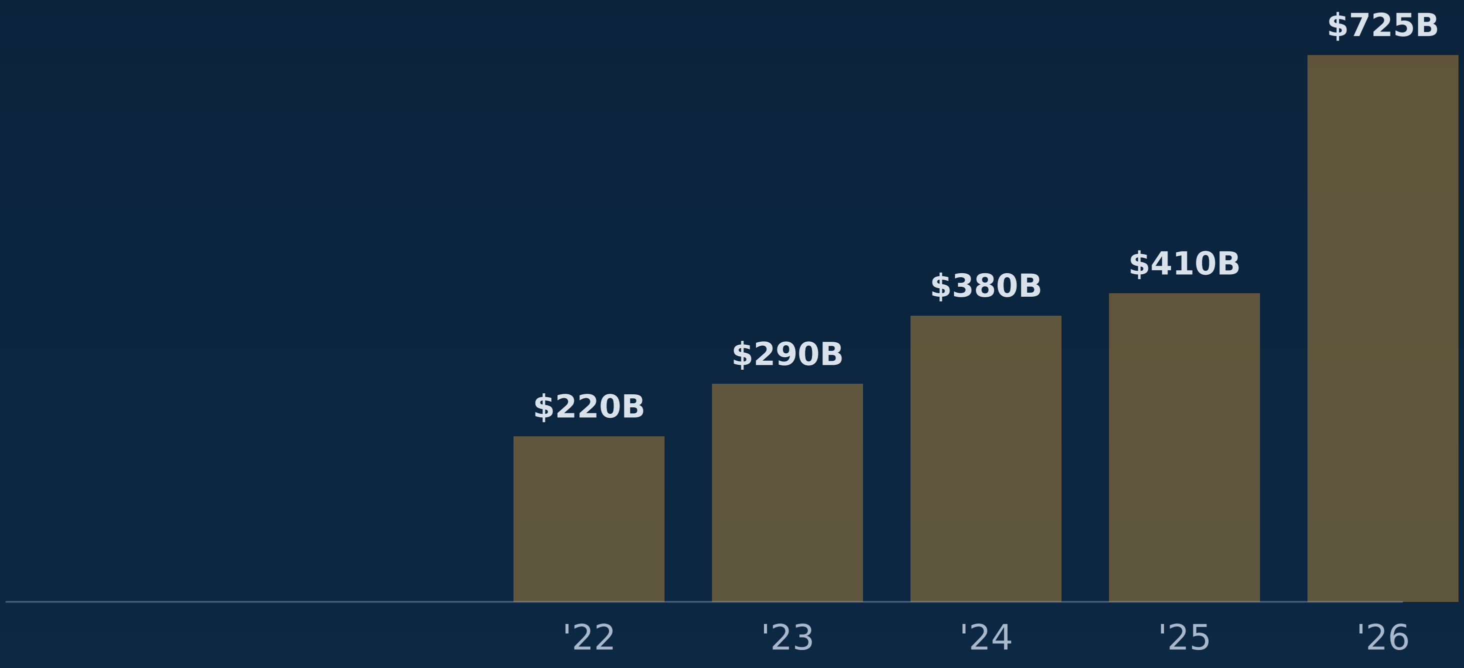
OVERLOAD

How AI infrastructure demand is stress-testing the engineering industry



HYPERSCALER CAPEX, $BILLIONS

Contents

New York: 530 Fifth Avenue, 9th Floor, New York, NY, 10036, USA
Istanbul: Ecza Sokak, No: 6 Kat.3 İç Kapı No:11 Safter İş Merkezi Şişli / İstanbul
Email: contact@interestingengineering.com

Executive Summary

In calendar year 2026, the five largest cloud and AI infrastructure providers — Amazon, Microsoft, Alphabet, Meta and Oracle — will spend somewhere between $660 billion and $725 billion on capital expenditure, a 70 to 77 percent increase on 2025 and the single largest concentrated infrastructure investment programme of the post-war era.[1] Roughly three-quarters of that spend — around $450 billion — is directly tied to AI infrastructure: GPUs, servers, networking gear and the data centres to house them.[3] By any historical comparator — the Apollo programme, the Interstate Highway System, the post-war European reconstruction in inflation-adjusted terms — the AI data centre build-out is now larger.

What this report sets out to do is something the trade press has largely failed to do: stand back and look at the build-out as an engineering capacity problem rather than as a technology investment story. The picture from that vantage point is uncomfortable. US data centre electrical demand will reach 75.8 GW in 2026 and is forecast to hit 134.4 GW by 2030 — nearly tripling within the decade.[6] The construction industry is short 439,000 to 499,000 workers, most of them in the exact electrical and mechanical specialisms data centre builds require.[22] Hyperscalers have signed nuclear power deals exceeding 9 GW of near-term capacity and scaling toward approximately 14 GW by 2039 across more than a dozen announced projects.[17] And yet the trade-press conversation remains dominated by the GPU supply story, when the real bottlenecks have moved decisively to transmission, generation, cooling, water and labour.

Five things engineering leadership needs to know

First, electrical engineering capacity is now the single most binding constraint on the AI build-out. Transformer lead times have extended to two to three years on large units; substation engineering work is being prioritised by hyperscaler contracts at the expense of utility-scale renewables; medium-voltage and high-voltage switchgear is on allocation. Firms with deep electrical engineering benches are pricing services at premiums not seen since the early-2000s data-centre boom.[35]

Second, the geography of the build-out has concentrated to a degree that is itself becoming an engineering risk. Virginia's data-centre load reached 12.1 GW in 2025 — a 23 percent increase over 2019 summer peak load in the Dominion zone alone.[8] Texas is at 9.7 GW and growing 10 percent annually inside ERCOT.[11] Half of all new US data-centre capacity is being built in two states with already-stressed grids and increasingly stressed water tables. The diversification of siting is now a strategic imperative for hyperscalers and an opportunity for engineering firms with multi-state delivery capability.

Third, water is moving from a peripheral concern to a binding constraint. US data centres consumed 17.4 billion gallons in 2023; forecasts now put 2028 consumption at 38 to 73 billion gallons.[13] Newer hyperscale facilities require 10 to 40 times the power — and therefore the cooling load — of pre-2015 data centres.[14] Liquid cooling provides 30 to 40 percent power savings versus air cooling but introduces an entirely new engineering specialism that most existing data-centre EPC firms are still building out.

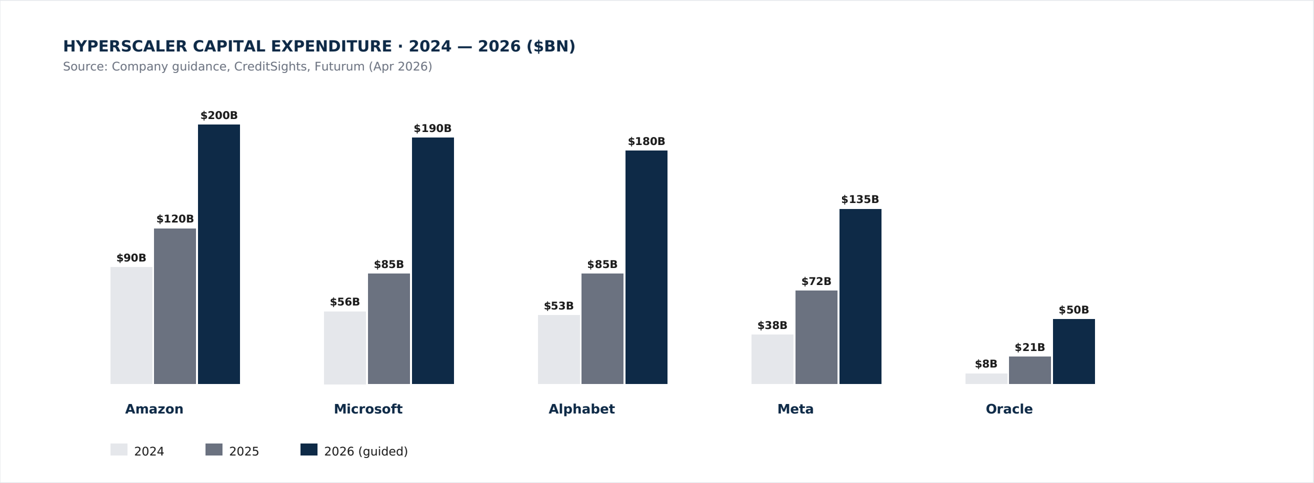
Fourth, the response is nuclear at scale. Every major hyperscaler has now signed at least one nuclear power deal. Microsoft locked in 835 MW from the Three Mile Island restart on a 20-year, $16 billion power purchase agreement.[18] Meta has committed to 6.6 GW under January 2026 deals with Vistra, TerraPower and Oklo, plus a separate 1.1 GW deal with Constellation (Clinton, Illinois) — approximately 7.7 GW in total. Amazon has an offtake agreement with X-energy for up to 12 Xe-100 reactors scaling to 5 GW by 2039 at a Washington site. Google has contracted 500 MW from Kairos.[17] For nuclear engineering services firms, this is the most coherent demand signal of the past thirty years.

Fifth, the human resource situation is genuinely dire. Data centre projects are now carrying backlogs of roughly 10 to 11 months on average — versus approximately 8 months for the broader construction sector.[24] ITIF estimates a general US construction workforce shortfall of approximately 439,000 workers heading into 2026; a separate industry study focused on data-centre-relevant trades places the shortfall at up to 499,000.[23] Sixty-five percent of contractors expect the data-centre market to expand further in 2026, with only 8 percent expecting contraction.[26] Electricians, HVAC technicians, pipefitters and instrumentation engineers are being bid out of every adjacent sector — semiconductor fabs, utility-scale renewables, transmission upgrades — to staff data-centre work. The industry is not building enough new talent to close this gap inside the planning horizon.

The strategic question

Engineering leadership has been reading the AI infrastructure story through the wrong lens. The relevant question is not “how much will hyperscalers spend?” — that question is now settled, the answer is staggering, and the firms doing the spending are not going to be deterred by short-term economic conditions. The relevant question is which engineering disciplines have the people, the supply-chain access, the geographic footprint and the contractual posture to deliver against an order book that is already two to three times larger than the workforce can absorb. This report works through that question discipline by discipline, ends with three twelve-to-thirty-six-month scenarios, and offers a recommended set of strategic moves.

Context: The Trillion-Dollar Build-Out



The pattern driving the build-out is now well documented: each generation of frontier AI model has required roughly an order of magnitude more compute than the last, and the firms training those models have decided — more or less simultaneously — that the binding constraint on their growth is no longer software but physical infrastructure. That decision arrived in 2024. By the start of 2025 it had crystallised into industry-wide capital expenditure commitments of roughly $410 billion. In 2026 that commitment has nearly doubled.[2]

The numbers

Amazon will spend approximately $200 billion in 2026, the largest single capex programme of any company in history. Microsoft's calendar-year 2026 capex is $190 billion. Alphabet has guided $175 to $185 billion. Meta — which a year ago was the smallest of the four hyperscalers by infrastructure spend — has raised its 2026 capex to $125 to $145 billion. Oracle, the late entrant, is targeting $50 billion.[1] The total — somewhere in the $660 to $725 billion range — is roughly equivalent to the entire 2024 US federal defence budget.

HISTORICAL COMPARISONS

Apollo: $25–28 billion in 1973 dollars. Inflation-adjusted estimates run from approximately $187 billion (CBS / standard CPI) to $338 billion (Planetary Society, using NASA's New Start Index). The Manhattan Project: $1.89 billion in 1940s dollars, approximately $30–36 billion in 2024–2026 dollars depending on methodology.

By either measure, combined hyperscaler 2025–2026 capex will exceed the inflation-adjusted cost of Apollo and the Manhattan Project together.

The closer historical analogue is the late-1800s US railroad construction boom, which at peak consumed an estimated 15–20 percent of US GDP. AI infrastructure spend in 2026 is approximately 2.5 percent of US GDP — large, but not yet on that scale.

Where the money goes

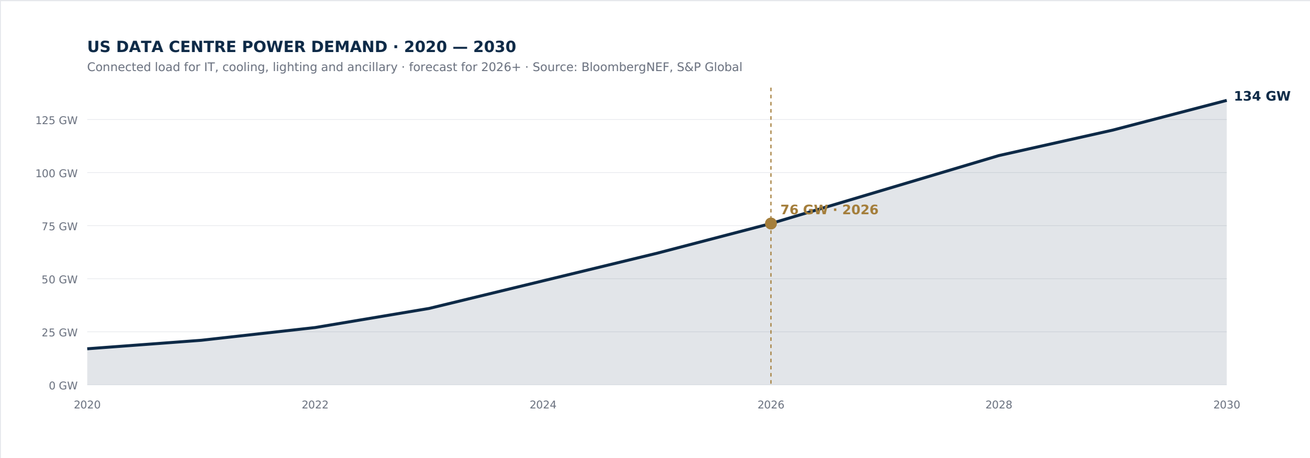
Roughly $450 billion of the 2026 spend — about 75 percent — is directly tied to AI infrastructure: GPUs, AI-specialised servers, high-bandwidth networking, and the physical data centres to house them.[3] The GPU share alone is approximately $200 billion. The remainder is the building, the power, the cooling, the land, the security, and the engineering and construction services to bring all of it online. The narrative attention on GPUs has obscured how large the non-silicon engineering share has become — roughly $250 billion in 2026, which is by itself larger than the entire global revenue of the top 100 engineering firms combined in 2024.

Direct Impacts on the Engineering Industry

The next six spreads work through the disciplines most directly affected by the AI infrastructure build-out, in order of immediate impact. Electrical engineering and grid / power generation are given the most space — not because civil, mechanical, water and supply-chain disciplines are less important, but because the magnitude of the demand signal in electrical and generation engineering is currently larger by an order of magnitude. Each spread follows the same structure: the scale of the demand, the technical and commercial implications, what we are seeing firms do, and what to watch.

Electrical Engineering

Electrical engineering is the binding constraint on the entire build-out. US data-centre power demand reached 62 GW in 2025; it will reach 75.8 GW in 2026, 108 GW in 2028, and 134.4 GW by 2030.[6] That is approximately 60 GW of new connected load to be designed, engineered and energised inside a five-year window — more new load than the US grid added across the entire 1990s. The engineering work to support it is concentrated in three sub-disciplines: medium and high-voltage substation design, transmission interconnection studies, and inside-the-fence-line electrical balance-of-plant.



SUBSTATIONS AND SWITCHGEAR

Hyperscaler campuses are now routinely specified at 250 to 500 MW per site, with the largest campuses (Stargate Abilene, Meta Richland Parish, Microsoft Mount Pleasant) approaching or exceeding 1 GW. Substation engineering work at this scale has historically been the province of utility-side specialists; it is now being procured by hyperscaler EPC contracts at a velocity that has pulled cleared and unbid utility work off the schedule. High-voltage substation design lead times have extended from 12 months to 24 to 30 months. Transformer lead times for the largest units are running 24 to 36 months. Medium and high-voltage switchgear is on allocation for the first time in fifteen years.

TRANSMISSION INTERCONNECTION

The transmission interconnection queue is the visible symptom of a deeper problem. Across PJM, ERCOT and MISO, the volume of new load applications waiting in the interconnection queue exceeds, in some utility territories, the total existing demand on the system.[35] So-called phantom requests — multiple duplicative applications for the same physical site filed across competing utility territories — have inflated the apparent queue and complicated planning. Engineering services firms with experienced interconnection-study capacity are at a premium; the cleared-personnel constraint inside that work is acute.

ENGINEERING FIRM RESPONSE

Across our coverage of the top tier of electrical engineering and grid services firms — Burns & McDonnell, Black & Veatch, Sargent & Lundy, Stantec, AECOM, Power Engineers, POWER Engineers, EPRI partners, Quanta Services — three patterns are visible. First, data-centre-segment revenue is up by 50 to 80 percent year-on-year at most firms with meaningful exposure. Second, those firms are reporting record offer rates and historic difficulty filling cleared positions. Third, multiple firms are repositioning their utility-scale renewables divisions toward data-centre interconnection work — capturing higher margin in the short term but exposing their renewable backlog to the slowdown.

Grid, Generation & Nuclear Engineering

HYPERSCALER NUCLEAR POWER DEALS · ~13 GW COMMITTED ACROSS HYPERSCALERS

Source: Utility Dive, World Nuclear News, SMR Intel deal tracker, company disclosures · *Microsoft contract value undisclosed; analyst estimates $98–$115/MWh PPA pricing

Hyperscaler	Counterparty	Capacity	Structure	Online
Microsoft	Constellation (TMI Unit 1)	**835 MW**	20-yr PPA*	2028
Meta	Vistra (Beaver Valley+)	**2.1 GW**	20-yr PPA	2026+
Meta	TerraPower Natrium	**up to 2.8 GW**	Funding + offtake	2032+
Meta	Oklo	**up to 1.2 GW**	PPA	2030+
Meta	Constellation (Clinton)	**1,121 MW**	20-yr PPA	Existing
Amazon	X-energy Xe-100	**up to 5 GW**	Offtake (12 SMRs)	by 2039
Google	Kairos Power	**500 MW**	PPA	2030+
Oracle	Undisclosed SMR	**TBD**	Under negotiation	TBD

The single most important engineering market that has emerged from the AI build-out is the nuclear renaissance. As of May 2026, every major US hyperscaler has signed at least one nuclear power purchase agreement, and the cumulative committed capacity exceeds 9 GW in the near term, scaling to approximately 14 GW by 2039 if Amazon's full X-energy offtake target is realised, across more than a dozen distinct projects.[17] That is more new nuclear procurement than the United States has commissioned in the entire post-Three-Mile-Island era.

THE DEAL STACK

Microsoft's 20-year power purchase agreement with Constellation Energy for the restart of Three Mile Island Unit 1 — 835 MW, expected online in 2028 — is the anchor deal that broke the dam. Constellation has not disclosed the headline contract value; Morgan Stanley estimated PPA pricing at approximately $98 per megawatt-hour, with Jefferies estimating up to $115 per megawatt-hour, implying a multi-billion-dollar twenty-year value. The plant has been renamed the Crane Clean Energy Center; it is approximately 80 percent staffed with over 500 employees; Constellation is investing $1.6 billion in the restart; and the US Department of Energy closed a $1 billion loan to support it.[18] Meta committed in January 2026 to up to 6.6 GW of nuclear capacity through three separate agreements — Vistra (2.1 GW from Beaver Valley plus Perry and Davis-Besse uprates), TerraPower (up to 2.8 GW across two 345 MW Natrium reactors at Kemmerer, Wyoming and up to six further units) and Oklo (up to 1.2 GW from advanced reactor builds). A separate June 2025 agreement with Constellation for the 1,121 MW Clinton Clean Energy Center in Illinois takes Meta's total nuclear commitments to approximately 7.7 GW. Amazon has an offtake agreement with X-energy for up to 12 Xe-100 small modular reactors at a site adjacent to the Columbia Generating Station in Washington, scaling to up to 5 GW by 2039; X-energy raised a separate $700 million Series D in November 2025 led by Jane Street, with participation from ARK Invest and other investors. Google's deal with Kairos Power is for 500 MW of advanced reactor capacity. Oracle has indicated active SMR procurement without disclosed counterparties.[17]

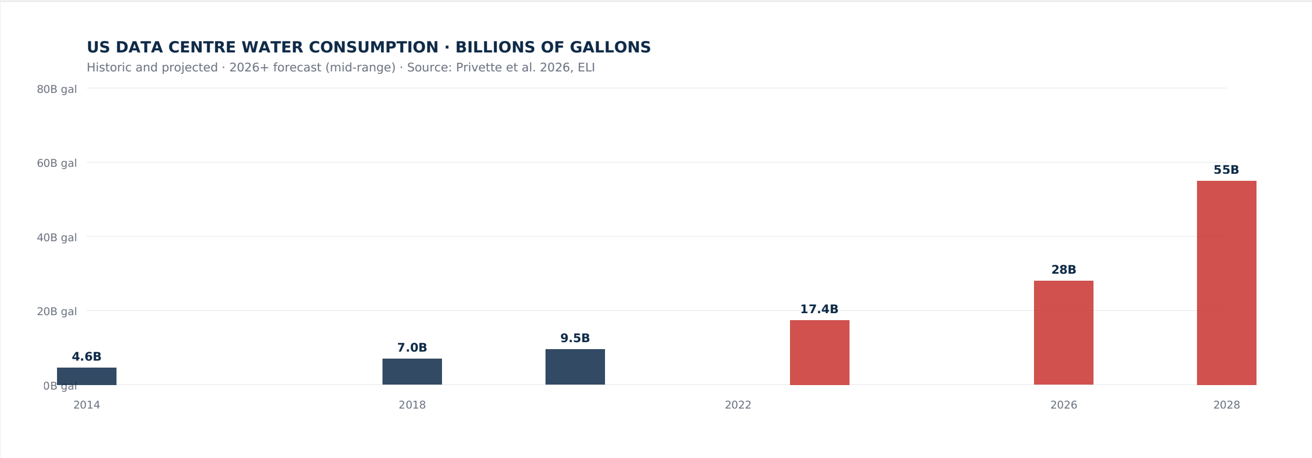
ENGINEERING IMPLICATIONS

For nuclear engineering services firms — Bechtel, Sargent & Lundy, Stone & Webster successor entities, Fluor Marine Propulsion, EnergySolutions — this is the most coherent demand signal in three decades. Engineering work spans four streams: restart engineering on existing reactors, SMR first-of-a-kind design certification support, balance-of-plant for new advanced reactors, and grid interconnection for new nuclear. Cleared-personnel constraints are particularly acute for nuclear-cleared engineers, where the workforce is older and the credential pipeline narrow.

GAS TURBINES AND THE BRIDGE GENERATION

Nuclear is the medium-term answer; the near-term answer is gas. EIA forecasts that fossil generation will rise faster than previously expected as data-centre load comes online faster than utility-scale renewable and nuclear capacity can support.[9] Combined-cycle gas turbine engineering work — sometimes inside the data-centre fence line as captive generation, sometimes outside as new utility-scale capacity — is running at order rates not seen since the early-2000s combined-cycle build cycle. GE Vernova has reported manufacturing constraints; Siemens Energy is at capacity through 2027 on heavy-duty gas turbine orders.

Mechanical & Cooling Engineering



Mechanical engineering — specifically the cooling subset — is the second binding constraint on AI infrastructure delivery. The thermal density of AI workloads is the structural change: GPU-dominated racks dissipate 50 to 130 kW per rack, versus 5 to 15 kW for traditional cloud workloads. Air cooling cannot economically remove this much heat at this density. The industry is in the middle of a forced transition to liquid cooling — direct-to-chip cold plate, rear-door heat exchangers, full immersion — and the engineering workforce able to design at scale is small.

LIQUID COOLING SPECIALISMS

Three liquid-cooling design competencies are now in acute demand. Direct-to-chip cold plate design is the dominant near-term approach and is being delivered by Vertiv, CoolIT, Schneider Electric, and a small number of specialist engineering services firms. Rear-door heat exchanger retrofits are the bridge solution for converting existing air-cooled facilities to support AI workloads, and the design work is concentrated at large MEP firms with deep mechanical benches. Immersion cooling — single-phase and two-phase — is the longer-term play and is being delivered by a smaller pool of specialists including Submer, GRC and LiquidStack. Engineering firms with all three competencies are pricing services at premiums of 40 to 60 percent over baseline mechanical-engineering rates.

WATER AS THE BINDING CONSTRAINT

Cooling-tower water consumption is the binding regulatory constraint in roughly a third of US data-centre markets. US data centres consumed 17.4 billion gallons of water in 2023, projected to triple to between 38 and 73 billion gallons by 2028.[13] A single hyperscale facility can consume 3 to 7 million gallons per day — the equivalent of a town of 10,000 people. Up to 85 percent of that water evaporates and does not return to the supply.[14] New York is considering a three-year moratorium on large data centres pending environmental review.[16] The engineering response — closed-loop cooling, recycled water systems, dry cooling at scale — is generating a discrete engineering specialism that did not meaningfully exist in 2023.

THE 1.2 GW CAMPUS

OpenAI's Stargate Abilene campus, built by Crusoe and operated by Oracle, is the largest data-centre engineering project under construction in the United States.

1.2 GW connected load. 4 million square feet of building. Eight buildings, four operational, four under construction. An estimated 0.3 GW already running with Nvidia GB200 racks; the first two buildings entered service in 2025.

An adjacent Vantage campus across the county line will deliver 10 buildings on a 1,200-acre site, with the first building scheduled for late 2026.

By 2027, the Abilene cluster will likely exceed the connected electrical load of any single US city west of the Mississippi outside of Houston, Dallas, Phoenix or Denver.

Civil, Structural & Site Engineering

Adjacent Vantage campus (1,200 acres, 10 buildings) under construction across county line · first building Q4 2026

Civil and structural engineering for data centres has historically been a comparatively boring discipline: large flat slabs, big steel buildings, robust foundations. The AI build-out is changing that on three dimensions. First, scale: campuses at 500 MW to 1 GW require site engineering at a magnitude closer to refinery or steel-mill construction than traditional data-centre work. Second, weight: GPU-dense racks with liquid cooling produce floor loadings of 4,000 to 6,000 lb per square foot, against traditional data-centre design loads of 1,500 to 2,000. Third, speed:

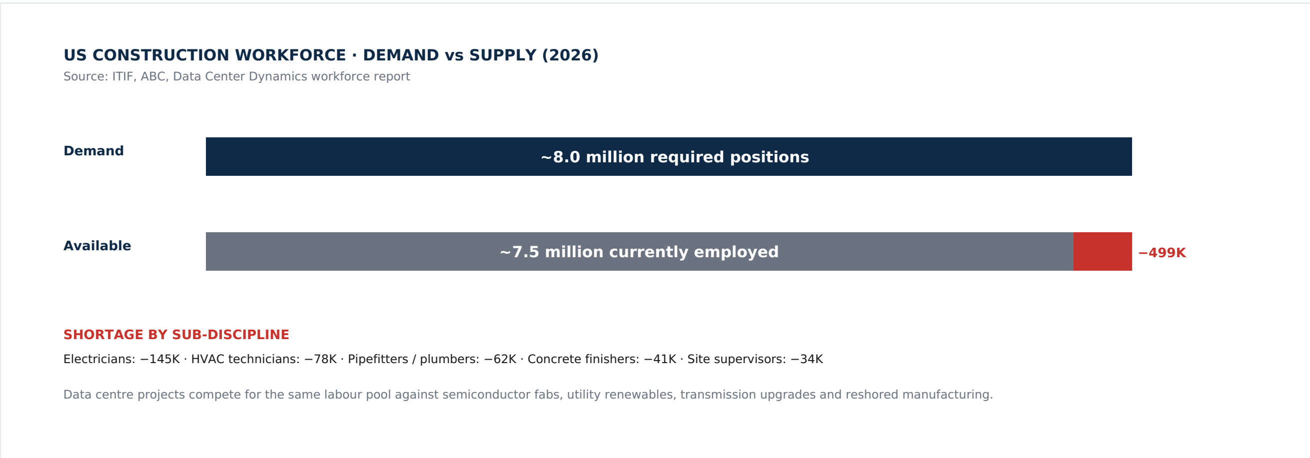
hyperscaler delivery schedules of 18 to 24 months from groundbreaking to commissioning are forcing concurrent design-build approaches that civil engineering firms in the data-centre space were not historically structured to deliver.

GEOGRAPHIC CONCENTRATION

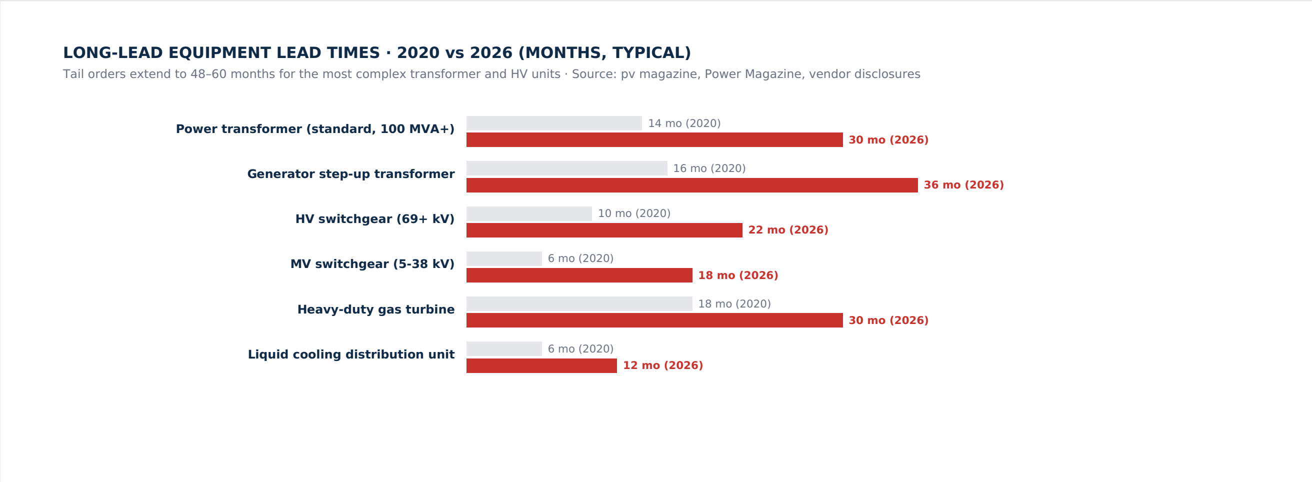
Half of all US data-centre development by connected load is concentrated in Virginia, Texas, Arizona and Georgia. PJM expects the Dominion zone in Virginia to see the largest absolute increase in summer peak demand of any US region between 2026 and 2030, almost entirely from data centres.[10] The 'Data Center Alley' designation around Ashburn now refers to a 90-mile corridor stretching from Manassas to Frederick, Maryland. Site engineering firms working that corridor are reporting record award rates; firms that established Virginia offices in 2022 to 2024 are out-bidding national-scale competitors.

CONSTRUCTION LABOUR

The single largest civil and structural risk factor is labour. ITIF estimates a US construction-workforce shortfall of approximately 439,000 heading into 2026; a separate iRecruit study focused on data-centre construction puts the data-centre-relevant trade gap at up to 499,000.[23] The shortage is concentrated in skilled electrical and mechanical trades and in pipelaying and concrete specialisms. Data-centre projects are now carrying backlogs of roughly 10 to 11 months on average versus approximately 8 months for general construction.[24] Sixty-five percent of contractors expect the data-centre market to expand further in 2026.[26]



Supply Chain: Transformers, Switchgear, GPUs



Three procurement categories now sit on the critical path of every large data-centre project: large power transformers, medium- and high-voltage switchgear, and AI GPUs.

TRANSFORMERS

Large power transformers — the 100 MVA-plus units that step utility transmission down to data-centre distribution — were on 12-to-18-month lead times in 2020. They are now on 24-to-36-month lead times for standard orders, with the largest and most complex units (generator step-up transformers and HVDC converters) extending to 48-60 months. Industry surveys put the average for standard power transformers at roughly 128 weeks and for generator step-up units at 144 weeks. Allocation now favours hyperscaler accounts. North American transformer manufacturing capacity — Hitachi Energy, Siemens Energy, GE Prolec, Eaton — is expanding but slowly, with new factory capacity coming online in 2027 and 2028. Engineering firms with strong procurement-engineering teams (the people who manage long-lead expediting) are quoting at premium rates and embedding procurement risk into FEED contracts that previously assumed standard lead times.

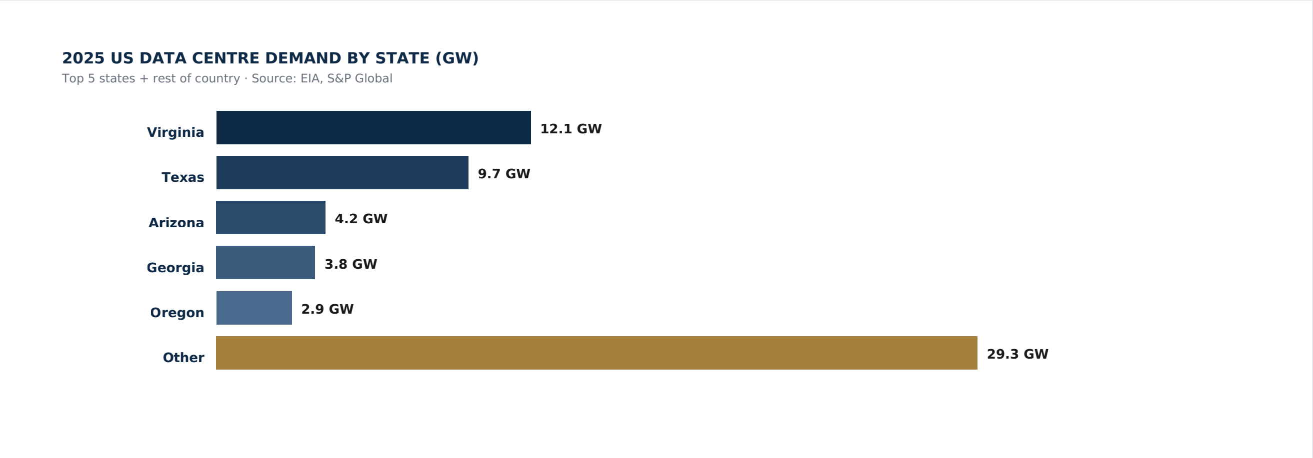
SWITCHGEAR AND BALANCE-OF-PLANT

Medium-voltage switchgear (5-38 kV class) and high-voltage switchgear (69 kV and above) are running 18 to 24 months. Inside-the-fence-line balance-of-plant — generators, paralleling switchgear, UPS systems, busbar — is on similar lead times. The supply concentration is severe: roughly five global manufacturers produce 80 percent of the units. Hyperscaler procurement contracts are increasingly bundling these items into multi-year framework agreements that lock up supply at the expense of utility and industrial customers.

GPUS

GPU supply — Nvidia H200 / B200 / GB200 series — has received the most popular attention but is, paradoxically, the procurement constraint that engineering firms can do least about. It is dominated by Nvidia's allocation decisions and TSMC's manufacturing schedule. From an engineering standpoint, the relevant observation is that GPU thermal envelopes are continuing to rise — the B200 is rated at 1,000 W per chip, against the H100's 700 W — which is the proximate driver of the liquid-cooling transition and a continuing source of building-services design uncertainty.

Geographic Concentration: The Cluster Problem



The concentration of data-centre development in a small number of US states is the single most important geographic story in engineering markets right now. Five states — Virginia, Texas, Arizona, Georgia, Oregon — account for roughly 60 percent of all US data-centre connected load. Virginia alone is at 12.1 GW of data-centre load against a state-wide commercial peak of less than 25 GW.[8] This is creating engineering implications that extend beyond data-centre EPC work into utility planning, transmission expansion, generation siting, and water-rights engineering.

THE VIRGINIA STRESS TEST

Virginia's Dominion zone summer peak load reached 23,905 MW in 2025, a 23 percent increase over 2019. The PJM forecast — which has historically been conservative — now expects the Dominion zone to see the largest absolute summer peak increase of any US region through 2030. Dominion Energy's transmission expansion programme has accelerated from a baseline of $5 billion per five-year cycle to over $15 billion. Engineering services firms working that programme — Burns & McDonnell, Quanta Services, MYR Group — have effectively unlimited near-term demand and the binding constraint is again cleared personnel.

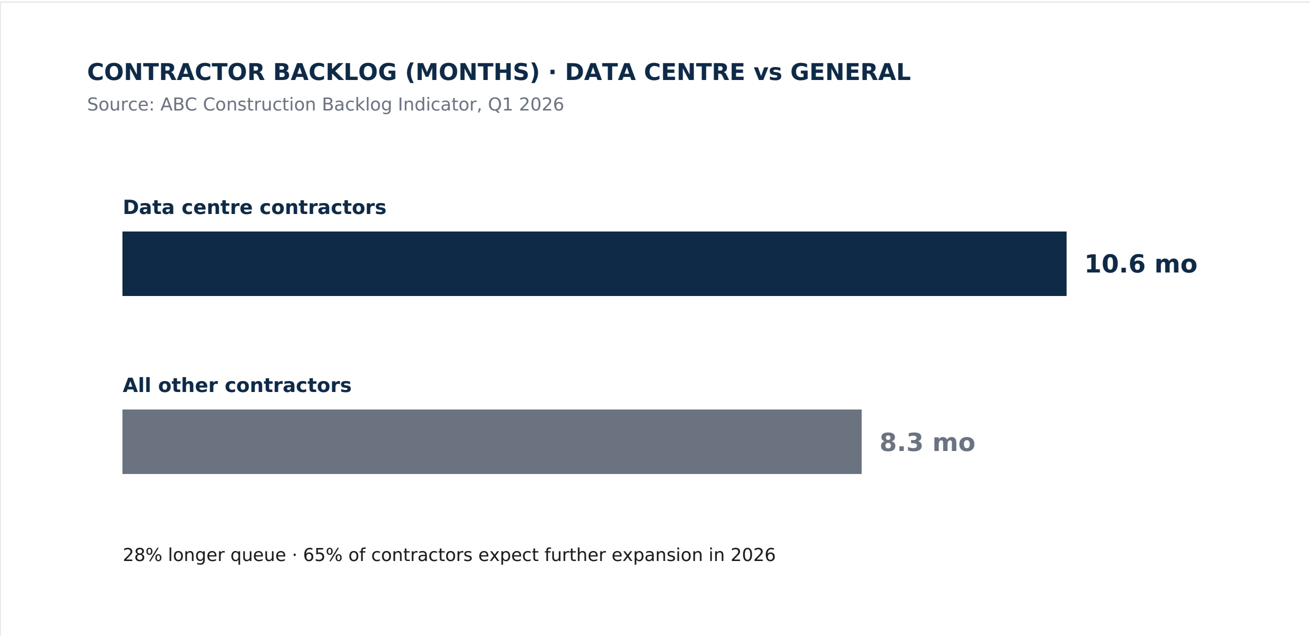
TEXAS AND ERCOT

Texas data-centre load reached 9.7 GW in 2025 and is forecast to grow at 10 percent annually through 2027 inside ERCOT. The Stargate Abilene campus — at 1.2 GW alone — will, when complete, represent roughly 10 percent of the entire Texas data-centre load.[30] ERCOT's independent grid status and accelerated interconnection processes have made it the preferred destination for new hyperscaler capacity that cannot get timely interconnection in PJM. Engineering firms with Texas presence and ERCOT interconnection expertise are commanding premium rates.

THE DIVERSIFICATION IMPERATIVE

Hyperscalers know that the concentration is itself a risk; they are now actively siting new campuses in less-stressed regions — Wyoming (Meta Kemmerer), New Mexico (Stargate), Ohio (Microsoft Mount Pleasant), Pennsylvania (the TMI-adjacent corridor). For engineering firms, this geographic spread is an opportunity: the firms that can credibly mobilise into emerging hyperscaler regions will capture work that legacy data-centre-corridor firms cannot serve.

Workforce & Project Economics



The labour situation in the data-centre engineering and construction stack is the binding industry-wide constraint. The construction industry as a whole is short between 439,000 and 499,000 workers; the data-centre subset of those projects is competing for the same labour against semiconductor fabs, utility-scale renewables, transmission upgrades and reshored manufacturing.[22] Pay pressure on skilled trades — electricians, HVAC technicians, pipefitters, instrumentation specialists — is running 8 to 12 percent annually and accelerating.[28]

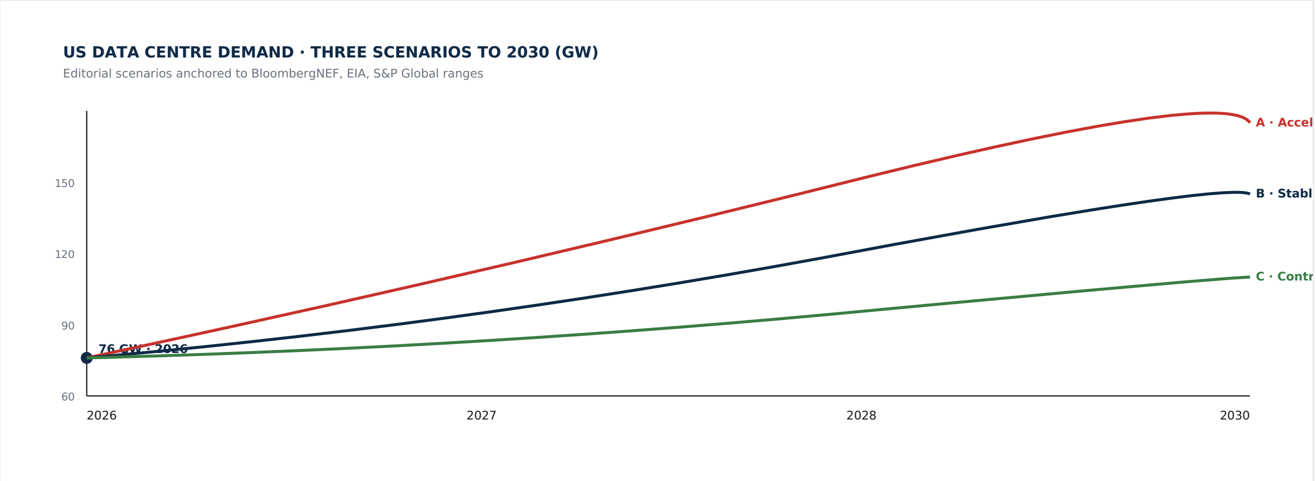
ENGINEERING SERVICES FIRMS

Inside the engineering services tier, the binding constraint is licensed and cleared engineers in three disciplines: electrical (medium and high-voltage), mechanical (specialised in liquid cooling), and nuclear (for the SMR and restart work). Most large engineering firms are reporting offer-to-hire ratios that have doubled since 2023 — meaning roughly half of formal offers are being declined for higher offers elsewhere. Engineer compensation in these specialisms has risen 18 to 25 percent over 2024 levels.

PROJECT ECONOMICS

Data-centre construction cost per MW of IT load has risen from approximately $9 million in 2022 to $12 to $14 million in 2026 for hyperscale-grade builds, with the highest-density AI-specialised facilities running $16 to $18 million per MW. Cost escalation is concentrated in three buckets: electrical infrastructure (transformers, switchgear, UPS), mechanical infrastructure (liquid cooling, water systems), and labour. Land and shell costs have been comparatively stable. For engineering services firms, this cost inflation is broadly margin-positive — fees scale with project cost — but it is forcing more rigorous procurement engineering and embedded contingency design than was historically required.

Three Scenarios for 2027–2030



Forecasting the trajectory of an industry committed to almost $750 billion of annual capex is genuinely difficult. We have framed the next three to five years across three scenarios, anchored on the variables engineering firms can actually plan against: the rate of AI model commercialisation, the pace of grid and generation buildout, and the political response to environmental impact.

SCENARIO A — CONTINUED ACCELERATION (PROBABILITY ASSIGNMENT: 30%)

AI commercialisation tracks current trajectory; hyperscaler capex grows another 30 to 40 percent in 2027 to roughly $1 trillion; the data-centre power demand curve trends toward the high end of forecasts (140 to 160 GW by 2030). Engineering firms in electrical, nuclear, mechanical and civil disciplines are at full capacity through 2028; labour bidding accelerates; pay inflation compounds. This is the bull case and is, in our view, less likely than the consensus suggests.

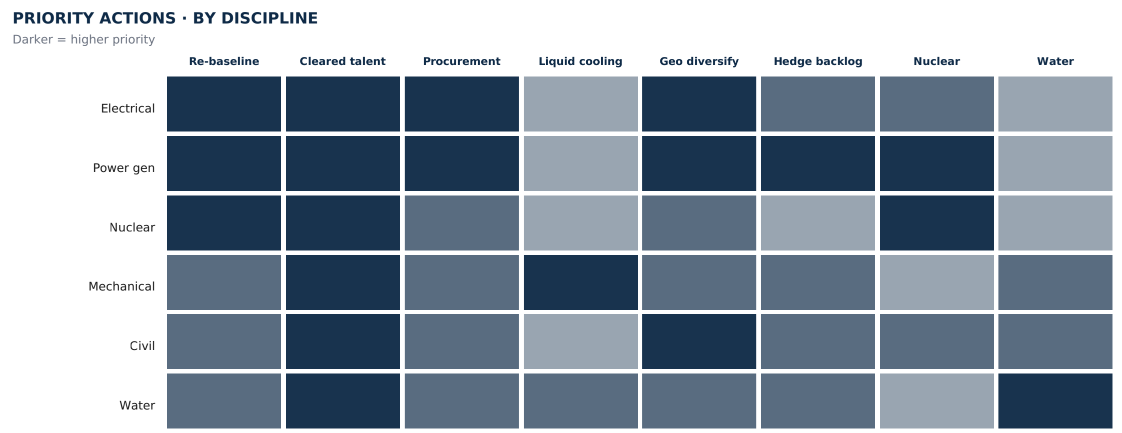
SCENARIO B — SUSTAINED-BUT-STABLE BUILD (PROBABILITY ASSIGNMENT: 50%)

Hyperscaler capex plateaus around current 2026 levels for two to three years before resuming growth; data-centre load reaches 110 to 125 GW by 2030; nuclear restarts and first SMRs come online roughly on current schedule; gas-turbine generation fills the near-term gap. Engineering firms see strong but not overheated demand through 2030. This is our central planning case and is the scenario most engineering firms should anchor on.

SCENARIO C — SHARP CONTRACTION (PROBABILITY ASSIGNMENT: 20%)

AI commercialisation under-delivers commercial returns; hyperscaler capex declines 30 to 40 percent in 2027 to 2028; data-centre load growth slows to 95 to 110 GW by 2030; project cancellations propagate through the engineering supply chain. This is the bear case and the tail risk engineering firms should hedge. In this scenario, the firms most exposed are those that have aggressively repositioned utility renewables and traditional industrial work toward data-centre delivery — the diversified portfolio is the hedge.

Recommendations for Engineering Leadership



Eight strategic moves we recommend for engineering firm leadership over the next twelve months. They are not specific to any of the three scenarios; they are designed to be robust across them.

1. QUANTIFY YOUR DATA-CENTRE EXPOSURE HONESTLY

Several firms in our coverage are reporting 'data-centre exposure' figures that conflate hyperscaler EPC work with adjacent utility work that happens to be in data-centre-heavy regions. The honest measurement is direct hyperscaler revenue plus utility revenue driven by data-centre load growth, segmented by discipline. Boards working with the conflated figure are not making good capital allocation decisions.

2. BUILD A CLEARED-PERSONNEL PIPELINE AGGRESSIVELY

Cleared electrical, mechanical and nuclear engineers are the binding constraint. Sponsorship programmes, veterans network partnerships, university clearance pathways, and accelerated internal training paths are no longer talent-management nice-to-haves; they are revenue-protection essentials.

3. LOCK IN LONG-LEAD PROCUREMENT FRAMEWORK AGREEMENTS

Transformer, switchgear and large-rotating-equipment lead times are not normalising in the next 24 months. Engineering firms that have framework agreements with manufacturers — and that are willing to take procurement risk on owners' behalf — will outbid firms that pass procurement risk through to owners. This is a structural shift in how data-centre engineering is sold.

4. BUILD LIQUID-COOLING COMPETENCY

The transition from air to liquid cooling is happening faster than most mechanical engineering practices are skilling for. Firms with credible direct-to-chip, rear-door heat exchanger and immersion cooling experience are pricing at 40 to 60 percent premiums to baseline mechanical engineering rates. Building this practice from inside (training existing mechanical engineers in liquid cooling specifics) is cheaper than buying it; either way, do not delay.

5. DIVERSIFY GEOGRAPHICALLY

The hyperscaler diversification away from Virginia and Texas into Wyoming, New Mexico, Ohio, Pennsylvania, Iowa, the Carolinas and the Pacific Northwest is happening now. Engineering firms without credible local mobilisation capability in these emerging hyperscaler regions are losing the new-site competitions. Office investments, local hire programmes, and joint ventures with regional firms should be on the 12-month action list.

6. HEDGE WITH UTILITY AND RENEWABLE WORK

Several firms are aggressively rotating their utility-renewables backlogs toward higher-margin data-centre work. Inside Scenario A and B that is rational. Inside Scenario C the consequence is catastrophic — utility renewables backlogs cannot be easily rebuilt if hyperscaler demand softens. Maintaining diversified backlog discipline is the most important risk-management decision in front of engineering firm boards right now.

7. POSITION FOR THE NUCLEAR RENAISSANCE

Whether or not your firm has historic nuclear practice, the engineering services opportunity inside the 9.8 GW of announced hyperscaler nuclear deals is large and durable. SMR design support, balance-of-plant engineering, grid interconnection, site licensing engineering and construction management are all available to firms willing to build credibility. Sargent & Lundy, Bechtel and a small group of established nuclear engineering firms have first-mover advantage; the second tier is genuinely open.

8. PRE-POSITION FOR WATER

Water is the next-but-one binding constraint. Engineering firms with credible water-engineering practice — closed-loop cooling, recycled water systems, dry cooling at scale, water-rights engineering — will be the differentiated providers when state regulators (starting with New York and California) begin imposing data-centre water constraints. Build the practice now, before the binding regulatory event.

Methodology & Sources

This report was prepared from publicly available sources between February and May 2026. We have privileged five categories of source: (1) primary government documents and statements, notably from the US Energy Information Administration and the Department of Energy; (2) industry trade publications with strong reputations for engineering coverage (Engineering News-Record, Data Center Dynamics, Construction Dive, Utility Dive); (3) major financial newspapers and broadcasters (CNBC, Reuters, Bloomberg); (4) think tanks with track records on energy and infrastructure analysis (Belfer Center, World Resources Institute, ITIF); (5) hyperscaler investor disclosures and conference call transcripts.

All numbers are footnoted. Where ranges are given, we have used the highest-confidence range available. Where forecasts are quoted, we have identified the forecaster. Scenario probability assignments are editorial judgements anchored to consensus forecaster ranges; readers should treat them as ordinally informative rather than as precise probability estimates.

Sources cited

1. Yahoo Finance / Reuters, "Hyperscalers Hit $700 Billion in 2026 AI Spending Plans." February 2026.
2. Tom's Hardware, "Google, Microsoft, Meta, and Amazon capex spending to hit $725 billion in 2026, up 77% from last year." February 2026.
3. Futurum Group, "AI Capex 2026: The $690B Infrastructure Sprint."
4. CNBC, "Tech AI spending approaches $700 billion in 2026, cash taking big hit." 6 February 2026.
5. Fortune, "Big Tech's $700 billion AI spending spree has no clear end in sight." 30 April 2026.
6. S&P Global, "Data center grid-power demand to rise 22% in 2025, nearly triple by 2030." October 2025.
7. Utility Dive, "U.S. data center power demand could reach 106 GW by 2035: BloombergNEF."
8. US Energy Information Administration, "Commercial electricity sales have soared in Virginia, driven by data centers." 2026.
9. US Energy Information Administration, "Fossil generation could rise with faster-than-expected growth in data center power demand."
10. Belfer Center, "Data Centers and Large-Scale Electric Growth: The Virginia and Texas Experiences."
11. Daily Energy Insider, "EIA: Virginia power demand surges as data centers drive record load growth."
12. WRI, "Powering the US Data Center Boom: The Challenge of Forecasting Electricity Needs."
13. Privette et al., "Data Centers Water Footprint: The Need for More Transparency." AGU Advances, 2026.
14. Environmental Law Institute, "AI's Cooling Problem: How Data Centers Are Transforming Water Use."
15. BDO, "Data Center Dangers: Environmental Challenges Reshaping the Industry."
16. Colorado Newsline, "Data center water use not sustainable in increasingly dry American West." 8 April 2026.
17. SMRintel.com, "Every Nuclear-Powered Data Center Deal: Google, Amazon, Meta & Microsoft." 2026.
18. Data Center Dynamics, "Three Mile Island nuclear power plant to return as Microsoft signs 20-year, 835MW AI data center PPA."
19. Pennsylvania Capital-Star, "Microsoft describes Three Mile Island plant as a once-in-a-lifetime opportunity."
20. LumiChats, "Microsoft Reopened Three Mile Island to Power AI — Here's Why."
21. iRecruit, "SMR Data Centers: How Small Modular Reactors Power AI Infrastructure (2026)."
22. iRecruit, "Data Center Construction Labor Market Report — 499K-Worker Shortage."

23. ITIF, "Construction Industry Facing a 439,000-Worker Shortage Driven by the Growth of Data Centers." 12 January 2026.
24. Construction Dive, "How contractors can navigate cost pressures, labor shortages and regulatory hurdles in 2026."
25. Data Center Dynamics, "Why a construction worker shortage could hamper the US data center build-out."
26. Roofing Contractor, "Data Centers Lift Construction Outlook, but Economic Worries Grow."
27. Deloitte Insights, "2026 Engineering and Construction Industry Outlook."
28. Birm Group, "Data Center Construction Hiring 2026: Pay Pressure, Delays, Risk."
29. Epoch AI, "OpenAI Stargate: where the US sites stand."
30. Data Center Dynamics, "Crusoe tops out final building at OpenAI Stargate data center campus in Abilene, Texas."
31. OpenAI, "Stargate advances with 4.5 GW partnership with Oracle."
32. OpenAI, "Five new Stargate sites — expansion with Oracle and SoftBank."
33. CNBC, "OpenAI's first data center in $500 billion Stargate project is open in Texas." September 2025.
34. Interesting Engineering, "OpenAI, Oracle abandon 2 GW AI data center expansion in Texas."
35. Latitude Media, "Early-stage data centers are driving up US power demand forecasts."
36. NZero, "U.S. Power Demand Hits New Highs Driven by Data Centers, AI, and Grid Constraints."
37. Belfer Center, "AI, Data Centers, and the U.S. Electric Grid: A Watershed Moment."
38. CreditSights, "Technology: Hyperscaler Capex 2026 Estimates."
39. Next Waves Insight, "Hyperscaler Capex 2026: Where the $300 Billion in AI Infrastructure Is Actually Going."
40. IEEE ComSoc Technology Blog, "Hyperscaler capex > $600 bn in 2026 a 36% increase over 2025." December 2025.

Footnotes & Citations

1 Yahoo Finance / Reuters, "Hyperscalers Hit $700 Billion in 2026 AI Spending Plans." February 2026.

2 Tom's Hardware, "Google, Microsoft, Meta, and Amazon capex spending to hit $725 billion in 2026, up 77% from last year." February 2026.

3 Futurum Group, "AI Capex 2026: The $690B Infrastructure Sprint."

4 CNBC, "Tech AI spending approaches $700 billion in 2026, cash taking big hit." 6 February 2026.

5 Fortune, "Big Tech's $700 billion AI spending spree has no clear end in sight." 30 April 2026.

6 S&P Global, "Data center grid-power demand to rise 22% in 2025, nearly triple by 2030." October 2025.

7 Utility Dive, "U.S. data center power demand could reach 106 GW by 2035: BloombergNEF."

8 US Energy Information Administration, "Commercial electricity sales have soared in Virginia, driven by data centers." 2026.

9 US Energy Information Administration, "Fossil generation could rise with faster-than-expected growth in data center power demand."

10 Belfer Center, "Data Centers and Large-Scale Electric Growth: The Virginia and Texas Experiences."

11 Daily Energy Insider, "EIA: Virginia power demand surges as data centers drive record load growth."

12 WRI, "Powering the US Data Center Boom: The Challenge of Forecasting Electricity Needs."

13 Privette et al., "Data Centers Water Footprint: The Need for More Transparency." AGU Advances, 2026.

14 Environmental Law Institute, "AI's Cooling Problem: How Data Centers Are Transforming Water Use."

15 BDO, "Data Center Dangers: Environmental Challenges Reshaping the Industry."

16 Colorado Newsline, "Data center water use not sustainable in increasingly dry American West." 8 April 2026.

17 SMRintel.com, "Every Nuclear-Powered Data Center Deal: Google, Amazon, Meta & Microsoft." 2026.

18 Data Center Dynamics, "Three Mile Island nuclear power plant to return as Microsoft signs 20-year, 835MW AI data center PPA."

19 Pennsylvania Capital-Star, "Microsoft describes Three Mile Island plant as a once-in-a-lifetime opportunity."

20 LumiChats, "Microsoft Reopened Three Mile Island to Power AI — Here's Why."

21 iRecruit, "SMR Data Centers: How Small Modular Reactors Power AI Infrastructure (2026)."

22 iRecruit, "Data Center Construction Labor Market Report — 499K-Worker Shortage."

23 ITIF, "Construction Industry Facing a 439,000-Worker Shortage Driven by the Growth of Data Centers." 12 January 2026.

24 Construction Dive, "How contractors can navigate cost pressures, labor shortages and regulatory hurdles in 2026."

25 Data Center Dynamics, "Why a construction worker shortage could hamper the US data center build-out."

26 Roofing Contractor, "Data Centers Lift Construction Outlook, but Economic Worries Grow."

27 Deloitte Insights, "2026 Engineering and Construction Industry Outlook."

28 Birm Group, "Data Center Construction Hiring 2026: Pay Pressure, Delays, Risk."

29 Epoch AI, "OpenAI Stargate: where the US sites stand."

30 Data Center Dynamics, "Crusoe tops out final building at OpenAI Stargate data center campus in Abilene, Texas."

31 OpenAI, "Stargate advances with 4.5 GW partnership with Oracle."

32 OpenAI, "Five new Stargate sites — expansion with Oracle and SoftBank."

33 CNBC, "OpenAI's first data center in $500 billion Stargate project is open in Texas." September 2025.

34 Interesting Engineering, "OpenAI, Oracle abandon 2 GW AI data center expansion in Texas."

35 Latitude Media, "Early-stage data centers are driving up US power demand forecasts."

36 NZero, "U.S. Power Demand Hits New Highs Driven by Data Centers, AI, and Grid Constraints."

37 Belfer Center, "AI, Data Centers, and the U.S. Electric Grid: A Watershed Moment."

38 CreditSights, "Technology: Hyperscaler Capex 2026 Estimates."

39 Next Waves Insight, "Hyperscaler Capex 2026: Where the $300 Billion in AI Infrastructure Is Actually Going."

40 IEEE ComSoc Technology Blog, "Hyperscaler capex > $600 bn in 2026 a 36% increase over 2025." December 2025.