Thematic Investing

Powering AI: Bring Your Own Power ('BYOP') as AI Party Continues - Data Center Insights v3.0

★ SIGNATURE

AI infrastructure remained a key theme across Tech, Industrial and Energy earnings. With speed-to-power key in the AI arms race, emphasis turns to bring-your-own-power (BYOP) solutions as the energy happy hour faces last call. We again compile recent Powering AI insights with 50+ company mentions.

The AI infrastructure party continues... Hyperscalers again raised AI infrastructure capex as data center demand continues to outpace supply. We highlight 1) hyperscaler capex trends, 2) strong public cloud demand, 3) elevated data center pre-leasing, 4) proliferation of AI training clusters, 5) surging AI inference demand, 6) US data center construction spend, and 7) hyperscaler electricity use.

...but speed-to-power fuels bring-your-own-power ('BYOP') solutions. Data center energy demand continues to outpace the ability to add grid power and infrastructure. This is leading to on-site power policies, including in Texas and Ireland, as speed-to-power fuels demand for distribution power across turbines, engines, fuel cells and energy storage, which we summarize in Figure 16. We compile companies exposed to the distributed power theme in Figure 17.

Company mentions include: Tech (**AAPL, AMZN, CRWV, GOOG, META, MSFT, ORCL**); Data centers (**CORZ, DLR, EQIX, IRM**); Industrials (**ABB, AAON, CARR, CAT, CLF, CMI, ENR, ETN, FIX, GEV, LR, MOD, NVT, OC, RR, SIE, SU, TSLA, TT, URI, WRT1V**); and Energy (**AEP, BE, BKR, CPK, GNRC, SEI, WMB**).

Thematic Investing

William Thompson
+1 212 526 8641
william.s.thompson@barclays.com
BCI, US

Hannah Greenberg
+1 212 526 2457
hannah.greenberg@barclays.com
BCI, US

Hiral Patel
+44 (0)20 3134 1618
hiral.patel@barclays.com
Barclays, UK

Laia Marin i Sola, CFA
+ 44 (0)20 3134 2531
laia.marin@barclays.com
Barclays, UK

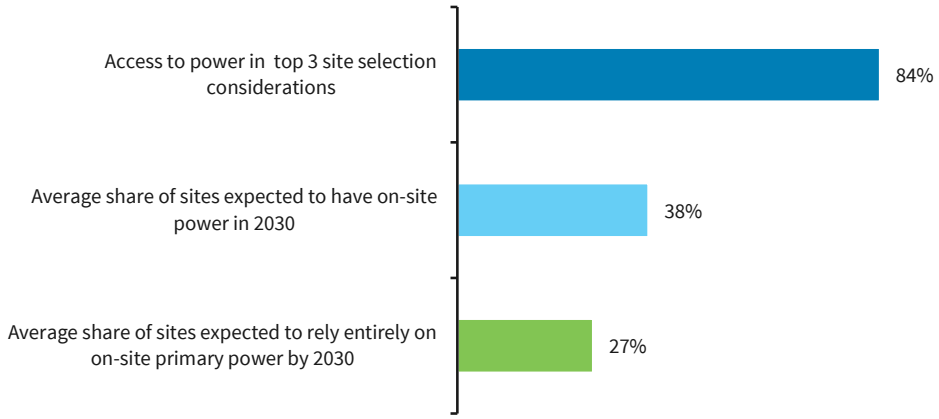
Katherine Ogundiya
+44 (0)20 3134 1391
katherine.a.ogundiya@barclays.com
Barclays, UK

Amy Lian
+44 (0)20 3134 0179
amy.lian@barclays.com
Barclays, UK

Siddharth Gupta
+91 22 6175 2540
siddharth.gupta1@barclays.com
Barclays, UK

Please see analyst certifications and important disclosures beginning on page 32.
Completed: 13-Aug-25, 02:59 GMT Released: 13-Aug-25, 04:10 GMT Restricted - External

FIGURE 1. Survey of data center leaders shows expectations for increasing on-site power needs



April 2025 survey of data center leaders
Source: Bloom Energy

CONTENTS

The AI Infrastructure Party Continues...

We outlined our rather bullish AI infrastructure views in Powering AI Chart Pack Summary (28 Slides) and Deep impact: The global AI arms race, which included why we thought AI was being underhyped (not overhyped), why the paradigm shift in AI scaling laws was leading to increasing AI computing and energy intensity despite hardware and algorithmic efficiencies (e.g. DeepSeek [private, not covered]), and the critical role of AI infrastructure and energy in the global AI arms race to general intelligence (AGI) and artificial superhuman intelligence (ASI).

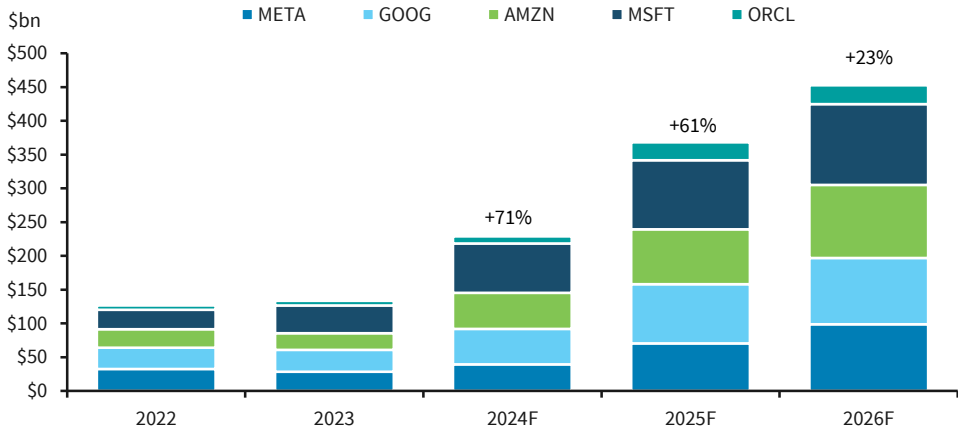
The post-DeepSeek bear thesis for AI infrastructure that emerged earlier this year suggested that ongoing AI efficiencies will continue to bend the AI compute and cost curves and commoditize AI, and this would lead to 1) hyperscalers pulling back on AI-directed capex amid low AI returns; 2) downward revisions to energy growth forecasts; and 3) an oversupply of data centers, as we detailed in Powering AI: Post-DeepSeek Bull/Bear Debate. This belief was also reinforced by speculation that MSFT was slowing and pausing leasing activity, which proved to be more noise than anything, as we note in MSFT data center plans. However, hyperscalers have raised (not cut) AI infrastructure-related capex and signaled out-year capex growth based on the following factors:

1. **Continued data center supply/demand tightness with power being the primary constraint.** Companies emphasized that strong cloud and AI demand continues to outpace supply, with Amazon (AMZN, covered by Ross Sandler) noting that power is the single-largest supply constraint.
2. **Optimisms on AI returns.** Companies are already seeing strong, measurable returns in terms of system efficiency, user engagement, and ad revenue and remain optimistic on generative AI opportunities amid compounding S-curves.
3. **Strength in conventional cloud demand.** While AI gets much of the attention, the hyperscalers continue to point to strength in their conventional cloud services.
4. **Proliferation of large AI training clusters.** The industry continues to pursue gigascale clusters of AI accelerators ('superclusters') to train frontier AI models. Anthropic (private, not covered) believes frontier AI training demand in America will reach 20-25 GW in total by 2028.
5. **Surging AI inference demand.** Hyperscalers have now started to disclose the rapidly rising growth in the number of tokens they're processing – a good proxy for AI inference computing/energy needs, in our view.

Hyperscaler Capex Continues to Climb

AI infrastructure spend continues to climb as cloud and AI demand outpaces supply. Meta (META, covered by Ross Sandler) raised the low end of its 2025 capex guidance range to $66-72 billion (~85% y/y vs. midpoint) after having raised its capex guide to $64-72 billion from $60-65 billion last quarter. Google (GOOG, covered by Ross Sandler) raised its 2025 capex guide to $85 billion (up ~60% y/y) from $75 billion. Microsoft (MSFT, covered by Raimo Lenschow) continues to guide for higher capex for fiscal 2026. Amazon has also implied 2025 capex would be up ~45% y/y. Collectively, hyperscaler consensus capex is now tracking over 20% growth in 2026 (vs. single digit growth expectations earlier this) following over 60% growth in 2025. The companies have emphasized that strong cloud and AI demand continues to outpace supply, with Amazon noting that power is the single largest supply constraint.

FIGURE 2. Hyperscaler historical capex and consensus capex forecasts



Amazon comprises AWS historical and Barclays AWS forecasts. MSFT (June 30 fiscal year) and ORCL (May 31 fiscal year) are shifted to better align with the calendar year.
Source: Company reports, Bloomberg, Barclays Research

- **Meta now guiding ~85% y/y 2025 capex growth and "significant" 2026 growth.** Meta raised the low end of its 2025 capex guidance range, which includes principal payments on finance leases, to $66-72 billion (up ~$30 billion, or ~85% y/y, vs. midpoint) after having raised the capex guide to $64-72 billion from $60-65 billion last quarter. Meta expects 2026 to be another year of "significant" capex growth as the company continues to aggressively pursue opportunities to bring additional AI capacity online to ensure enough capacity for both its core AI work (i.e., to support ad ranking and recommendation engines) and gen AI work (i.e., train frontier AI models and inference demand as it seeks to expand into consumer use-cases). Meta confirmed plans to "invest hundreds of billion of dollars in compute to build superintelligence" with "several multi-GP clusters," which we provide more detail on in section Proliferation of Large AI Training Clusters. Meta expects a greater mix of its 2025-2026 capex to be in shorter-lived assets (e.g, servers, networking) than in prior years and is exploring opportunities to attract significant external financing to support large-scale data center projects. Meta is already seeing strong, measurable returns for its core AI efforts with AI unlocking greater ad system efficiency (+5%/+3% ad conversions on Instagram/Facebook), meaningful ad revenue from campaigns using AI features, and more user engagement (+5%/+6% time spent on Instagram/Facebook). While Meta doesn't expect meaningful gen AI revenue this year or next year as gen AI is much earlier on the return curve, it remains very optimistic about monetization opportunities, particularly owing to its conviction that superintelligence is going to improve every aspect of what Meta does. Marc Zuckerberg outlined that developing superintelligence is now in sight, adding that the "fastest assumptions [to superintelligence] have been the ones that have most accurately predicted what would happen," and he expects this to continue. We previously detailed AI performance benchmarks and progress towards gen AI in AI Model Performance Gains & Insights.

> "We also expect that developing leading AI infrastructure will be a core advantage in developing the best AI models and product experiences. So we expect to ramp our investments significantly in 2026 to support that work."
>
> *— Susan Li, Meta CFO during 2Q25 conference call*

- **Microsoft continues to guide higher fiscal 2026 capex with demand outpacing supply.** MSFT has reiterated guidance for higher capex in fiscal 2026 (ends June 30th), but at more moderate growth with a greater mix of short-lived assets vs. ~50% y/y growth in fiscal 2025 to $85 billion, including lease financing. MSFT also expects higher capex growth in fiscal 1H26 vs. 2H26 due to timing of data center capacity deliveries, including large finance lease sites. The company stood up over 2 GW of data center capacity over the last 12 months and believes it continues to scale owned data center capacity faster than competitors, but emphasized that demand still exceeds supply; demand signals are strong across its cloud and AI offerings despite accelerated capex and the company "trying to pull leases in and get CPUs and GPUs in the system as quickly as we can." MSFT feels good about returns on its capex spend with a set of compounding S-curves that are continuously improving efficiency and performance for its customers, but "needs the teams to execute at their very best to get the capacity in place as quickly and effectively as they can." These comments are in stark contrast to market speculation earlier this year that MSFT was pulling back on data center leasing activity (Reuters: Microsoft pulls back from more data center leases in US and Europe, analysts say). In our view, MSFT's slowing or pausing some early-stage projects was arguably business-as-usual amid multi-year planning cycles owing to project-specific issues, the company having many long-dated lease options, needing to change regional capacity to ensure sufficient infrastructure is in the right places, and the company maintaining its evolving relationship with OpenAI (private, not covered). Note that AMZN has also emphasized the substantial logistics challenge of matching supply/demand for various services (conventional cloud products, AI training and AI inference) across many regions and data center zones globally. See Powering AI - Data Center World event: bullish sentiment, MSFT theories, power constraints.

> “We continue to lead the AI infrastructure wave and took share every quarter this year. We opened new DCs across six continents and now have over 400 data centers across 70 regions, more than any other cloud provider...We stood up more than 2 gigawatts of new capacity over the past 12 months alone. And we continue to scale our own data center capacity faster than any other competitor.”
>
> *— Satya Nadella, Microsoft CEO and Chairman during fiscal 4Q25 earnings call*

- **Google now guiding ~60% y/y 2025 capex growth and further increases in 2026.** GOOG recently raised its 2025 capex guide to $85 billion (up ~60% y/y) from $75 billion due to strong cloud demand and expects higher capex in 2026 amid tight supply/demand going into 2026 with strong customer demand across the company. The company noted that the vast majority of 2Q25 capex was invested in technical infrastructure with about two-thirds in servers and one-third in data centers and networking equipment. GOOG also emphasized its confidence in its capex returns given increasing customer satisfaction, very low customer churn rates, and greater investment efficiency in driving revenue growth.

- **Amazon implies ~45% capex growth in 2025.** AMZN expects its $31.4 billion 2Q25 capex to be a reasonable proxy for its quarterly capex rate for 2H25, implying ~$120 billion of 2025 capex, up ~45% y/y vs. ~$83 billion in 2024. Capex spend is being driven by AWS as the

company continues to invest in chips, data centers, and power to pursue the "unusually large" gen AI opportunity. We highlight AWS spend of $16.0 billion during 2Q25 (~52% total capex vs. 63% in 2024) was up 26% y/y, but down 21% q/q, which is likely timing related. AMZN noted that AWS demand continues to exceed supply and expects this to continue for several quarters, though is optimistic that the supply/demand imbalance will improve each quarter. The company noted strong demand across both its gen AI and non-gen AI services, and that power is the single largest supply constraint, among other constraints, including chips and other server components. Interestingly, AMZN suggested that next generation chips don't always yield what they expected.

> "On the supply constraints as it relates to AWS and what we see there, as I mentioned, we have more demand than we have capacity at this point. And I think that – and you see some of the constraints and they kind of exist in multiple places. The single biggest constraint is power. But you also see constraints off and on with chips and then some of the components that once you have the chips to actually make the servers, sometimes you have new generations of chips that are a little bit later than they're supposed to be and sometimes you get the chips and the yield you get in making servers isn't what you expect when you get to ramp."
>
> *— Andy Jassy, Amazon President and CEO during 2Q25 earnings call*

- **Oracle guided to higher fiscal 2026 capex.** Oracle (ORCL, covered by Raimo Lenschow), which doesn't report fiscal 1Q26 results until Sep. 9th, previously guided for fiscal 2026 capex (ends May 31st) to exceed $25 billion (up >18% y/y) to meet demand from backlog with demand being "almost insatiable." This follows ORCL having more than tripled capex in fiscal 2025 to ~$21 billion. ORCL emphasized that the vast majority of its capex is for revenue-generating equipment that is going into data centers and not for land or buildings as it relies on leased data center capacity. In a recent filing, Oracle disclosed a cloud deal "expected to contribute more than $30 billion in annual revenue starting in FY28." For comparison, Oracle collectively sold $24.5 billion worth of cloud services in its fiscal 2025 to all customers. OpenAI confirmed that it's the counterparty behind the deal and announced an agreement with ORCL to "develop 4.5 GW of additional Stargate data center capacity in the US," which with the 1.2 GW Stargate I site in Abilene, TX, will bring over 5 GW of Stargate data center capacity and will "run over 2 million chips." [1] [2]

- **Other notable mentions.** In our view, the initial intentions for the **Stargate** JV joint venture between OpenAI, Oracle, Softbank and Emirati state-owned investment firm MGX to spend $500 billion on AI infrastructure in the US over the next 4 years, including $100 billion

[1] Data Center Dynamics: OpenAI confirmed to be behind $30bn-a-year Oracle cloud deal, 4.5GW expected across multiple data center sites, accessed July 2025

[2] OpenAI: Stargate Advances with 4.5 GW partnership with Oracle, accessed July 2025

"immediately," were rather unrealistic (implies 45 GW of total capacity assuming $12 per MW). That said, Stargate is currently building a 1.2 GW AI-focused data center campus ('Stargate I') in Abilene, TX, west of Dallas-Fort Worth. **Crusoe** (private, not covered), which is the developer of Stargate I, indicated that the first phase of the project with over 200 MW began in June 2024 and is expected be to energized in 1H25, and the company began construction on the second phase with another roughly 1 GW across eight buildings.[3] OpenAI confirmed that parts of Stargate I are now up and running; **Oracle** began delivering the first Nvidia Blackwell racks last month (June 2025), enabling OpenAI to begin running early training and inference workloads.[4] Reports indicate that Crusoe is also pursuing a 1.8 GW data center campus in Wyoming in partnership with **Tallgrass** (private, not covered).[5] This follows headlines that Stargate was scaling back on a project likely in Ohio. [6] **Apple** (AAPL, covered by Tim Long) stated that it expects continued capex growth (Bloomberg reports fiscal 2026 capex consensus of ~$15.5 billion vs. ~$12.6 bn in fiscal 2025) while "significantly" increasing its investment in AI, which it described as "the most profound technologies of our lifetime," with investment in private cloud compute to preserve user privacy. Even with this slightly higher capex profile, AAPL still dramatically lags other large AI players when it comes to investment in infrastructure. At its recent analyst day, **Equinix** (EQIX, covered by Brendan Lynch) outlined plans to double its data center capacity by 2029 vs. ~2 GW today (1.5 GW in retail and 500 MW xScale capacity for hyperscaler tenants). EQIX expects $4-5 billion per year of spend, comprising $3-4 billion per year to expand capacity and support its portfolio (~80% for new capacity), plus $1 billion per year of strategic capital investment (land purchases plus xScale JV contributions). It has secured utility connection agreements for over 1 GW through 2026 and submitted for an additional 2 GW through 2029 with an active development pipeline for an additional 4 GW. **Digital Realty** (DLR, covered by Brendan Lynch) has about 2.9 GW of IT capacity in place with another ~5 GW of buildable IT capacity, including ~730 MW under construction, and expects $3.0-3.5 billion of 2025 development capex, up ~45% y/y vs. the midpoint. DLR reports that its first US Hyperscale Data Center Fund, which expands its pool of capital to support hyperscale data center capacity, is oversubscribed with over $3 billion of LP commitments to date and has the potential to support ~$10 billion of total data center investment. **Coreweave** (CRWV, covered by Raimo Lenschow) announced its intent to acquire Core Scientific (CORZ, not covered) in an all-stock transaction, which will provide CRWV with direct ownership of CORZ's 1.3 GW of gross power capacity (~840 MW tied to CRWV's active contracts across 5 sites and ~500 MW of crypto mining that can be converted to data centers or divested), plus 1 GW of potential expansion, including over 700 MW of incremental power available at current CRWV sites. CRWV emphasized that the deal provides "greater control over a critical and scarce power footprint" with more than 2 GW of gross power across 9 distinct sites. CRWV has also announced over a $6 billion commit to equip a new AI-dedicated data center in Lancaster, PA, starting with 100 MW with potential to expand to 300 MW.

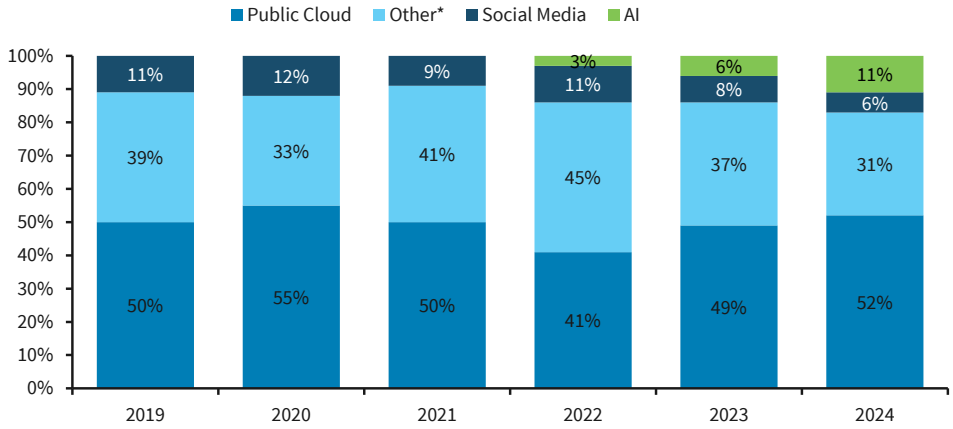
Public Cloud Underpins Data Center Demand

Hyperscalers continue to cite strength in conventional cloud demand. While AI gets much of the attention, the hyperscalers continue to point to strength in their conventional cloud services. Amazon noted that 85-90% of worldwide IT spend is still with on-premise enterprise data centers and it believes that will continue to flip to the cloud, accelerated by AI. The company added that AWS continues to help "organizations of all sizes accelerate their transition to the cloud," noting its new agreements with a number of major corporations. According to DC

[3] Crusoe: Crusoe Expands AI Data Center Campus in Abilene to 1.2 Gigawatts, accessed July 2025
[4] OpenAI: Stargate Advances with 4.5 GW partnership with Oracle, accessed July 2025
[5] Data Center Dynamics: Crusoe plans 1.8GW data center campus in Wyoming, accessed July 2025
[6] Reuters: SoftBank and OpenAI's Stargate aims building small data center by year-end, WSJ reports, accessed July 2025

Byte, public cloud still accounts for ~50% of global data center capacity, where as AI made up only 11% in 2024, up from 6% in 2023.[7]

FIGURE 3. Mix of global data center IT demand



*Other includes financial institutions, public sectors, MSP/SaaS, etc.
Source: DC Byte, Barclays Research

Strong Pre-leasing Supports Colocation Expansion

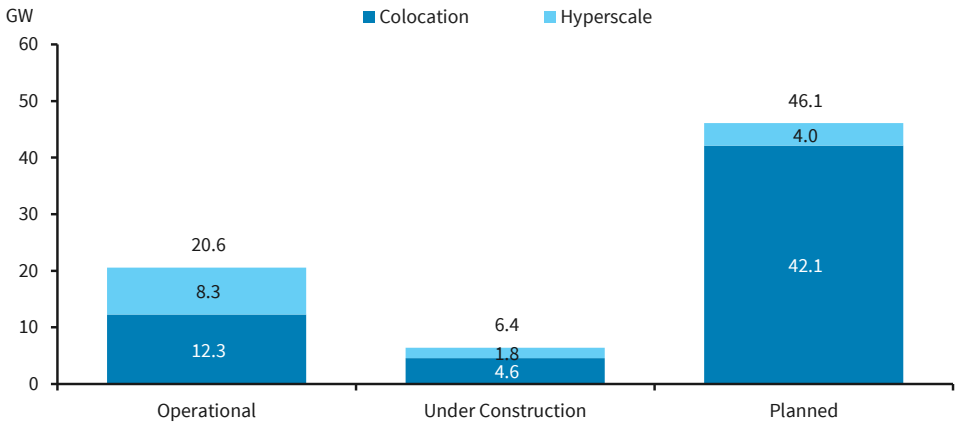
High pre-leasing of capacity under construction signals ongoing market tightness. According to Cushman & Wakefield, 76% of America data center capacity under construction is already leased (83% when including hyperscaler self-builds) and another 17% of planned capacity is already "spoken for."[8] This includes 79% pre-leasing in established data center markets (85% when including hyperscaler self-builds). Cushman & Wakefield pegs Americas data center vacancy at only ~5%, but notes that even in regions with higher vacancy rates, high vacancy doesn't tell the full story. High vacancy paired with strong pre-leasing can suggest the issue is not a lack of regional demand, but rather a lack of adequate capacity, since finding contiguous space in a large data hall, or building-sized availability, has become increasingly difficult. This has forced companies requiring 10+ MW of contiguous capacity to target the construction pipeline.

> "As this rapid expansion continues, questions of oversupply arising from potentially slowing demand have emerged. Our trend analysis on demand—by account of volume and increasingly advanced pre-leases—indicates otherwise. On all accounts, demand has undoubtedly exceeded supply, and can be expected to continue on this trajectory with added demand drivers alongside increasing supply bottlenecks."
>
> *— DC Byte: Global Data Centre Index 2025*

[7] DC Byte: 2025 Global Data Centre Index, accessed August 2025
[8] Cushman & Wakefield: 2025 Global Data Center Market Comparison, accessed July 2025

During 2Q25, 98% of Digital Realty's 96 MW of new capacity delivered was pre-leased. Equinix indicated that over 85% of its xScale capacity for hyperscaler tenants was leased or pre-leased. Iron Mountain (IRM, covered by Brendan Lynch) indicated that 60% of its 202 MW of under construction capacity is pre-leased.

FIGURE 4. Americas data center IT capacity by status and type



Excludes enterprise data centers.
Source: Cushman & Wakefield, Barclays Research

According to Cushman & Wakefield, North America operational IT capacity is split about ~60%/40% colocation/hyperscale, under construction capacity is split ~70%/30% colocation/hyperscale, and planned capacity is split ~90%/10% colocation/hyperscale.[9] DC Byte believes the surge in pre-leasing of committed capacity indicates that demand continues to outpace supply, particularly with reservations now well in advance of construction.[10]

> "I just remind you that the vast majority of our data center growth capital is going to support the construction of pre-leased assets. If you look at the next few quarters, we have assets in Arizona that will be completing construction, 100% leased. If you look at what we've got in London, same situation. If you look at what we've got in Northern Virginia, same situation, all those assets are 100% pre-leased and the clients are looking forward to having the ability to turn on. So no change in terms of the way we're thinking about capital deployment, it's vastly going to support pre-leased construction."
>
> *— Barry Hytinen, Iron Mountain Executive VP and CFO during 2Q25 earnings call*

[9] Cushman & Wakefield: Americas Data Center H2 2024 Update, July 2025
[10] DC Byte: 2025 Global Data Centre Index, accessed August 2025

Proliferation of Large AI Training Clusters

Bigger is still better for frontier models. Despite expected implications from DeepSeek, there continues to be a trend towards large data centers to support massive clusters of AI accelerators for training frontier AI models. **Anthropic** projects that "2GW and 5GW data centers will be needed to develop single advanced AI models for Anthropic in 2027 and 2028, respectively," and expects total frontier AI training demand in America to reach 20-25 GW in total by 2028 amid several other US companies developing frontier AI models.[11] Combined with interference demand, the company believes the US needs at least 50 GW of AI capacity by 2028 to maintain global AI leadership.[12] It plans to train its next-generation Claude models on “hundreds of thousands” of AWS Trainium2 chips. **Meta** noted that a big driver of its 2026 capex will be for scaling "GenAI capacity as we build out training capacity" with plans to "invest hundreds of billion of dollars in compute to build superintelligence" with "several multi-GW clusters." This includes Meta's 1-GW Prometheus project in Ohio that is set to be operational in 2026 and its multi-GW Hyperion project in Louisiana (1.5 GW operational as early as 2026-2027, but designed to potentially scale to 5 GW over several years). Meta has also noted "multiple more Titan clusters" in development. Elon Musk's **xAI** (private, not covered) startup trained Grok 3 at its 'Colossus' 1 data center in Memphis, Tennessee. Colossus 1 has ~230k Nvidia GPUs (mostly H200s, but also 30k GB200s) in a single "supercluster" dedicated to training, while xAI's inference needs are provided by its cloud providers. Musk has also revealed plans for Colossus 2, which he described as the "the first Gigawatt AI training cluster." Colossus 2 will be near Memphis, TN, and will initial deploy ~550k GPUs (GB200s and GB300s) to train future Grok models.

> “Building AI in the United States is a national security and economic imperative. As AI systems grow more capable, the energy and computational requirements to train and deploy frontier AI are surging. Recent estimates by outside experts and our own research at Anthropic suggest that the U.S. AI sector is on track to require at least 50 gigawatts of electric capacity by 2028, much of which will be needed to train the world’s most capable models.”
>
> *— Anthropic: Build AI in America, July 2025*

The original AI scaling law, now referred to as pre-training scaling, indicated that AI performance gains would require continually increasing the size of models and the amount of data they are trained on. As both model size and training data increase, there is an exponential increase in training computing requirements and costs, holding all else constant in terms of model and hardware architecture. Some industry executives indicated that the cost to train a single large language model (LLM) could approach $100 billion within the next five years, requiring hundreds of thousands, if not millions, of the most advanced AI accelerator chips and gigawatts of power.

[11] Anthropic: Build AI in America, accessed July 2025

[12] Anthropic: Build AI in America, accessed July 2025

> “If you sort of even subscribe to this point of view that intelligence is basically log of compute, that means compute is going to grow and you've got to use it as efficiently as possible to just keep creating intelligence.”
>
> *— Satya Nadella, Microsoft CEO and Chairman during fiscal 4Q25 conference cal*

DeepSeek, however, put this all into question as the emergence of post-training and time-test scaling opened the door for smaller, cheaper, and better performing AI models. While GPT-4 (~1.8 trillion parameters) was believed to usher an era of trillion parameter models, many recent state-of-the-art foundation models, including DeepSeek V3 (617 billion parameters), have been smaller than GPT-4 despite significant improvements in model performance. This reflects a shift to over-training smaller models with more tokens; advancing the performance of smaller, more inference-efficient ‘student’ models by using high-quality synthetic data distilled from larger 'teacher' models; and allocating additional compute during the inference phase via time-test reasoning to improve model performance without increasing model size. That all said, as we outlined in Powering AI: Post-DeepSeek Bull/Bear Debate, the industry would likely continue to pursue sizeable LLMs trained on large training clusters to achieve artificial generative intelligence (AGI) and eventually artificial superintelligence (ASI) owing to the following:

- **Time-test reasoning has it limitations.** While time-test reasoning can efficiently improve model performance for complex, reasoning-heavy tasks, such as math, science, and coding, it is less effective for non-reasoning tasks, particularly those that don't have verifiable rewards with exact correct answer, such as chatbots, creative writing, general Q&A, editing, and summarization.

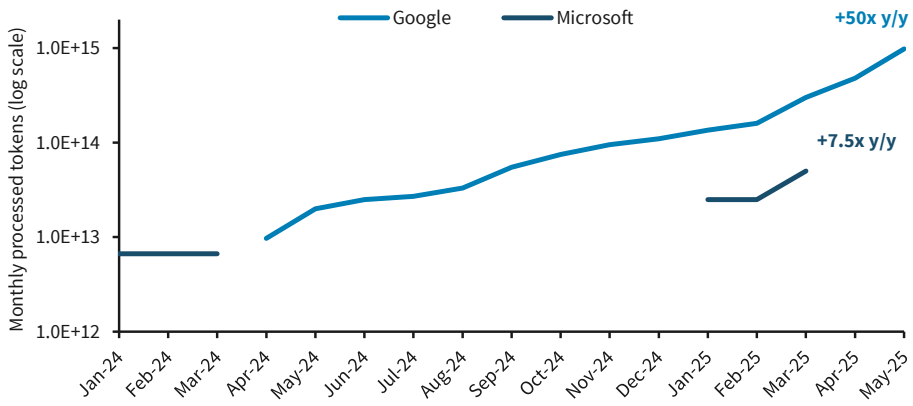
- **Large teacher models are still needed to develop smaller student models.** Hyperscalers and frontier AI labs have been training much larger models that won't be released to the public due to high inference latency and costs, but are used as 'teacher' models to advance the performance of smaller, more inference-efficient ‘student’ models via distillation. For example, Meta's Llama 4 'Behemoth' model has nearly 5x more parameters trained on about 2x more tokens than Llama 3, implying nearly 10x more computing need before any model and hardware architecture advancements. Meta recently noted a trend toward models "so big that they're just not practical for a lot of other people to use," but the reason to pursue big models like Behemoth is that these large frontier models ('teacher models') can be distilled into smaller, more efficient models ('student models'). Llama 4 Maverick (400 billion parameters) and Llama 4 Scout (109 billion parameters) were distilled from Llama 4 Behemoth (4 trillion parameters) in order "to get the 90%, 95% of the intelligence of the large model in a form factor that is much lower latency and much more efficient."

Surging AI Inference Demand

New hyperscaler disclosure confirms surging inference demand. Some of the investor pushback to our bullish Powering AI views has been the lack of tangible evidence of AI inference demand. While we can point to the size of models (parameters) and training data (tokens) to quantify AI training needs, we haven't had the same level of disclosure for AI inference. Hyperscalers have now started to disclose the number of tokens they're processing—a proxy for AI inference computing/energy needs. In May, **Google** indicated that it was processing 480 trillion monthly tokens (+50x y/y) and recently indicated that this has doubled to 980 trillion monthly tokens. **Microsoft** indicated it processed 100+ trillion tokens in fiscal 3Q25 (ended

March 31st), a 5x increase y/y, with 50 trillion tokens in March alone, implying ~8x y/y increase vs. the F3Q24 monthly average. It also disclosed that it processed over 500 trillion tokens in fiscal 2025, up over 7x, by Foundry APIs, which it believes "is a good indicator of true platform diffusion of AI with application becoming embedded in our daily work and life, while its family of Copilot apps has surpassed 100 million monthly active users across commercial and consumer. See Powering AI: Evidence of Strong AI Inference Demand Growth.

FIGURE 5. Monthly processed tokens for Google and Microsoft



The monthly MSFT numbers reflect the reported quarterly average for F3Q24 and implied average for Jan/Feb 2025 (100+ T tokens for F3Q25 less 50 T tokens in March divided by 2 months).
Source: Company reports, Barclays Research

Anthropic has emphasized the distinct energy and AI infrastructure challenges between AI training and AI inference.[13] While frontier models are trained primarily within single data centers that contain enormous clusters of AI accelerators requiring "massive volumes of electricity delivered specifically to those places," Anthropic notes that inference does not need such highly concentrated electricity delivery. The challenge with inference, however, is "building infrastructure to handle lighter computational workloads located physically near the customers seeking to use AI," particularly for AI workloads that require low latency (i.e. the lag time between between AI users and the data centers serving them). Amazon noted during its 2Q25 earnings call that "today so much of the cost is in training because customers are really training their models and trying to figure out to get their applications into production. But at scale, 80% to 90% of the cost will be in inference, because you only train periodically, but you're spitting out predictions and inferences all the time. And so what they're going to care a lot about the compute and the hardware they're using."

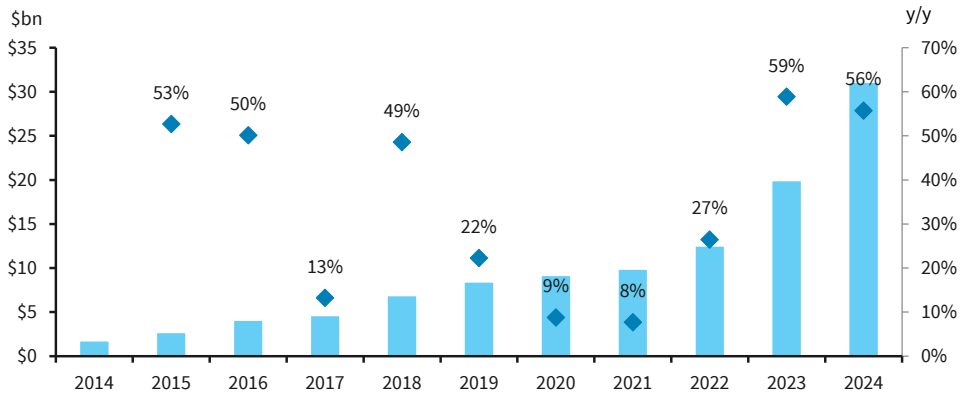
US Data Center Construction Spending

Data center construction spending remains quite strong. The US Census Bureau has started breaking out private data center construction spending (dating back to January 2014) from its office construction spending figures.[14] US data center construction spending grew 59% y/y in 2023 to $20 billion and another 56% y/y in 2024 to $31 billion. Construction spending includes the cost of labor and materials, architectural and engineering work, overhead, interest and taxes paid during construction, and contractor's profits, but excludes land acquisition and data center racks and servers.

[13] Anthropic: Build AI in America, accessed July 2025

[14] US Consensu Bureau: Construction Spending, accessed July 2025

FIGURE 6. Annual value of US private data center construction put in place



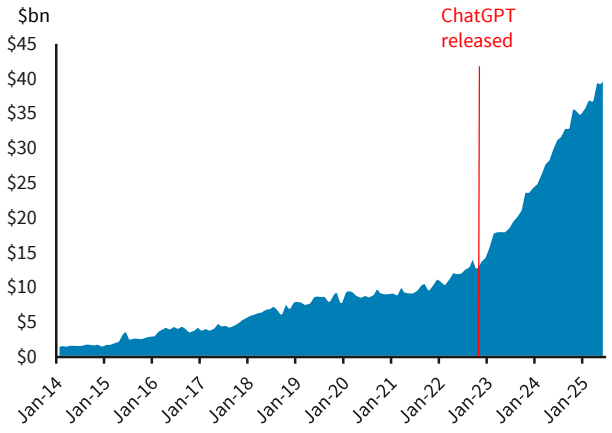
Includes value of construction installed but excludes land acquisition, racks and servers.
Source: US Census Bureau, Barclays Research

For June 2025, the seasonally adjusted annual run-rate in US data center construction spending exceeded $40 billion, up 28% y/y. **United Rentals** (URI, covered by Adam Seiden) talked positively about data centers with new projects kicking off. **Comfort Systems** (FIX, not covered), which provides mechanical and electrical contracting services, reported a 66% y/y revenue increase from its Technology customers during 2Q25, adding that the "increase in demand has been especially strong in the technology sector, particularly for data centers." HVAC suppliers echoed this strength in data center demand. **Carrier** (CARR, covered by Julian Mitchell) is on track to double its data center revenues to $1 billion this year and is continuing to build its backlog for next year and beyond. CARR's Americas data center orders increased over 300% y/y during 2Q25 and the company disclosed a $45 million data center order from a Middle East customer. CARR emphasized that hyperscaler customers are looking for supplier capacity availability on a global scale. **Johnson Controls** (JCI, covered by Julian Mitchell) noted "very high" data center demand and is focused on improving chillers lead times to serve the data center market. **Trane** (TT, covered by Julian Mitchell) emphasized how fast the data center market is moving from a technology standpoint. **Modine Manufacturing** (MOD, not covered) reported 15% y/y growth in North America data center revenue driven by both hyperscaler and colocation customers, announced a $100 million North America capacity expansion to support data center demand, and raised its fiscal 2026 outlook for data center revenue growth to 45%. **AAON** (AAON, not covered) reported a 127% y/y increase in BASX-branded data center sales during 2Q25 (+269% YTD), adding that the "data center market continues to demonstrate exceptional strength." **Owens Corning** (OC, covered by Matthew Bouley) indicated strong insulation demand for data center construction to help control temperature and moisture, which is an extensive HVAC requirement.

Electrical equipment suppliers have also indicated strong global data center demand. **Eaton** (ETN, covered by Julian Mitchell) reported ~55% and 50% y/y growth in data center orders and revenues, respectively, during 2Q25 with particular strength from multi-tenant data center customers. ETN cited the biggest pain points for its data center customers were power availability, the speed to build a data center and efforts to increase return on capital. **nVent** (NVT, covered by Julian Mitchell) noted accelerating data center growth to support the AI buildout with strength across both data center white and gray space. It identified a trend towards modular data centers using large outdoor enclosures to house all the IT hardware, including cooling, plus a trend towards trying to expand the white space and maximize the IT footprint by moving the gray space outside, requiring outdoor enclosures. **Vertiv** (VRT, covered by Julian Mitchell) echoed that there is strength across data center gray and white space,

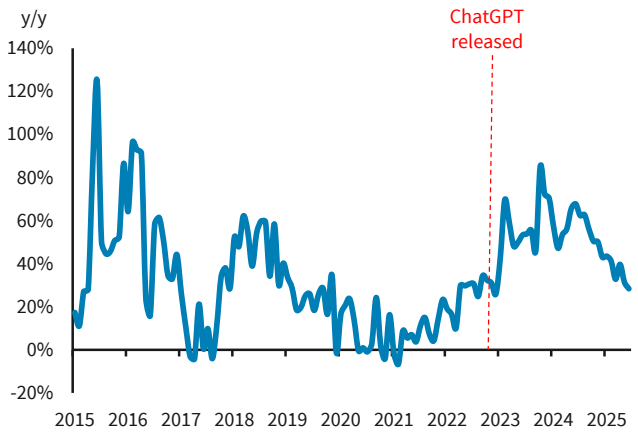
stating during its 2Q24 earnings call that, "Gray space and white space no longer are separate spaces....With increase in rack density, the physical integration and interoperability between these spaces has become absolutely evident and critical." **ABB** (ABBN.SW, covered by George Featherstone) reported double digit y/y growth in data center orders during 2Q25 with strong demand across all clients and geographies. The company expects order strength to continue, but noted more normalized pricing with a better supply/demand balance amid supply additions. **Legrand** (LR.FP, covered by George Featherstone) reported 9% y/y organic revenue growth in 1H25, driven by data centers, and now expects double-digit average annual organic growth in its data center market between 2025 and 2030. **Schneider Electric** (SU.FP, covered by George Featherstone) reported strong double-digit growth in Energy Management with contribution across end-markets and segments, most notably in Data Center, which saw "particular traction for the Group's modular offering." SU still sees a double-digit growth opportunity for data centers with a strong pipeline of deals. **Siemens** (SIE.GY, covered by Vlad Sergievskii) reported 16% y/y fiscal 3Q25 revenue growth for its Smart Infrastructure's electrification business driven by data centers. While SIE's electrical produce order fell 2% due to fewer major orders for data centers in the US compared to the prior year, the company emphasized "activity in the data center vertical remained sound, with a book-to-bill clearly above 1x".

FIGURE 7. Monthly annual run-rate of value of US private data center construction put in place (seasonally adjusted)



Includes value of construction installed but excludes land acquisition, racks and servers.
Source: US Census Bureau, Barclays Research

FIGURE 8. Monthly growth of annual run-rate of value of US private data center construction put in place (seasonally adjusted)



Includes value of construction installed but excludes land acquisition, racks and servers.
Source: US Census Bureau, Barclays Research

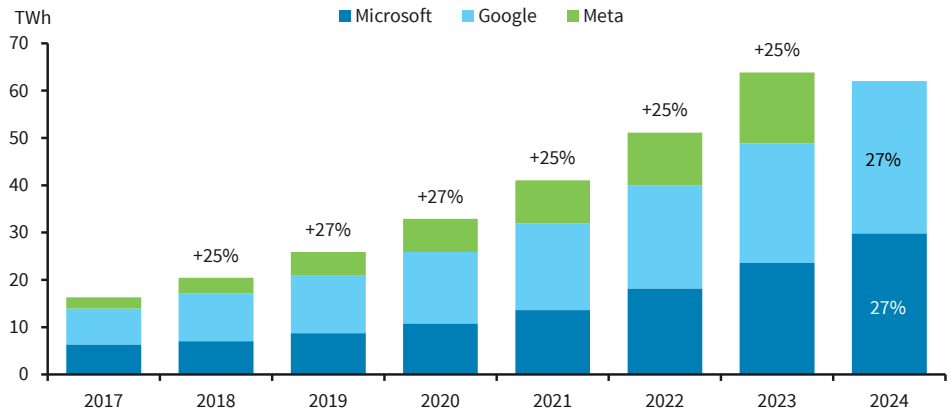
We struggle, however, to reconcile the US Census Bureau data center construction spending data with US/North America data center capacity addition estimates. According to Cushman & Wakefield, data center construction costs now average around $12 million per MW, ranging from $9-15 million per MW depending on the region.[15] We heard similar average figures at a recent data center event (see Powering AI - Data Center World event: bullish sentiment, MSFT theories, power constraints), while IT equipment (i.e., racks and servers) adds another $20-30 million per MW. Based on $12 million per MW construction costs, the ~$40 billion annual run-rate in US data center construction spending implies only 3.3 GW of capacity additions, which is well below third-party estimates (Cushman & Wakefiled, DC Byte, BloombergNEF, IEA) indicating closer to 5 GW added in 2024 and 6.4-11 GW currently under construction.

[15] Cushman & Wakefield: Data Center Development Cost Guide 2025, accessed July 2025

Hyperscaler Electricity Use

Hyperscalers trending towards seventh consecutive year of 25%+ electricity use. Google and Microsoft both reported a 27% y/y increase in global electricity use during 2024. This suggests hyperscalers are on track for their seventh consecutive year of 25%+ y/y electricity demand, even before surging AI inference demand.

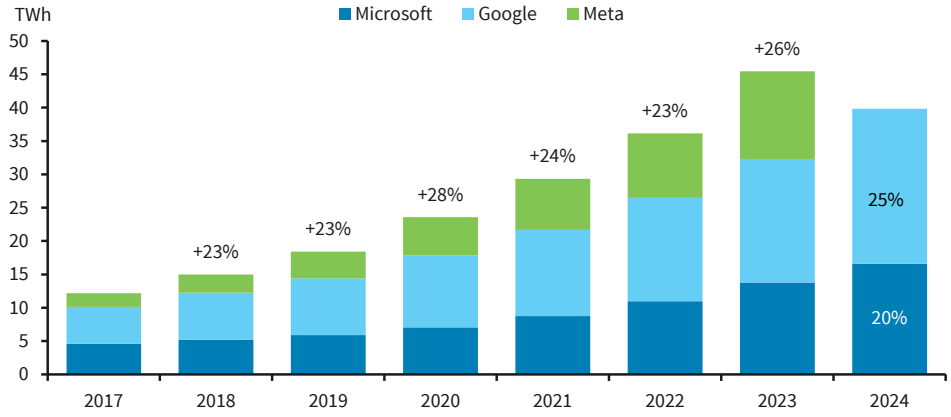
FIGURE 9. Hyperscaler global electricity use



Source: Company reports, Barclays Research

North America/US electricity use increased 20% y/y and 25% y/y for Microsoft and for Google.

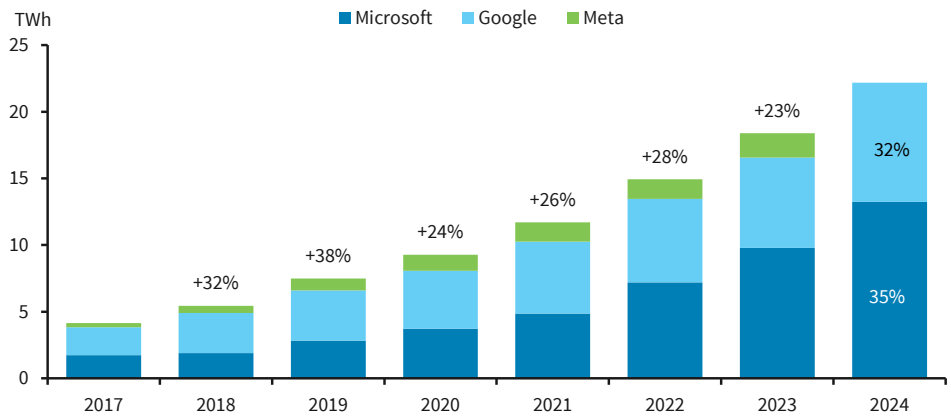
FIGURE 10. Hyperscaler North America/US electricity use



Source: Company reports, Barclays Research

International electricity use increased 35% y/y and 32% y/y for Microsoft and for Google. See Powering AI: MSFT electricity use +27% in FY2024, transitioning to liquid cooling and Powering AI: Google Reports Surging 2024 Electricity & Water Use.

FIGURE 11. **Hyperscaler International electricity use**



Source: Company reports, Barclays Research

...But the Energy Happy Hour Faces Last Call

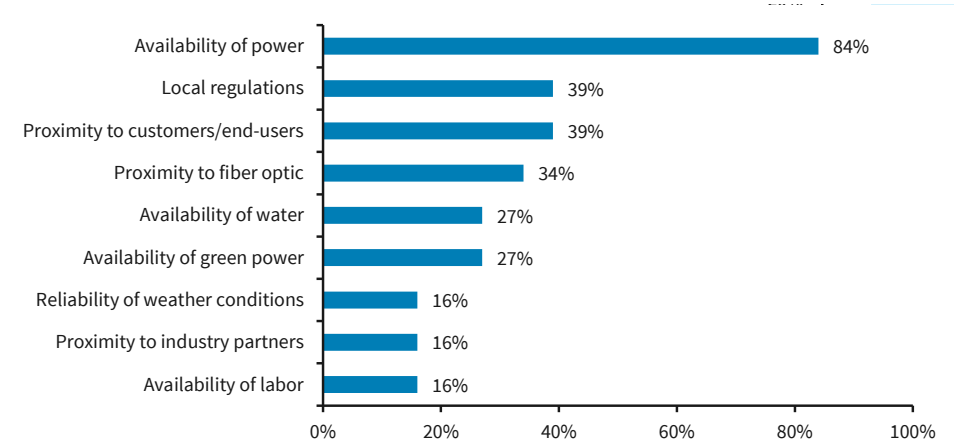
Pivoting to Bring-your-own-power ('BYOP') Solutions

Amid grid power constraints, data centers are turning to distributed power. The scale and speed of new data center energy demand is outpacing the ability to develop new generation and grid infrastructure, as we describe in Powering AI Chart Pack Summary (28 Slides) and Electric Grid (Un)lock: Electrification 'Pick & Shovel' Players. Power is often cited as the single largest supply constraint for data centers, which we detail in Powering AI: The Power & Land Arms Race. An example of this is xAI's use of mobile gas turbines to help power its Colossus 1 data center in Memphis, TN, and its purchase of an existing overseas power plant that it seeks to import to the US to help power Colossus 2, as we discuss below.

In a April 2025 survey conducted by Bloom Energy (BE, not covered), 84% of data center leaders ranked access to power in their top three considerations in data center site selection, far exceeding local regulations (39%), proximity to customers/end-users (39%), fiber optics (34%), availability to water (27%) and green power (27%).[16]

[16] Bloom Energy: 2025 Data Center Power Report: Mid-Year Pulse, accessed July 2025

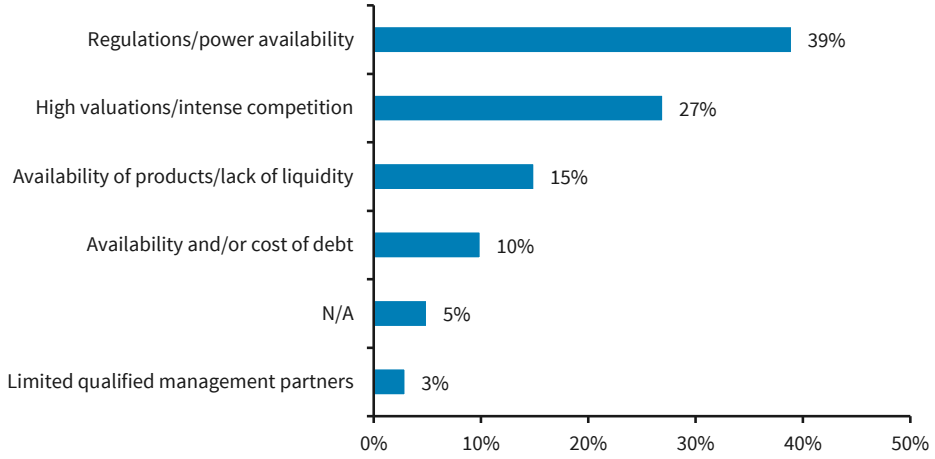
FIGURE 12. Bloom Energy survey of data center leaders: Power rated in top 3 considerations when selecting a data center location



Source: Bloom Energy

In a recent survey of major data center investors conducted by CBRE, 39% of investors listed power constraints and permitting delays as the primary risk to their current data center investment strategy, replacing cost and availability of debt, which was cited by only 10% of respondents this year vs. 59% last year. [17] Furthermore, only 73% of investors indicated that ESG considerations are at least somewhat important (includes only 22% believing it is highly important), which down from 93% in 2024.

FIGURE 13. CBRE survey of data center investors: Power availability and permitting delays were rated as the top risk to their current data center investment strategy



Source: CBRE

The US Department of Energy (DOE) has now joined the North America Reliably Commission (NERC) in raising concerns about future grid reliability, stating that "Absent decisive intervention, the Nation's power grid will be unable to meet projected demand for manufacturing, re-industrialization, and data centers driving artificial intelligence (AI) innovation."[18]

[17] CBRE: 2025 Global Data Center Investor Intentions Survey, accessed August 2025

[18] US Deptarment of Energy: Resource Adequacy Report - Evaluating the Reliability and Security of the United States Electric Grid, accessed July 2025

> "A failure to power the data centers needed to win the AI arms race or to build the grid infrastructure that ensures our energy independence could result in adversary nations shaping digital norms and controlling digital infrastructure, thereby jeopardizing U.S. economic and national security."

— US Dept. of Energy: Resource Adequacy Report - Evaluating the Reliability & Security of the United States Electric Grid, July 2025

In its "Winning the Race - America's AI Action Plan", the White House argues that the US grid needs to be upgraded to support data centers and other energy-intensive industries of the future and calls for 1) stabilizing the current grid 2) optimizing existing grid resources; 3) prioritizing the interconnection of reliable, dispatchable power sources and embracing new energy generation sources (e.g., enhanced geothermal, nuclear fission, nuclear fusion); and 4) creating a strategic blueprint to navigate the complex energy landscape of the 21st century. According to PJM, the fact that capacity pricing hit the $329/MW-day price cap during its recent auction underscores the region's "tightening supply-demand balance" amid data center expansion and will result in 1.5-5% y/y increases in some customers' bills.[19] Anthropic also recently detailed the importance in AI infrastructure in its Build AI in America report, echoing many points that OpenAI made it is AI in America Economic Blueprint report.

> "Historically, data centers have used onsite power mostly for backup purposes, but we are seeing a shift toward onsite power generation as a primary source of power."

— Bloom Energy: 2025 Data Center Power Report

We previously detailed the data center arms race for large tracts of land with untapped power in Powering AI: The Power & Land Arms Race. **Cleveland Cliffs** (CLF, not covered) indicated that data centers developers have expressed interest in its recently idled facilities that are uniquely positioned geographically with access to power and water infrastructure to support data centers. However, indications are that such sites are increasingly few and far between, particularly amid competition with new industrial manufacturing plants and other large users of power. According to Anthropic, "Finding sites suitable for gigawatt-scale AI infrastructure is a significant challenge."

> "The power grid is the lifeblood of the modern economy and a cornerstone of national security, but it is facing a confluence of challenges that demand strategic foresight and decisive action. Escalating demand driven by electrification and

[19] PJM: PJM Auction Procures 134,311 MW of Generation Resources; Supply Responds to Price Signal, accessed July 2025

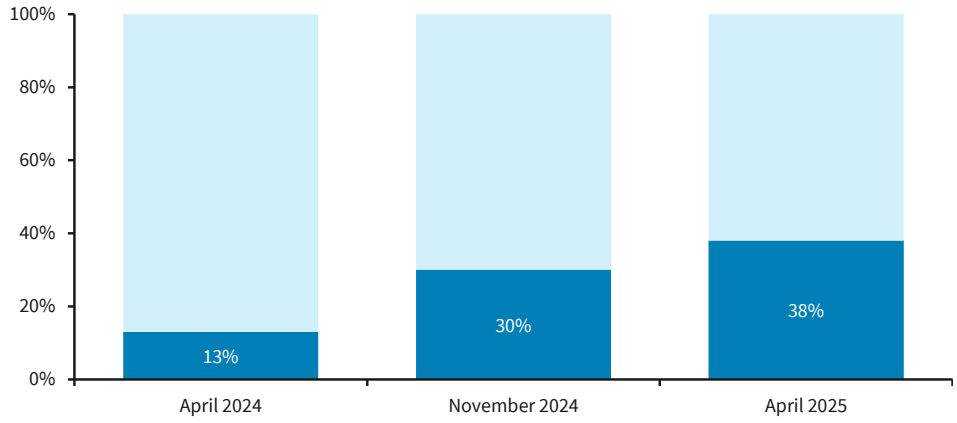
> the technological advancements of AI are increasing pressures on the grid. The United States must develop a comprehensive strategy to enhance and expand the power grid designed not just to weather these challenges, but to ensure the grid's continued strength and capacity for future growth."
>
> *— White House: Winning the Race - America's AI Action Plan, July 2025*

With speed-to-power critical in the AI arms race, there appears to be a push towards what can best be described as bring-your-own-power ('BYOP') policies and solutions. Bloom Energy's survey indicates a 1- to 2-year gap between when data center developers expect power and when utilities can realistically deliver it in many key US data center markets.

> "Data center developers are more optimistic than utility providers about how quickly grid power can be delivered. In key markets, this misalignment in expectations highlights a typical 1-to-2-year gap between when developers expect power and when utilities can realistically deliver it."
>
> *— Bloom Energy: 2025 Data Center Power Report, Mid-Year Pulse*

As a result, Bloom Energy found that data center leaders expect ~38% of all data center sites will use some on-site power as a primary energy source to supplement to grid power by 2030, up from 13% only a year ago. [20]

FIGURE 14. Bloom Energy survey of data center leaders: Average share of data center sites expected to have on-site power generation in 2030

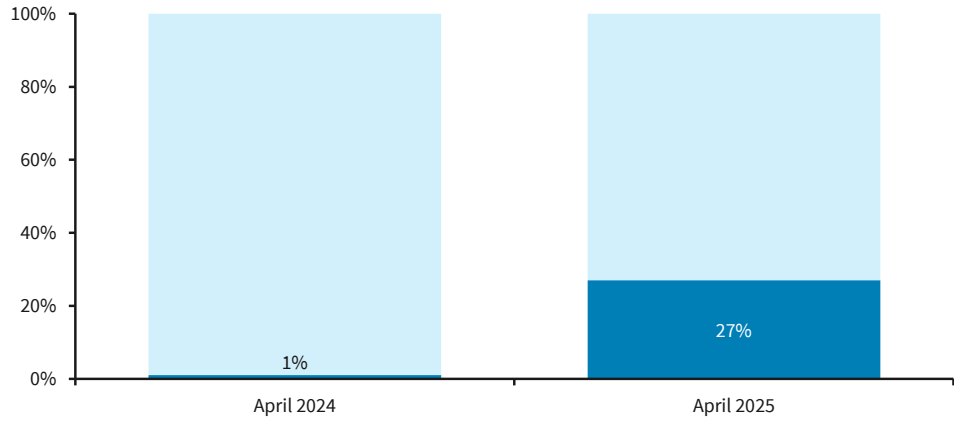


Source: Bloom Energy

[20] Bloom Energy: 2025 Data Center Power Report, accessed July 2025

Furthermore, data center leaders expect 27% of data center projects to rely entirely on on-site generation for primary power by 2030, up from just 1% a year ago.

FIGURE 15. Bloom Energy survey of data center leaders: Average share of data center sites expected to rely entirely on on-site generation for primary power by 2030



Source: Bloom Energy

At a recent Senate hearing on "Identify Challenges to Meeting Increased Electricity Demand", Jeff Tench, Executive VP for North America & Asia Pacific at **Vantage Data Centers** (private, not covered) argued that Vantage would prefer to source power from the grid, but is being forced to explore alternatives due to the following:

1. **Long interconnection timelines.** "Queue studies can take several years to complete. Restudies and limited transparency add further delay. This uncertainty makes it difficult for developers to plan and for customers to commit."
2. **Lagging transmission development.** "Many regions are not planning for the scale of growth driven by the digital economy. Without anticipatory planning, grid congestion worsens, and costs rise."
3. **Permitting is fragmented and sequential.** "Transmission and generation projects must navigate multiple federal, state, and local reviews, often with overlapping or duplicative requirements. This slows down projects that are otherwise ready to proceed."

As a result, Vantage has to consider on-site generation, co-locating near existing power plants, planning developments in phases, and coordinating with utilities and state partners to locate scarce near-term capacity. The company has built and now operates a large off-grid gas plant in Virginia as it would have taken 4-7 years, or longer, to obtain enough grid power.

> "The greatest challenge we collectively face to our leadership on artificial intelligence today is timely access to reliable electric power...In market after market across the U.S., our development teams all report the same issue: we cannot get the amount of electricity we need in the timeframe we need it to build our data centers."
>
> *— Jeff Tench, Vantage Data Centers, Executive VP for North America and Asia Pacific, July 2025*

Mr. Trench, however, notes that these alternatives are unsustainable for national infrastructure growth and diverts capital away from shared infrastructure that could strengthen the overall grid.

> "We may face a future scenario where large portions of load operate off-grid. That is a missed opportunity for national resilience, and it reflects a system out of sync with the scale of modern demand. When data centers are integrated into the grid with the right planning and policy framework, the load can be a source of stability. Large customers invest in shared infrastructure that improves grid resiliency and supports broader system needs without impacting rates on other customers."

— Jeff Tench, Vantage Data Centers, Executive Vice President for North America and Asia Pacific, July 2025

Additionally, policy makers are increasingly focused on trying to find ways to ensure data centers contribute to grid adequacy. Texas recently signed into law Senate Bill 6 ('SB 6'), which establishes standards for interconnecting large loads (75+ MW). SB 6, which is often referred to as the "Kill Switch Bill", requires large loads to have on-site backup power capacity meeting 50% of their power demand, and gives Texas grid operator ERCOT authority to mandate large loads to switch to their backup power, curtailment power use, or completely cutoff power remotely, with 24-hour notification, during periods of grid strain.

> "For example, Senate Bill 6 in Texas, which was recently signed into law. The new law requires enough co-located generation for large demand loads so that they can be self-sufficient off the grid. We have observed that this regulatory clarity is creating numerous potential commercial opportunities, as industry participants continue to acknowledge the reliability that our solutions can provide, both directly to the customer, but also to the broader power supply, transmission and regulatory ecosystem."

— Kyle Ramachandran, Solaris Energy Infrastructure President and CFO

Ireland, which arguably serves as a case study for the challenges in balancing the twin transitions of digitalization and decarbonization with data centers making up 21% in its

national electricity demand in 2023 and growing (as detailed in Green Data Centers: Ireland's electricity dilemma – where next), recently enacted a private wire policy that enables private investment in wire connections between power supply and users (e.g., a wind farm directly powering a data center). Ireland is also proposing that data centers be required to develop on-site or nearby "dispatchable" power supply or energy storage that meets a data center's max capacity after adjusting for a de-rating factor (i.e., ~120 MW of on-site gas capacity for 100 MW data center). The proposed policy also requires the on-site/nearby power or storage to participate in the wholesale electricity market and be centrally dispatched by the grid operator so the grid operator can facilitate the optimum low emissions solution on a whole system operational basis by taking advantage of renewable electricity during periods of high availability.

Distributed Power was a Major Earnings Theme

Demand for distributed and backup power solutions for data centers was an emerging theme during this recent earnings seasons. This includes strong orders for aeroderivative gas turbines, industrial gas turbines, industrial reciprocating gas and diesel engines, and gas/hyrodgen-powered fuel cells. We also see an increasing role for energy storage, not only for lithium-ion batteries, but also long-duration energy storage (LDES) technologies.

FIGURE 16. Overview on the different distributed power solutions

Technology	Power Output	Fuel Options	Pros	Cons	Target Use	Major OEM Suppliers
Heavy-duty gas turbines	Very high (100+ MW)	Natural gas, biogas & hydrogen	High power capacity with robust & reliable design; low relative cost per MW	Slow start-up times; requires large footprint (not mobile)	Ideal for utility-scale continuous base load power & combined cycle applications	GE Vernova, Mitsubishi Power, Siemens Energy
Industrial gas turbines	Broad (1-100+ MW)	Natural gas, biogas & hydrogen	Robust & reliable design; low relative cost per MW	Slow start-up times; poor part-load efficiency vs. reciprocating engines	Ideal for distributed base load power & combined cycle applications	Baker Hughes, GE Vernova, MAN Energy Solutions (Volkswagen), Mitsubishi Power, Siemens Energy, Solar (Caterpillar)
Aeroderivative gas turbines	Medium/High (30-140 MW)	Natural gas, biogas & hydrogen	Relatively fast start-up time (~5 mins); compact & light (derived from aircrasft engine); fast installation & can be mobile	Cost per MW;relatively high maintenance & less rugged; less optimal for baseload power; poor part-load efficiency vs. reciporcating engines	Ideal for quick start/shutdown (e.g., gas peaking) & distributed power	GE Vernova, Mitsubishi Power (acquired Pratt & Whitney Power Systems)
Reciprocating engines(gas & diesel)	Low (<20 MW)	Natural gas, biogas, hydrogen, diesel & renewable diesel	Very fast startup times, good part-load efficiency vs. turbines, mobile	Limited power output; large foot print per MW; high emissions	Backup power, industrial applications, oil & gas	Caterpillar, Cummins, Generac, Jenbacher (private), Man Energy Solutions (Volkswagen), Mitsubishi Power, MTU (Rolls-Royce), Wartsila, Waukesh (private)
Fuel cells	Modular	Natural gas, biogas, hydrogen	High relative efficiency vs. turbines & engines; low emissions; easy to site & quiet; low maintenance	High capital costs, high fuel purity	Distributed power	Bloom, Plug Power, Ballard Power
Energy Storage	Modular	Electricity	High efficiency; can be paired with renewables	High capital costs; limited duration; safety concerns (thermal runaway)	Balance grid supply/demand, peak shaving, power price arbitrage	Tesla, Fluence, LG Chem, BYD, CATL

Source: Barclays Research

Heavy-duty Gas Turbines

Siemens Energy (ENR, covered by Vlad Sergievskii) reported that 3 GW of its 9 GW in Gas Service orders during fiscal 3Q25 was related to data centers, primarily mid-sized F-class turbines, driven by the US, which accounted for ~50% of total Gas Service orders. Year to date, over 8 GW of its 14 GW of Gas Service orders was for data centers. ENR noted during its fiscal 3Q25 earnings call that F-class heavy duty turbines are a "good balance between size and speed" given faster start-up times than its HL-class turbines. It indicated that it is generally booked out until 2027 and "2028 already has a good orderbook." ENR also suggested power

needs in the Middle East for data centers is the "next big wave". ENR believes its partnership with **Rolls-Royce** (RR, not covered) for small modular reactors (SMR) positions it well for the SMR opportunity, but emphasized that it doesn't expect the SMR technology to be ready for the market until 2030+.

> "We currently see an extremely strong market in the United States for energy solutions, so both in terms of the replacement of outdated infrastructure and enormous demand for electricity for data centers, in particular are now driving very high demand for our products in the US. Gas Services, for example, has secured gas turbine orders totally 14 GW. 60% of those are for data centers. "
>
> *— Christian Bruch, Siemens Energy President and CEO during fiscal 3Q25 conference call*

GE Vernova (GEV, covered by Julian Mitchell) has secured an order for 7 HA-class gas turbines to provide up to 4.4 GW of data center power at the Homer City Development, which was previously the site of the largest coal plant in Pennsylvania. The turbine deliveries are expected to begin in 2026.

Aeroderivative Gas Turbines

GE Vernova also reported a sharp increase in orders for its aeroderivative gas turbines (27 units ordered in 2Q25 vs. only 1 unit in 2Q24), in large part to support data centers. This includes 19 units booked in June 2025 by Crusoe, the developer of the Stargate I data center project, bringing Crusoe orders to 29 LM2500XPRESS aero turbines with nearly 1 GW of power (~35 MW / unit) having previously ordered 10 units in December 2024.[21] **Lancium** (private, not covered), which owns the land that the Stargate I project is being developed on, has indicated that the site has undergone a rigorous, formal review process in collaboration with transmission service providers and has 1.2 GW of grid interconnection fully approved by ERCOT.[22] In addition to the 1.2-GW grid interconnection, the Stargate I project will include large-scale behind-the-meter natural gas turbines and solar plus energy storage to help ensure grid reliability.[23]

[21] Crusoe: GE Vernova and Crusoe Announce Major 29-unit Aeroderivative Gas Turbine Deal to Deliver Power to AI Data Centers, accessed July 2025

[22] Lancium: Abilene, TX Clean Campus, accessed July 2025

[23] Crusoe: Crusoe Expands AI Data Center Campus in Abilene to 1.2 Gigawatts, accessed July 2025

> "There's the need for incremental bridge power. And the beauty of Aeroderivatives is they can be commissioned faster and and that's needed in the environment today. And our customers are able to price at a premium for expedited power. So Aeroderivatives are a very attractive solution right now."
>
> *— GE Vernova CEO Scott Strazik during 2Q25 earnings call*

Industrial Gas Turbines

Baker Hughes (BKR, covered by David Anderson) noted "demand accelerating" for data centers with over $550 million in data center power generation orders booked in 2Q25. Data centers accounted more than 70% of Baker Hughes' 69 NovalLT gas turbines booked in 2Q25, including 30 units (~500 MW, implying ~17 MW/unit) for data centers in the US and another 17 units (up to 270 MW or ~16 MW/unit) from **Frontier** (private, not covered) for data centers in Wyoming and Texas. With about $650 million of data centers orders YTD, including 1.2 GW of NovalLT gas turbine capacity, Baker Hughes believes it can meet or exceed its 3-year (2025-2027) data center equipment orders target of $1.5 billion earlier than planned.

Solaris Energy Infrastructure (SEI, covered by David Anderson) reported ~600 MW of distributed power generation with equipment orders to scale to 1.7 GW by 2H17 with 67% contacted to data centers and 8% contracted to the energy sector. This includes a mix of gas turbines from GE Vernova and Solar (a division of Caterpillar). Solaris noted it continues to see 3-4 year paybacks on its distributed power equipment. This includes a 900-MW commercial contract in a JV for an initial term of 7 years to support "a new data center."

Williams Energy (WMB, covered by Theresa Chen) disclosed with 1Q25 earnings a long-term, fixed-price power purchase agreement to invest $1.6 billion in on-site power and supporting infrastructure for two 200 MW facilities (Socrates North and Socrates South) to power **Meta's** 1-GW Prometheus data center campus in New Albany, Ohio. The power generation reportedly includes 30 gas turbines. Williams noted it has two other projects that are utilizing the same model and has ordered equipment with the same agreements as used in the Socrates projects.

xAI has received permit approval for 15 stationary gas turbines with a combined 247 MW.[24] xAI's Colossus data center, which operates ~230k GPUs, initially had 50 MW of grid power, which increased to 150 MW earlier this year; the company expects to complete a second 150-MW substation by year-end. Elon Musk revealed Colossus 2, which he described as the "the first Gigawatt AI training cluster"; it is near Memphis, TN, and will initially deploy ~550k GPUs. He confirmed that xAI has purchased an overseas power plant and is seeking to import it to the US, where it will presumably be installed at a former Duke Energy site in Mississippi that is in closer proximity to the reported Colossus 2 site.[25] [26]

[24] Data Center Dynamics: xAI granted permits for 15 gas turbines at Memphis data center, accessed July 2025

[25] Data Center Dynamics: xAI importing power plant from abroad for new Memphis data center, Elon Musk claims, access July 2025

[26] Data Center Dynamics: Elon Musk's xAI buys former gas power plant site in Southaven, Mississippi, accessed July 2025

Industrial Gas & Diesel Reciprocating Engines

Caterpillar (CAT, covered by Adam Seiden) reported a 19% y/y increase in Power Generation sales during 2Q25, driven by strong demand for its large Solar reciprocating gas engines, primarily for data centers, more than offsetting a slight decline in turbines and turbine-related services "due to timing". CAT expects full year Power Generation growth as demand remains strong for both prime and backup power applications to support data center growth. The company suggested it still has some supply capacity for near-term orders as it continues to add capacity, but is also taking orders further out given strong line of sight from the largest data center customers, adding that, "We continue to stay close to our largest data center customers and receive regular feedback on their long-term demand." CAT announced an agreement with Joule Capital Partners and Wheeler Machinery Co. to power Joule's 4 GW data center campus in Utah with CAT's latest G3520K generator sets and 1.1 GWh of grid forming battery energy storage.

Generac (GNRC, covered by Christine Cho) noted strong reception with 2Q25 earnings to its new large MW diesel backup generators for data centers with over $150 million in global data center backlog already with initial shipments to international customers in 3Q25 and domestic customers in late 2025. It emphasized market supply constraints for backup power equipment, stating "over the last several months, nearly every data center developer, operator, owner, end customer has told us that there are two major components that they worry about in the lead time for construction of new data centers. The first is transformers and the second is backup generators." GNRC believes there is a structural deficit in 2026 alone of 5,000 backup power units based on current market capacity in the market and data center construction completion timelines. At $1 million all-in cost per unit, the company suggests this equates to $5 billion supply/demand gap. It continues to ramp its capabilities for large MW generators with expected annual production capacity of over $500 million. The company emphasized traction with traditional and hyperscaler data center owners, but its "more interesting conversations" are with the hyperscalers given their extensive power needs leading to the biggest part of the backup power deficit with planning discussions for 2027 and out to 2029. The company also believes increasing data center power requirements will lead for demand for data center backup power to "continue to grow at a dramatic rate for the foreseeable future."

> "In our early conversations here, over the last several months, nearly every data center developer, operator, owner, end customer has told us that there are two major components that they worry about in the lead time for construction of new data centers. The first is transformers and the second is backup generators...we believe, based on our conversations, that there appears to be about a structural deficit just in 2026 of something on the order of maybe 5,000 machines based on current capacity in the market and based on current construction completion timelines for the projects that are underway for data centers."
>
> *— Aaron Jagdfeld, Generac Holdings Chairman, President and CEO during 2Q25 earnings call*

Cummins (CMI, covered by Adam Seiden) reported a 25% y/y increase in North America power generation equipment sales during 2Q25, driven primarily by continued strong demand in data centers and mission-critical applications. The company also reported a 32% y/y increase in Chinese power equipment sales due to accelerating data center demand. CMI is doubling the capacity of its Power System business, which it expects to be fully online by the beginning of next year.

Rolls-Royce (RR, not covered) noted that it continues to capture growth in the data center market with "very strong" demand for backup power generators. The company reported 68% y/y Power Systems order growth in 2Q25, primarily driven by data centers where orders grew 85% y/y. RR now expects midterm revenue growth of ~20% annually vs. 15%-17% previously for its Power Generation segment and is expanding its US manufacturing capabilities "especially for the rapidly growing American data center industry".

Wartsila (WRT1V, covered by Vlad Sergievskii) secured its first US data center order during 2Q25. The company will supply ~280 MW of power for a data center in Ohio using 15 natural gas engines. Warsila has emphasized that its engines offer more flexibility than gas turbines and as little as 18 month lead times vs. 3-4 years for large gas turbines. It believes data centers continue to offer interesting business potential in both equipment and services.

Fuel Cells

Bloom Energy, which supplies on-site fuel cells powered by natural gas or hydrogen, has deployed over 400 MW to power data centers globally, announced it will supply its fuel cell technology to **Oracle** within 90 days to power an islanded data center in the US (i.e., not connected to the grid). Bloom plans to double its factory capacity from 1 GW per year to 2 GW per year by the end of next year. The company emphasized that its systems do not need multiple AC-to-DC converters or specialized equipment to suppress harmonics, which helps lower costs, increase reliability, and reduce carbon footprint. In addition, the company claims the systems generate virtually no air pollution and use no water, which helps expedite air permits and improve speed-to-power. In late 2024, **American Electric Power Co.** (AEP, covered

by Nick Campanella) announced an agreement with Bloom to acquire up to 1 GW of fuel cells to enable data center customers to continue to expand their operations while transmission infrastructure is built. In May 2025, AEP Ohio secured approval for customer contracts that utilize fuel cell technology. AEP described Bloom's fuel cells as "a low-risk approach to bridging data center load from first power to ultimate grid connection," adding that it has provide fuel cells to Amazon AWS and **Cologix** (private, not covered) while the company worked through the interconnection agreements with them, which is probably 5-7 years. **Chesapeake Utilities** (CPK, covered by Nick Campanella) has entered into an agreement with AEP to construct and operate a $10 million intrastate natural gas pipeline to power AEP's fuel-cell facility. The pipeline is expected to be online in 1H27.

> “The large hyperscalers put together are going to spend more than $1 billion a day on CapEx, weekday and weekend. It's more than $500 billion are going to be spent just in this calendar year by those people. So, you take that number of $500 billion and you say an order of magnitude down, at least $50 billion of power capital equipment needs to be spent to electrify that additional demand that's going to come on. And you take that and you do a simple math and say more than one sizable nuclear power plant's worth of baseload is what is needed every month. And we all know what can be done by the existing legacy electric infrastructure in this country. It cannot move at AI speed.”
>
> *— KR Sridhar, Bloom Energy Chairman and CEO during 2Q25 earnings call*

Energy Storage

Tesla's Megapacks (TSLA, covered by Dan Levy) are being used by xAI. According to Elon Musk, xAI's Colossus 2 data center will include 168 Tesla Megapacks, noting their capacity to "run an entire neighborhood." Tesla deployed 9.6 GWh of energy storage products during 2Q25.

"The Energy business is more critical than ever. The availability of clean, reliable energy is necessary for economic growth and an imperative for the development and commercialization of AI enabled products and services. As electricity demand grows, our Megapack product helps to increase utilization of existing generation and transmission capacity, resulting in a more efficient use of the electric grid. When paired with solar PV, Megapack is cost competitive with traditional fossil fuel generation assets and can be deployed 4x faster than traditional fossil fuel plants of the same capacity. Trailing twelve-month Energy storage deployments achieved their 12th consecutive quarterly record."

— Tesla 2Q25 update

Google recently announced that it has partnered with **Energy Dome** (private, not covered) to globally deploy Energy Dome's CO_2 Battery, a long-duration energy storage (LDES) technology that can store and dispatch renewable energy for up to 24 hours. This marks Google's first commercial LDES deal and supports its goal to operate entirely on carbon-free energy by 2030. The CO_2 Battery uses a closed-loop system with carbon dioxide, offering a scalable and cost-effective alternative to lithium-ion batteries. The partnership includes multiple projects globally, aiming to enhance grid reliability and accelerate the transition to 24/7 clean energy. Google has also made a strategic investment in Energy Dome. As we detailed in Utility-scale Energy Storage: Long-duration Energy Storage Technologies, rising energy storage duration requirements are increasing urgency to advance LDES technologies or risk stunting the energy transition.

FIGURE 17. **Select companies with distributed power thematic exposure**

Company	Distributed Power Focus	Ticker	Market Cap	Analyst
Bloom Energy	Fuel cells	BE.US	$10 bn	Not Covered
Caterpillar	Industrial turbines, recip engines	CAT.US	$194 bn	Adam Seiden
Cummins	Recip engines	CMI.US	$54 bn	Adam Seiden
Generac	Recip engines	GNRC.US	$12 bn	Christine Cho
Liberty Energy	Distributed power provider	LBRT.US	$1.8 bn	David Anderson
Plug Power	Fuel cells	PLUG.US	$1.7 bn	Not Covered
Rolls-Royce	Recip engines	RR.LN	£92 bn	Not Covered
Solaris Energy Infrastructure	Distributed power provider	SEI.US	$2.1 bn	David Anderson
Baker Hughes	Industrial turbines	BKR.US	$42 bn	David Anderson
GE Vernova	Aero turbines, industrial turbines	GEV.US	$178 bn	Julian Mitchell
Siemens Energy	Industrial turbines	ENR.GY	€83 bn	Vlad Sergievksii
Tesla	Energy storage	TSLA.US	$1,094 bn	Dan Levy
Mitsubish Heavy Industries	Aero turbines, industrial turbines, recip engines	7011.JP	JPN13,360 bn	Not Covered
Fluence	Energy storage	FLNC.US	$1.3 bn	Christine Cho
Wartsila	Recip engines	WRT1V.HE	€14 bn	Vlad Sergievskii
Williams Energy	Distributed gas power & infrastructure	WMB.US	$70 bn	Theresa Chen

Source: Barclays Research

Related Research

- Powering AI Chart Pack Summary (28 Slides) (06.06.25)
- Powering AI: Google Reports Surging 2024 Electricity & Water Use (01.03.25)
- Powering AI: MSFT electricity use +27% in FY2024, transitioning to liquid cooling (11.06.25)
- Powering AI: Evidence of Strong AI Inference Demand Growth (24.06.25)
- Powering AI: Post-DeepSeek Bull/Bear Debate (06.03.25)
- AI revolution: Meeting massive AI infrastructure demands (16.01.25)
- Powering AI: Insights across Economics, Securitized Credit, Industrials, Tech & Energy (17.01.25)
- Powering AI: The Power & Land Arms Race (23.07.24)
- Powering AI: Calibrating US Data Center Energy Demand (27.06.24)
- Powering AI - Data Center World event: bullish sentiment, MSFT theories, power constraints (17.04.25)
- Powering AI: AI arms race vs. China, IEA data center energy forecasts & MSFT data center plans (10.04.25)

Analyst(s) Certification(s):
We, William Thompson, Hiral Patel, Katherine Ogundiya, Hannah Greenberg, Amy Lian, Laia Marin i Sola, CFA and Siddharth Gupta, hereby certify (1) that the views expressed in this research report accurately reflect our personal views about any or all of the subject securities or issuers referred to in this research report and (2) no part of our compensation was, is or will be directly or indirectly related to the specific recommendations or views expressed in this research report.