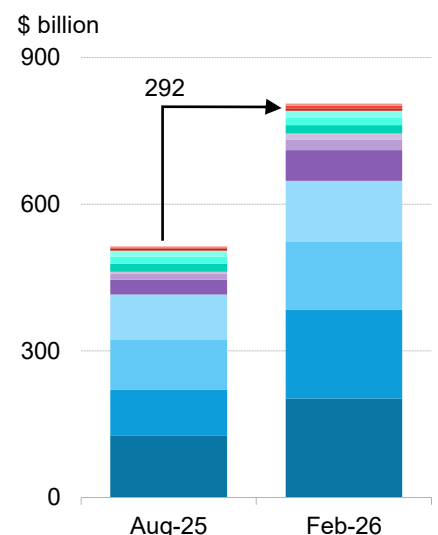


Figure 20: FY 2027 capital expenditure of largest publicly owned data center operators, change in consensus view

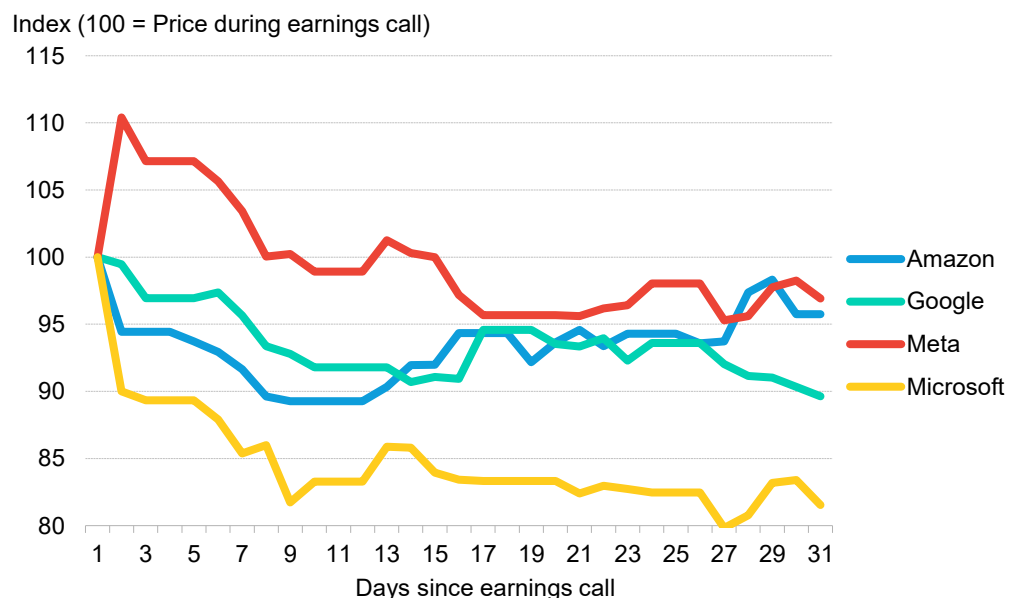


Source: BloombergNEF. Note: Uses headline capex from FA<GO>. Companies do have other capex activities, most notably Amazon. Only includes public companies that are in the top 20 owners of live IT capacity and have 200 megawatts or more of IT capacity under construction, according to BNEF’s Global Data Center Live IT Capacity Database (1.2.0). \*2026-2028 figures are third-party analyst estimates collated by Bloomberg and available from FA<GO>.

Capex of the four hyperscalers – Alphabet, Amazon, Meta, Microsoft – in FY2025 was marginally higher than analyst consensus, at \$357 billion versus the \$351 billion expected. This made up 80% of the \$446 billion BNEF tracked in total, although it is worth highlighting that the hyperscalers are also the most likely to have non-data-center capex, most notably Amazon.

The biggest surprise from Q4 earnings calls was the upward revision in projected capex. Between August 2025 and February 2026, analyst consensus capex for FY2027 has been revised upwards by nearly \$300 billion – a 57% hike (Figure 20). The response of equity investors has been largely negative. Microsoft’s stock price fell significantly following its last earnings call, losing 10% of its value in a single day of trading. Shareholders were skeptical of its high capex to support AI, which is unlikely to generate positive cashflow for several years (see Section 0). Meta, which released its earnings call the same day, rallied on higher-than-expected earnings despite the fact it also revised capex upwards. Google and Amazon both slid following their earnings calls, on February 4 and February 5 respectively, although this was more pronounced for Amazon, which closed down 6% on Feb 6.

**Figure 21: Share price of hyperscalers in month following most recent earnings call**



Source: BloombergNEF, Bloomberg Terminal. Note: Google refers to parent company Alphabet. Meta and Microsoft reported earnings on January 28, Google on February 4 and Amazon on February 5. Google, Meta and Amazon reported Q4 FY25 earnings. Microsoft reported Q2 FY26.

**Battles over depreciation schedules are influencing financing appetite for data centers**

The question of GPU depreciation made headlines in November 2025, with a number of related concerns tied together into one debate. At present, BNEF considers this to be a neutral signal on data center buildout, but an important indicator to understand appetites for financing data centers, and thus a signal for capital investment.

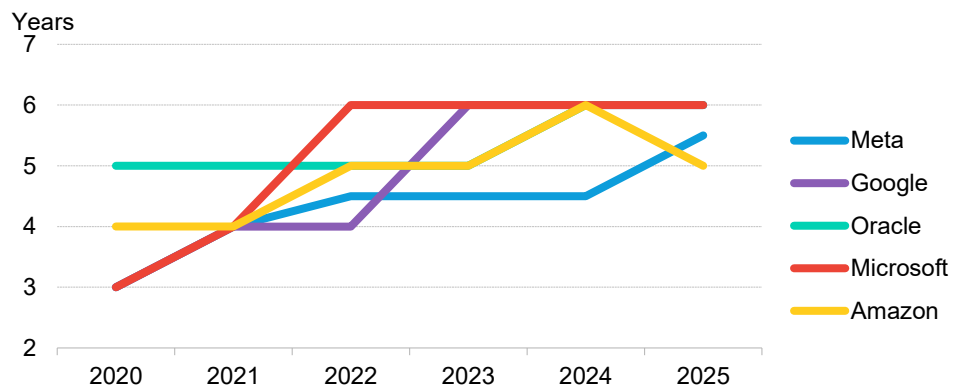
There are three things that people might be referring to when talking about GPU depreciation:

- **Physical depreciation:** The time over which a GPU physically degrades, as a result of use.

- **Technical depreciation:** The time over which the market value of a GPU declines, due to more advanced hardware being released.
- **Accounting depreciation:** the time over which a company spreads the cost of a capital asset, recording a portion of that cost as an expense each year of its useful life.

**Accounting depreciation** was raised by investor Michael Burry (made famous by book *The Big Short*) who said that the big tech companies are extending the useful life of GPUs to artificially boost earnings. This was based in the observation that the five largest tech companies have all increased the “useful life” of chips since 2020, despite the fact that Nvidia has in that time shifted from a two year release cycle to a one year release cycle (Figure 22). Whether or not this is a problem is a combined function of the physical, and technical, depreciation of these assets.

**Figure 22: Accounting depreciation period for GPUs at hyperscalers and Oracle**



Source: BloombergNEF, company SEC filings.

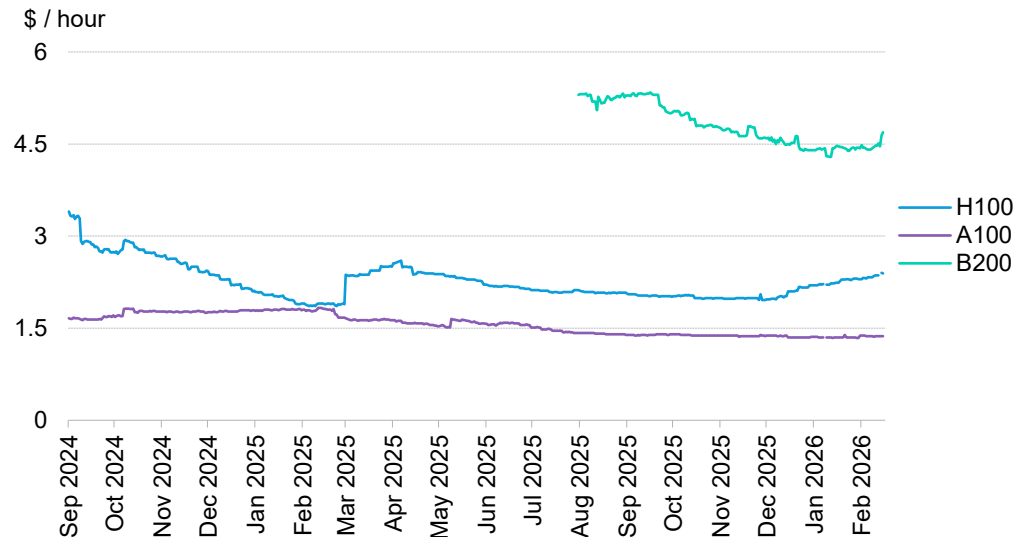
**Physical depreciation** for state-of-the-art GPUs is poorly understood. The power draw of the current generation of Nvidia chips is an order of magnitude higher than those in a typical data center. The workloads of a GPU are significantly “spikier” than those of a CPU (the chips traditionally used in data centers), meaning that they undergo larger and more rapid power swings than most existing chips. There are open questions about the material impact of this over multi-year horizons – not just to the silicon, but to wider hardware including packaging, interconnects and power delivery.

**Technical depreciation** is similarly unknown. Nvidia announced its intention to shift from a biannual release cadence to an annual one in its Q1 FY2025 earnings call, which will theoretically double the rate at which chips are superseded. The jump between chip generations is generally massive, both on overall compute performance and on energy efficiency.

**One signal that depreciation concerns are overblown** is that AWS and Google are both using chips from 2020 and earlier. AWS CEO Matt Garman recently said it has never retired an Nvidia A100 server, a chip first unveiled in 2020. Google Vice President Amin Vahdat has said that demand is out-stripping supply, with Google still operating 8-year old AI chips at “100% utilization”. Chip workloads have changed in this period, following a fairly typical trajectory – new chips are used for model training in a period that may be as short as 12-18 months, and then move to (typically less intensive) inference workloads.

Despite current demand for chips five-plus years old, this is not guaranteed to remain the case. The rental price of GPUs is generally volatile (Figure 23). For now, demand remains strong – and, in the case of H100s, prices recently rallied – but there are challenges in forecasting what this will look like in a year.

Figure 23: GPU hourly rental market spot price, Nvidia flagships



Source: BloombergNEF, Bloomberg Terminal, Silicon Data. Note: H100 refers to SDH100RT Index on Bloomberg terminal. A100 refers to SDA100RT Index. B200 refers to SDB200RT Index.

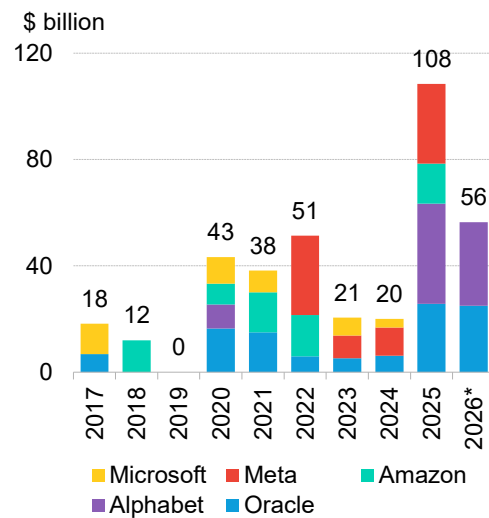
### The growing role of corporate debt

Data center developers continued to raise capital via debt markets, with BNEF tracking \$108 billion in bond issuances from Alphabet (Google’s parent company), Amazon, Meta and Oracle in 2025, and \$56 billion so far in 2026, from Alphabet and Oracle. Microsoft has not issued bonds since a \$3.3 billion offering in 2024. This is a change in strategy from companies which have very strong cash flows and have historically chosen to finance capex on balance sheet. The pivot to bond markets is a function of the scale of capital required for the current wave of infrastructure build-out.

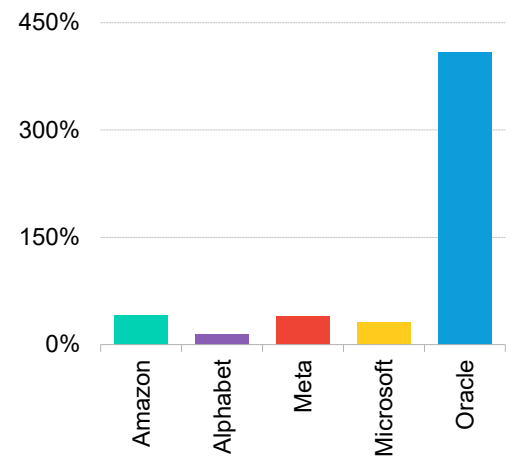
The value of bonds issued by the hyperscalers (including Oracle) reached \$108 billion in 2025 – more than double the \$51 billion raised in 2022, the second-most active year in the past decade for bond issuance by these companies (Figure 24). Despite the growing reliance on debt, these companies typically have low debt-to-equity ratios, due to strong cashflows and historical aversion to corporate debt. Oracle stands out for having a much higher debt-to-equity ratio than its peers, which was a factor in its shaky stock market performance over the past few months.

*The corporate debt discussed in this section is not inclusive of project finance, which is frequently used by data center developers to develop individual sites.*

**Figure 24: Bond issuances by hyperscalers including Oracle**



**Figure 25: Debt-to-equity ratio of hyperscalers including Oracle**



Source: BloombergNEF, Bloomberg Terminal. Source: BloombergNEF.  
 Note: \*2026 is YTD. Uses data from FIW <GO>.

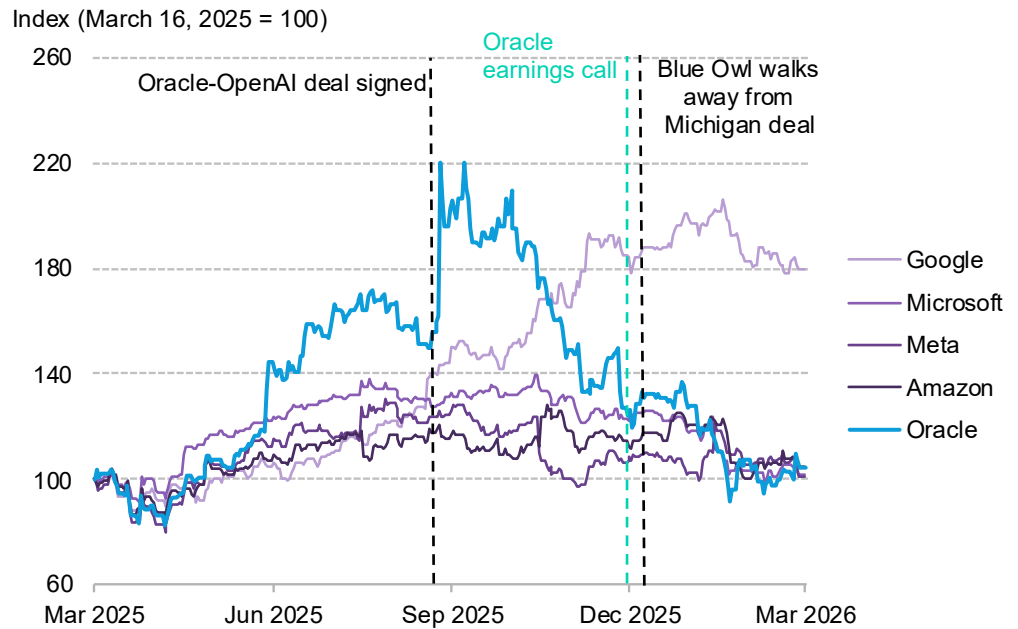
Google issued £1 billion (\$1.4 billion) of 100-year bonds in early February, grabbing headlines despite being a small part of the total \$32 billion offering. Century bonds are a highly unusual offering for any company, and especially rare for a tech firm – BNEF is aware of just two prior sales, IBM in 1996 and Motorola in 1997. This triggered a lot of commentary, much of which was broadly negative. Despite speculation, the move makes sense – Google has announced plans to increase capex to as much as \$185 billion in the next financial year (section 3.1). As such, it makes sense to tap all available sources of capital. A century bond can attract interest from long-term institutional investors, such as university endowments and life insurance providers. It proved popular, with the £1 billion offering nearly 10-times oversubscribed.

**3.2. Oracle is increasingly a bellwether for AI financing appetite**

Oracle, a company with its roots in database software, has in many ways been at the center of AI-focused datacenter buildout. Even compared to the other big tech companies, it had a significant rally in the first half of 2025, more than doubling its valuation between March and August (Figure 26). Since then, it has slid significantly, finishing down 1.5% on its valuation from a year prior.

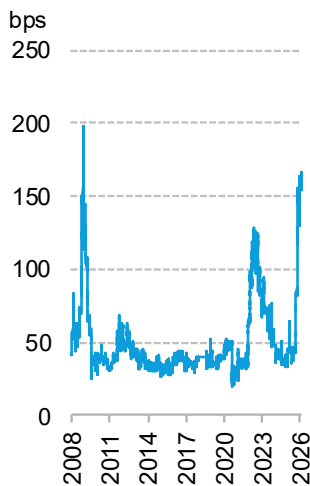
Oracle is closely involved with OpenAI, both through its role in Stargate and more recently via additional offtake agreements. This included a \$300 billion, five-year contract announced on September 10. In its earnings call on December 10, Oracle announced \$455 billion in remaining performance obligations (RPO). RPO refers to the total value of contracted revenue that has not yet been recognized, as the contracts refer to future years. Of this, over \$300 billion – two thirds – is tied to OpenAI. This would represent a high level of concentration risk regardless of counterparty. The risk is especially large with a company like OpenAI that expects to burn in excess of \$100 billion in the coming years (see section 0).

Figure 26: Share price, large tech companies



Source: BloombergNEF, Bloomberg Terminal.

Figure 27: Credit default swap spread for Oracle



Source: BloombergNEF, Bloomberg, ICE Data Services. Note: bps refers to basis points. 1 bp = 0.01%. Shows spread on 5-year senior unsecured Oracle debt (ORCLCP CDS USD SR 5Y D14).

As “AI bubble” concerns escalated through November 2025, credit default swaps (CDS) on Oracle bonds climbed to their highest value since the financial crisis. A CDS is a credit derivative which functions like insurance on the underlying debt. They are typically purchased as a way of hedging risk and speculation. Investor appetite for CDS against tech companies is generally growing, but it has been the most extreme in Oracle’s case, in a sign that Oracle’s ability to meet its financial obligations is a growing concern for investors. This is partially tied to concerns around Oracle’s reliance on OpenAI.

Blue Owl, an infrastructure financier, decided not to fund a 1GW Oracle data center in Michigan in December 2025, in a move which weighed on Oracle’s share price (Figure 26). Blue Owl has previously backed several data centers tied to Oracle, including the Stargate campus in Abilene, Texas and its New Mexico campus, under development. The FT, citing people close to the deal, reported that lenders “pushed for stricter leasing and debt terms amid shifting market sentiment around enormous AI spending including Oracle’s own commitments and rising debt levels.”

Oracle raised \$25 billion in a heavily oversubscribed bond offering on February 2. The level of interest – investors placed orders for more than \$129 billion – represented a vote of confidence in the company, after its debt had been “trading more like junk bonds” for several weeks.

Similar to the neoclouds, Oracle increasingly serves as an indicator of wider market attitudes toward AI infrastructure generally, and OpenAI in particular. Its share price, and ability to raise capital through debt markets, reflects broader views on long term AI demand. In the past six months, it has suffered from some jitters, but overall market support remains strong. Any further slippage is likely an indicator of attitudes on Wall Street on AI infrastructure souring.

### 3.3. Unusual deal structures play increasingly important role

The first half of 2025 saw multiple data center project financing agreements in the tens of billions of dollars, including a \$29 billion special purpose vehicle (SPV) for a Meta data center. BNEF did not track any similar-sized packages in 2H 2025, although other unusual debt structures were used to finance new infrastructure, mostly by the neoclouds. These were:

- **Big-tech-backed deals:** Agreements that are ultimately underwritten by a very large tech company, such as Nvidia or Google. This involves the data center developer using an investment-grade partner as a guarantor on a contract, to secure financing.
- **Convertible notes:** Debt that can convert into equity at a pre-set conversion price is a popular form of capital for the public neoclouds, with Nebius raising \$2.8 billion in convertible notes in September 2025, and CoreWeave raising \$2.6 billion in December. These are a desirable financing instrument for high-growth, capex-intensive companies because they typically carry lower cash coupons than straight debt and defer equity dilution. However, convertibles still add leverage and refinancing risk. If the share price does not trade above the conversion price at maturity, the notes behave like conventional debt.
- **GPU-backed-debt:** A loan secured against the advanced chips being deployed in data centers. CoreWeave helped popularize the structure starting in late 2023, and issuance activity accelerated notably through 2025, including multiple large secured facilities. The FT reports that these loans offer returns in the high single digits to mid-teens, which is generally higher than traditional tech debt. Both neoclouds and established players have been raising capital in this manner, often via SPVs that take on the debt and lease the GPUs to the data center operator. This can keep the debt off balance sheet. Although currently very popular, accelerated GPU depreciation rates do pose a potential risk to this as a financing structure in the long term. This is mitigated by the fact that the leases are typically under a “take or pay” agreement, meaning tech companies have limited ability to terminate the contract early without significant penalties.

These novel financial instruments are an indicator both of the volume of capital which is required for data center buildout and the growing challenges the neoclouds face in accessing capital. The willingness of investors to provide debt to the neoclouds will be an important indicator of medium-term buildout, and serves as a proxy for wider investor sentiment. If data center demand remains high and GPUs hold their value, GPU-backed debt is likely to become more standardized and lower cost. Conversely, if depreciation accelerates or utilization assumptions soften, access to debt may become a constraint on infrastructure development.

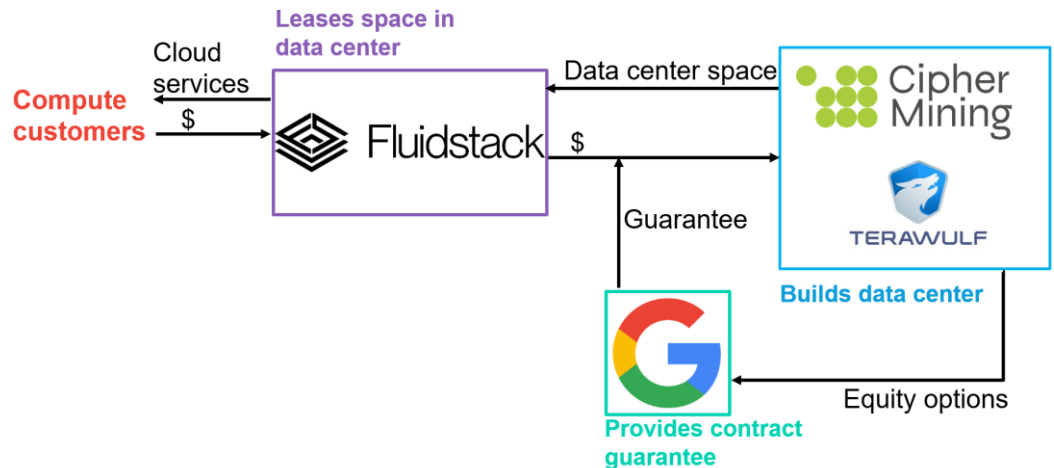
#### Google and Nvidia backstop neocloud contracts

Through conversations with industry players, BNEF understands that there are growing challenges in securing leverage on projects purely tied to the neoclouds. Having a tech giant backstop the tenant’s obligations means that it can effectively use the backstop provider’s credit rating, accessing better terms. Google agreed to act as a credit backstop for at least two computing leases involving neocloud Fluidstack.

In these agreements, Fluidstack leases data center space from a data center developer. The examples this year were Terawulf – providing 360MW at Lake Mariner in New York – and Cipher Mining – providing 207MW at Barber Lake in Texas. This capacity is then sold by Fluidstack to customers with demand for GPU compute. In both cases, Google agreed to provide a financial guarantee: if Fluidstack is unable to meet its contractual obligations, Google will step in and take over the lease. Google received warrants – a type of option – in the host companies, in return for providing this backstop. Both contracts are for capacity over a 10-year period. If demand for AI

infrastructure continues to rise in this time, these warrants give Google equity upside without requiring it to commit to long-term capacity offtake. The structure also strengthens its influence within the broader AI infrastructure ecosystem.

**Figure 28: Structure of Google-Fluidstack offtake deals**



Source: BloombergNEF, company websites.

The backstop structure is still relatively unusual, although not unique to Google. Nvidia has provided a similar lease-obligation guarantee to CoreWeave. If challenges in financing deals solely tied to the neoclouds remain, this type of arrangement may become more common. That would further increase the importance of the hyperscalers in the overall AI infrastructure stack and could concentrate risk around a small number of companies.

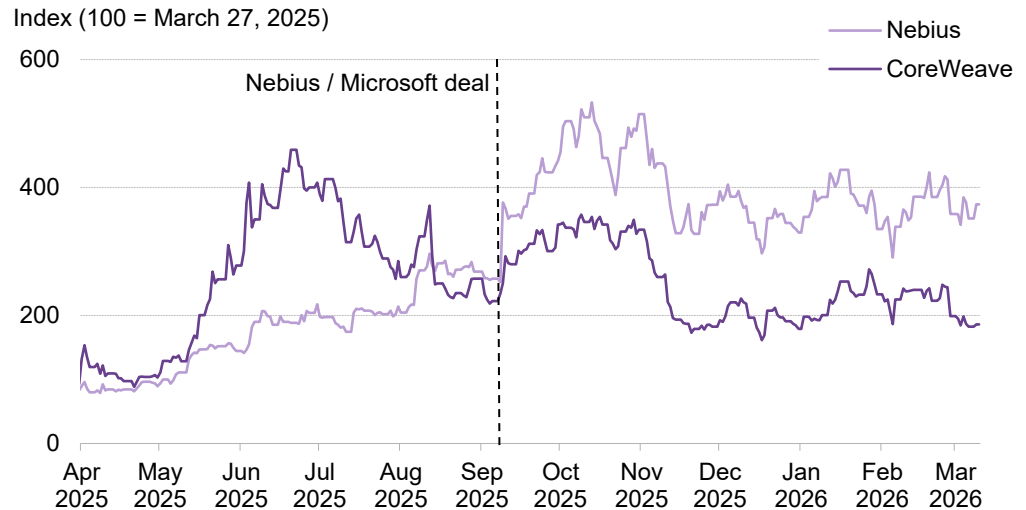
#### 4. Leasing drives confidence in neoclouds as depreciation remains a risk

The health of neocloud companies – notable for their GPU-as-a-service (GPUaaS) business model, and high optimization for AI workloads – continues to be an important indicator of broader market sentiments towards demand for AI compute. In the previous iteration of this note, BNEF included Crusoe, CoreWeave, Nebius Group and Lambda. On this iteration, this group is extended to also include Nscale and Fluidstack.

A flurry of new leases signed between hyperscalers, especially Microsoft, and neoclouds in the second half of 2025 drove positive sentiment toward the companies, although they were hit especially hard by GPU depreciation concerns (section 3.1).

The share prices of CoreWeave and Nebius are closely correlated, with major events for one company impacting the other’s share price. Investors remain jittery, with price swings common as sentiment around the wider health of the AI sector moves. CoreWeave has slightly underperformed the wider US market in the past six months, although it remains significantly elevated since its IPO on March 27 2025. Nebius continues to significantly out-perform the wider market, boosted by a major offtake agreement with Microsoft signed in September 2025, although it has slid marginally since its peak (Figure 29).

**Figure 29: Share price of neoclods since CoreWeave IPO**



Source: BloombergNEF, Bloomberg Terminal.

**Microsoft and Meta sign the most neocloud leases in recent months**

Offtake contracts signed by hyperscalers are a strong positive indicator of demand for neocloud services. Between September and November 2025 alone, BNEF identifies over \$50 billion in disclosed commitments from hyperscalers. Microsoft accounted for four agreements (Table 4). The majority of the contracts are on five-year terms, with capacity being delivered from 2026. For neoclods, hyperscaler offtake has been critical to securing project financing and scaling buildouts, although these contracts are for much shorter periods than total asset life, which is particularly relevant given the discussion of GPU depreciation rates.

For hyperscalers this represents a less capital-intensive path to deploying AI capacity. Using relatively short contracts means that they can allow contracts to expire at maturity should demand be lower than expected, so that they are not left with potentially stranded data center capacity. Microsoft is taking a more conservative approach to infrastructure buildout than its hyperscaler peers, signified by its lower capex figures (Figure 19) and greater use of leasing.

**Table 4: Data center capacity contracts between hyperscalers and neoclods, 2025**

Date	Offtaker	Provider	Size (\$ billion)	Term length (years)	Start date	Location of data center	Source
September 8, 2025	Microsoft	Nebius	17.4, expandable to 19.4	5	In “several tranches” through 2025, 2026.	Vineland, New Jersey	(SEC)
September 17, 2025	Microsoft	NScale	6.2	5	Staged deployments beginning in 2026	Narvik, Norway	(PR Newswire)
September 25, 2025	Meta	CoreWeave	14.2	6, extendable to 7	Not disclosed	Not disclosed	(SEC)
October 15, 2025	Microsoft	Nscale	14*	5	Phased delivery from 1Q 2026	Texas, Portugal, UK, and Narvik	(Nscale)

No portion of this document may be reproduced, scanned into an electronic system, distributed, publicly displayed or used as the basis of derivative works without the prior written consent of Bloomberg Finance L.P. For more information on terms of use, please contact sales.bnef@bloomberg.net. Copyright and Disclaimer notice on page 28 applies throughout.

Date	Offtaker	Provider	Size (\$ billion)	Term length (years)	Start date	Location of data center	Source
November 2, 2025	Microsoft	IREN	9.7	5	2026	Childress, Texas	(SEC)
November 3, 2025	Microsoft	Lambda	"Multibillion", not disclosed	"Multi-year", not disclosed	Not disclosed	Not disclosed	(Lambda)
November 3, 2025	Amazon	Cipher Mining	5.5	15	August 2026	Black Pearl, Texas	(SEC)
November 11, 2025	Meta	Nebius	3	5	November 2025 – February 2026	Not disclosed	(Nebius)
March 16, 2026	Meta	Nebius	"Up to" 27	5	2027	Several sites "across multiple locations"	(Nebius)

Source: BloombergNEF, SEC filings, company press releases. Note: \*October 15 Microsoft / Nscale deal did not disclose a deal value; the value provided is an estimate based on the volume of chips and terms of September 17 deal.

### Big Nebius contract boosts investor confidence

The Nebius-Microsoft deal, worth at least \$17.4 billion over a five-year period, drove a 61% rally in Nebius' share price the next trading day. This boosted confidence in the neoclouds more broadly, with CoreWeave posting 32% gains on the news.

Under the agreement, Nebius will provide over 100,000 Nvidia GPUs for Microsoft at a site in Vineland, New Jersey until 2031, with Microsoft having an option to acquire additional services and/or capacity, increasing the contract value to \$19.4 billion. Assuming the contract has a flat structure, this translates to annual revenues of at least \$3.5 billion. For comparison, Nebius' total revenue for 2025 was \$530 million.

### Depreciation fears hit the neoclouds

CoreWeave and Nebius were both hit particularly hard by GPU depreciation fears (section Figure 29) through late October and early November. This wider conversation was partially triggered by Nvidia CEO Jensen Huang [commenting](#) that "When Blackwell [next generation chips] starts shipping in volume, you couldn't give Hoppers [previous generation chips] away" in early October. In their worst two-week period, the companies fell by 42% and 36% respectively, mostly wiping out the gains seen since the Microsoft deal in early September. The fact that share prices have not seen significant growth since this point is an indicator that some worries remain.

Shifts in share prices remain frequent – bond offerings in December and wider worries around an overbuild of AI infrastructure in January both caused sudden dips, although prices have rebounded quickly. Broadly speaking, Nebius and Coreweave are trading at the level they hit after their November slide.

### Nscale argues that digital sovereignty could drive tailwinds for European neocloud developers

Nscale has grown rapidly since being spun out from Australian bitcoin miner Arkon energy in 2024. The London-based neocloud has focused on developing AI infrastructure in Europe, and is one of the largest AI-optimized developers in the continent. It raised some of the largest ever series B and series C rounds in Europe, at \$1.1 and \$2 billion respectively. Its March 2026 series C valued it at \$14.6 billion – the highest of the privately held Neoclouds tracked in this report. Further to these large equity raises, it [also raised](#) \$433 million via simple agreement for future

equity (SAFE) financing in October 2025, and a \$1.4 billion delayed draw loan to fund GPU deployment across Norway, Portugal, Iceland and the UK.

Whilst the company is active globally – it has a large data center in Texas – it is increasingly positioning itself as a European “sovereign AI” champion. Sovereign AI typically refers to developing and deploying AI using in-region data, infrastructure and governance to meet data-residency and security requirements. This is an issue of growing political importance in Europe, as AI is viewed as having long-term economic significance, and concern about relying on US hyperscalers for critical workloads grows. This provides a potential tailwind for data center development, and developers, in the region. The political will is evidenced by the EU’s EuroHPC “AI Factories” program, and national AI programs in key markets such as France and the UK.

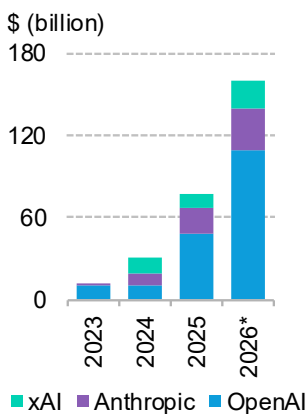
In the medium term, sovereign AI ambitions may allow other non-American neoclouds to access significant pools of capital, and win contracts – both in Europe, and potentially in other regions such as the Middle East. It will not, however, offer a guaranteed fix for other headwinds, such as poor power availability or complex land permitting.

For more on the regional headwinds and tailwinds of data center development in Europe see *Data Center Development: Regional Drivers and Hurdles* ([web](#) | [terminal](#)).

## 5. AI lab fundamentals

Appetite for equity investment into the frontier AI labs has remained extremely high. This is an indicator that investors still believe in the potential of these companies and are willing to finance additional cash burn. However, as spending commitments continue to climb and projected revenue growth is steep, there is little room for error as these companies scale.

**Figure 30: Fundraising for leading AI labs**



Source: BloombergNEF.

Note: Fundraising includes both debt and equity rounds. 2026 is up to February 28.

### 5.1. AI labs raise \$160 billion so far in 2026

BNEF tracked four fundraising rounds for AI labs in the past six months, totaling \$162 billion. Three of these were in the space of two months. These were:

- OpenAI**, the maker of ChatGPT, announced \$110 billion of new investment on February 27. This round alone is more than what the entire climate-tech sector raised in 2025, across public and private equity markets. The company received a post-money valuation of \$840 billion – a 180% increase over the valuation of \$300 billion it had in August 2025. The round was led by Amazon, which committed \$50 billion. The first \$15 billion was paid immediately, with the remaining \$35 billion tied to unspecified conditions, to be paid over the coming months. Under the deal, OpenAI has committed to leasing Amazon Trainium chips, with \$100 billion to be paid over the next eight years. This is additional to the \$38 billion over seven years OpenAI had previously announced. Nvidia and SoftBank each committed \$30 billion.
- Anthropic** announced it had closed a \$30 billion Series G round on February 12, giving it a post money valuation of \$380 billion. The round included contributions from Nvidia and Microsoft, as well as the Qatar Investment Authority. While it was not the first time that Anthropic has taken money from the Middle East, Anthropic has previously expressed concern about accepting investment from the region.
- xAI** merged with SpaceX in an all-stock deal in February, reportedly valuing xAI at \$250 billion, alongside SpaceX’s \$1 trillion. Both companies are controlled by Elon Musk. In January 2026, xAI raised \$20 billion at a \$230 billion valuation, with notable investments including up to \$2 billion from Nvidia, according to [Bloomberg News](#) and \$3 billion from Saudi AI company Humain.

- **Mistral**, a European AI lab, announced it had raised €1.7 billion (\$2 billion), giving it a post-money valuation of €11.7 billion (\$13.7 billion). The round was led by European semiconductor equipment manufacturer ASML, with Nvidia also contributing.

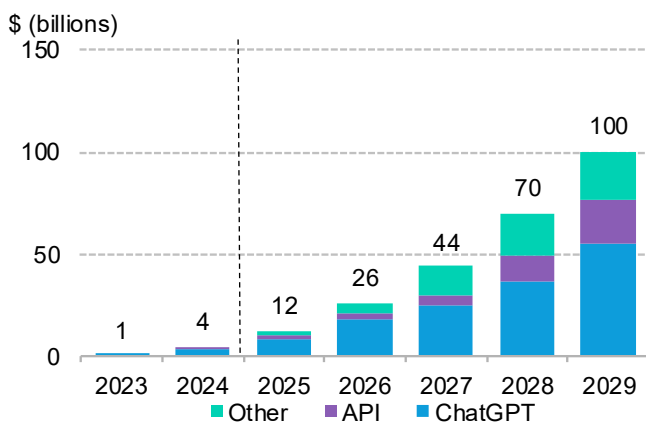
### 5.2. Spending strategies diverge, risks lie ahead

The top labs are increasingly diverging in their revenue strategies as cash burn remains high across the board. They do not expect to be profitable for years. However, gross margins are reportedly positive for current models, suggesting current capabilities could be profitably delivered. Ongoing negative cash flows relate to spending on the development of new training infrastructure and model development.

#### OpenAI bets big on rising revenues

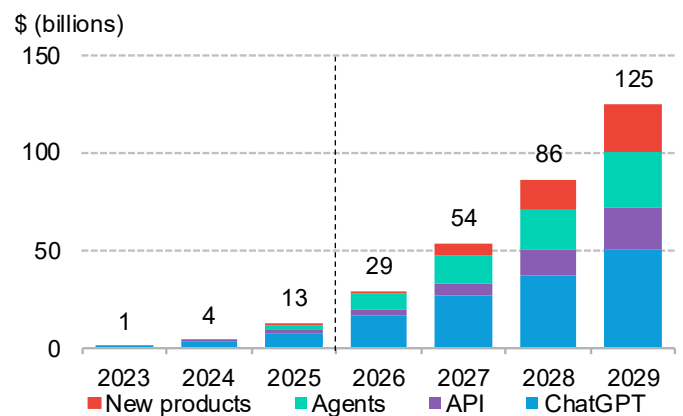
OpenAI has the most aggressive revenue growth projections of the top AI labs. It revised its already high cash burn projections upwards in February, according to [reporting](#) by The Information. Cash burn in 2025 was \$8 billion, approximately \$0.5 billion below its prior forecast. It expects to burn \$25 billion this year, however, and \$57 billion next, an upward revision of \$30 billion. The company simultaneously raised revenue forecasts and continues to project positive cash flow of \$40 billion by 2030. This is the second time in a year that the company has increased its cash burn forecast, highlighting the rising cost of compute access.

Figure 31: OpenAI revenue projections, autumn 2024



Source: BloombergNEF, OpenAI via The Information. Note: Data is actual to 2024, projected from 2025.

Figure 32: OpenAI revenue projections, Winter 2025



Source: BloombergNEF, OpenAI via The Information. Note: New products includes free user monetization. Data is actual to 2025, projected from 2026.

OpenAI's high cash burn is a function of its very high spending commitments. In November 2025, CEO Sam Altman [said](#) that the company was "looking at commitments of about \$1.4 trillion over the next 8 years". In February 2026, it played down these claims, [telling investors](#) it is targeting \$600 billion in infrastructure spending to 2030. This is a much more minor revision than the headline figures suggest, as they refer to a four-year rather than eight-year time horizon, and it is likely that infrastructure spending will gradually ramp over that period.

#### Anthropic takes (slightly) more conservative approach to scaling

Anthropic is taking a slightly more conservative approach to spending than OpenAI and expects to turn cashflow positive sooner, in 2028. As part of this, it has committed to leasing less compute

than its rival. It reportedly expects to pay \$80 billion to the major cloud providers through 2029, and separately announced \$50 billion in infrastructure spending with Fluidstack.

The challenges in scaling an AI lab were highlighted in February by Anthropic CEO Dario Amodei, speaking about the importance of timing expenditures. The current projected revenue growth for AI labs is amongst the steepest in history. The Anthropic CEO is discussing growing at a 10x growth rate, while at a \$10 billion annual recurring revenue or ARR. There is uncertainty in this growth rate. Due to the lead-times on contracting data center capacity, companies need confidence in their revenue projections. Amodei said that relatively small errors in this projection could lead to bankruptcy.

The risk of mistiming revenues is most acute for the AI labs themselves – the big tech companies could likely absorb this kind of mis-calculation into their existing free cash flows for a period. This is also not a major risk to long term infrastructure demand – if an AI lab failed due to overextension, its data center capacity would likely still be attractive to other players in the market. That does not mean that there is no risk, and changes in cash flow projections, and business models, are important to understand the future of data center demand.

### 5.3. Monetization strategies and user adoption

The divergence in monetization strategies between the major AI labs has continued since the last iteration of this note. Growth rates are generally high, although there have been some misses. This points to rising AI demand, although the ultimate scale of that demand, and how it translates into energy demand, remains uncertain.

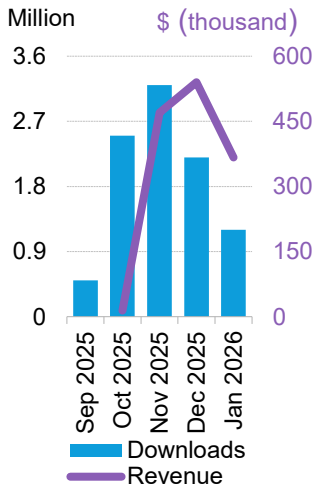
OpenAI has been pursuing its mass-market strategy, including by launching a consumer video app called Sora and introducing advertisements to the free tier of its platform. Recent reporting suggests a strategy change, although details are yet to emerge. Anthropic meanwhile is targeting the enterprise market and has grown its market share significantly. Google and xAI continue to integrate their AI tools with other products.

#### OpenAI questions consumer market positioning

OpenAI has been actively pursuing consumer growth, and seen gains – its user base continues to grow rapidly, with ChatGPT hitting 900 million weekly active users in February. This is 29% growth since July 2025, with growth in paying users even steeper – the company now has 50 million subscribers, up from 20 million in July. It has also begun efforts to monetize its non-paying subscribers, by introducing advertisements to its free, and low-cost “Go” offerings. This roll-out began in February, and so it is too soon to see the impact this has on margins or user numbers.

The company may be planning a major strategy shift, pulling back on its consumer bets, and targeting professional users instead, The Wall Street Journal reported a major strategy shift at the company on March 17. It could “deprioritize” certain “side quests”, to focus on business productivity offerings. This would bring its strategy more in line with rival Anthropic. Depending on implementation, this may translate into lower data center demand, although at present BNEF considers that unlikely. The development of its strategy over the coming six months will be important for understanding long-term energy demand from the sector.

**Figure 33: Sora app downloads and revenue generation**



Source: BloombergNEF, Appfigures Intelligence via TechCrunch.

**OpenAI’s video generation app launches but downloads have stalled**

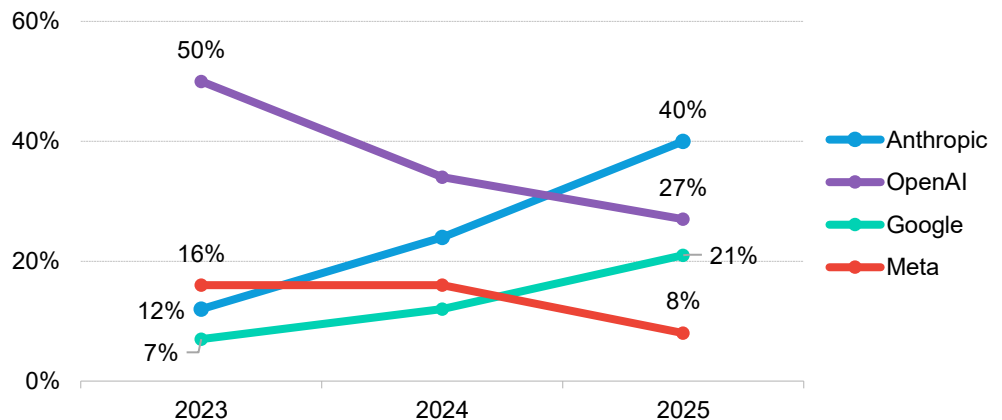
OpenAI released the app ‘Sora’, which gave users access to its energy-intensive video generation model. The app is free up to a video generation limit, and users can pay for use beyond this. It has proven relatively unpopular with users. It hit one million downloads in less than five days (faster than ChatGPT), but downloads have slid since a peak in November, and its revenue generation has been weak (Figure 33). This relatively poor uptake is a result of both competition from the likes of Google and Meta, and copyright challenges from Hollywood. For now, interest seems relatively low, but growth could translate into significant amounts of energy demand, due to the computational intensity of generating video.

**Anthropic consolidates enterprise leadership**

Anthropic, which has long positioned itself as the most safety-focused of the AI labs is being rewarded in enterprise sales. According to Menlo Ventures, it finished 2025 with 40% of enterprise large language model (LLM) API market share, and 54% market share for coding applications. This makes it by far the largest provider of enterprise AI services via API (Figure 34).

**An API** (application programming interface) is a way for different software systems to communicate and share data with each other, enabling services and applications to work together. In the AI context, it refers to a B2B revenue model where providers sell programmatic access to their models so other companies can integrate AI directly into their own products and workflows.

**Figure 34: Enterprise LLM API market share, leading providers**



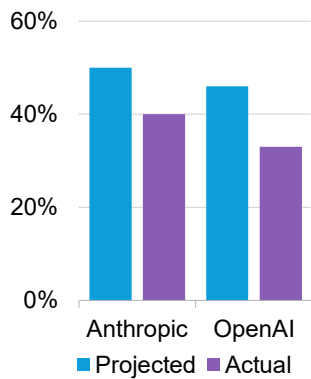
Source: BloombergNEF, Menlo Ventures.

**Google apps gain traction with launch of Gemini 3**

Google’s Gemini app has over 750 million monthly annual users (MAU), according to Google’s 4Q 2025 earnings call. This is up from 450 MAU in July 2025, a 67% increase. This is significantly higher than the growth that ChatGPT saw over the same period (although ChatGPT numbers started higher, and refer to weekly, not monthly, users).

Google is also seeing “significantly higher” engagement from users since the launch of Gemini 3 in December, and has seen revenue growth from AI-integrated products, notably search. The company has sold 8 million Gemini enterprise licenses since launching the product four months

**Figure 35: AI lab 2025 gross margins, projected versus actual**



Source: BloombergNEF, Anthropic and OpenAI via The Information. Note: Projected refers to internal margin projections from summer 2025.

ago, and has 325 million paying subscribers across consumer services including Google One, which can include premium AI access.

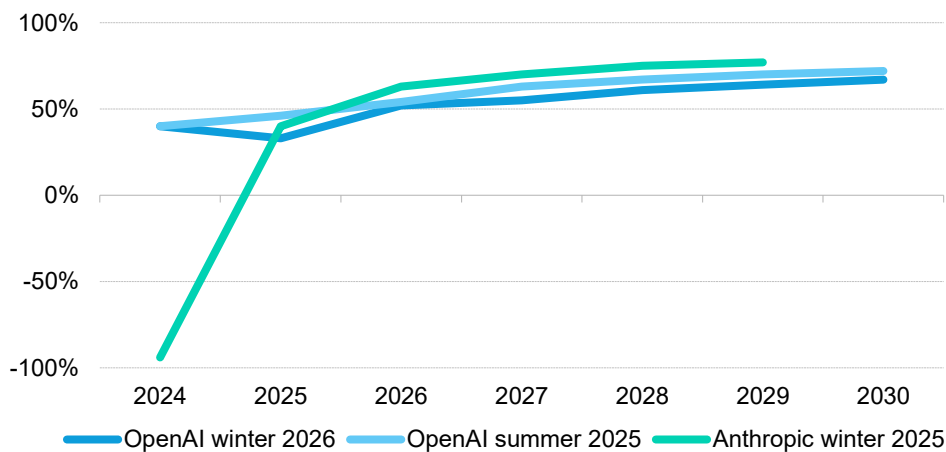
**5.4. AI gross margins are positive, but miss targets**

There is still uncertainty around the margins of providing AI services – for now, AI inference is much more energy intensive, and requires more expensive hardware, than traditional software products. This might result in lower gross margins, although there is wide scope for improvements in model and chip efficiency to improve these.

OpenAI's adjusted gross margin – revenue minus the cost of inference – fell to 33% in 2025, against a projection of 46%. The company told investors that its inference costs quadrupled in 2025. This was reportedly because it was forced to buy extra compute at the last minute due to rising demand for its products. If that is accurate, its margins are likely to improve medium term, if it is able to lock compute in further out.

Anthropic reportedly had an adjusted gross margin of 40% in 2025. Whilst higher than OpenAI's, likely due to its targeting of paid enterprise users, this was still 10% down on its projections. Like its rival, it was hit by higher-than-expected inference costs.

**Figure 36: OpenAI and Anthropic, expected gross margin**



Source: BloombergNEF, OpenAI and Anthropic via The Information. Note: OpenAI summer 2025 and winter 2026 refer to outlooks produced by OpenAI in summer 2025 and 2026 respectively. Anthropic winter 2025 refers to the "optimistic" outlook produced by Anthropic in winter 2025.

Margins at both Anthropic and OpenAI have fallen as their user bases increase. This is at least partially because cloud providers charge a premium for on-demand access to servers versus reserving them in advance. It is likely also partially due to growth in free-tier users, which are effectively subsidized by paying customers. OpenAI is introducing ads to drive revenue from these users, and Anthropic is not targeting the mass consumer market so has a much smaller free-tier offering (section 5.2).

## About us

### Contact details

#### Client enquiries:

- Bloomberg Terminal: press <Help> key twice
- Email: [support.bnef@bloomberg.net](mailto:support.bnef@bloomberg.net)

Lloyd Arnold	Analyst, Technology and Innovation	<a href="mailto:larnold46@bloomberg.net">larnold46@bloomberg.net</a>
Mark Daly	Head of Technology and Innovation	<a href="mailto:mdaly74@bloomberg.net">mdaly74@bloomberg.net</a>
Vandana Gombar	Senior Editor	<a href="mailto:vgombar@bloomberg.net">vgombar@bloomberg.net</a>

### Copyright

© Bloomberg Finance L.P. 2026. This publication is the copyright of Bloomberg Finance L.P. in connection with BloombergNEF. No portion of this document may be photocopied, reproduced, scanned into an electronic system or transmitted, forwarded or distributed in any way without prior consent of BloombergNEF.

### Disclaimer

The BloombergNEF ("BNEF"), service/information is derived from selected public sources. Bloomberg Finance L.P. and its affiliates, in providing the service/information, believe that the information it uses comes from reliable sources, but do not guarantee the accuracy or completeness of this information, which is subject to change without notice, and nothing in this document shall be construed as such a guarantee. The statements in this service/document reflect the current judgement of the authors of the relevant articles or features, and do not necessarily reflect the opinion of Bloomberg Finance L.P., Bloomberg L.P. or any of their affiliates ("Bloomberg"). Bloomberg disclaims any liability arising from use of this document, its contents and/or this service. Nothing herein shall constitute or be construed as an offering of financial instruments or as investment advice or recommendations by Bloomberg of an investment or other strategy (e.g., whether or not to "buy", "sell", or "hold" an investment). The information available through this service is not based on consideration of a subscriber's individual circumstances and should not be considered as information sufficient upon which to base an investment decision. You should determine on your own whether you agree with the content. This service should not be construed as tax or accounting advice or as a service designed to facilitate any subscriber's compliance with its tax, accounting or other legal obligations. Employees involved in this service may hold positions in the companies mentioned in the services/information.

The data included in these materials are for illustrative purposes only. The BLOOMBERG TERMINAL service and Bloomberg data products (the "Services") are owned and distributed by Bloomberg Finance L.P. ("BFLP") except (i) in Argentina, Australia and certain jurisdictions in the Pacific islands, Bermuda, China, India, Japan, Korea and New Zealand, where Bloomberg L.P. and its subsidiaries ("BLP") distribute these products, and (ii) in Singapore and the jurisdictions serviced by Bloomberg's Singapore office, where a subsidiary of BFLP distributes these products. BLP provides BFLP and its subsidiaries with global marketing and operational support and service. Certain features, functions, products and services are available only to sophisticated investors and only where permitted. BFLP, BLP and their affiliates do not guarantee the accuracy of prices or other information in the Services. Nothing in the Services shall constitute or be construed as an offering of financial instruments by BFLP, BLP or their affiliates, or as investment advice or recommendations by BFLP, BLP or their affiliates of an investment strategy or whether or not to "buy", "sell" or "hold" an investment. Information available via the Services should not be considered as research analysis or research recommendation or an opinion concerning any financial products or securities or entity or as information sufficient upon which to base an investment decision. Bloomberg makes no claims or representations, or provides any assurances, about the sustainability characteristics, profile or data points of any underlying issuers, products or services, and users should make their own determination on such issues. The following are trademarks and service marks of BFLP, a Delaware limited partnership, or its subsidiaries: BLOOMBERG, BLOOMBERG ANYWHERE, BLOOMBERG MARKETS, BLOOMBERG NEWS, BLOOMBERG PROFESSIONAL, BLOOMBERG TERMINAL and BLOOMBERG.COM. Absence of any trademark or service mark from this list does not waive Bloomberg's intellectual property rights in that name, mark or logo. All rights reserved. © 2026 Bloomberg.