

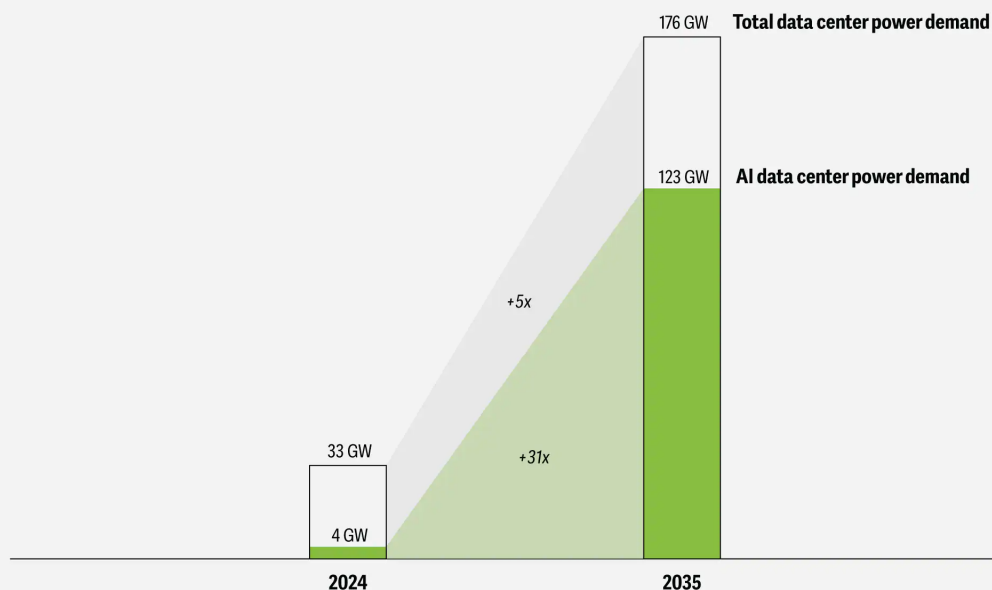
Can US infrastructure keep up with the AI economy?

The AI age is expected to require scaling data centers, grid capacity, and supply chains. Deloitte's 2025 AI Infrastructure Survey examines gaps and solutions.

The scale of AI data centers and their commensurate power needs are growing exponentially. By 2035, Deloitte estimates that power demand from AI data centers in the United States could grow more than thirtyfold, reaching 123 gigawatts, up from 4 gigawatts in 2024 (figure 1).¹ AI data centers can require dramatically more energy per square foot than traditional data centers. For example, a five-acre data center augmenting central processing units with specialized graphics processing units might see its energy usage increase from 5 to 50 megawatts.²

Figure 1

US power demand from AI data centers is expected to boom



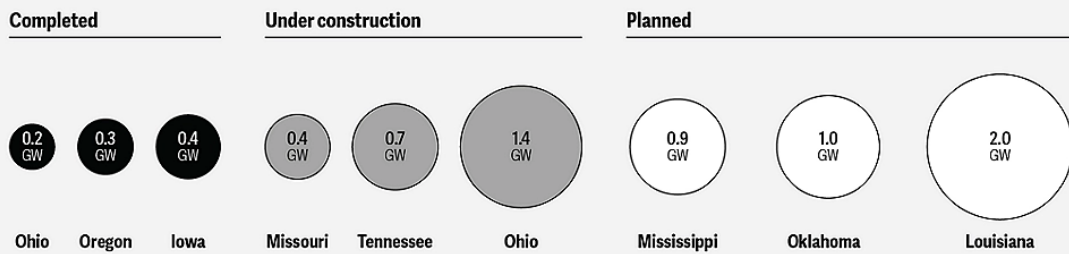
Source: Deloitte analysis of data from DC Byte, Wood Mackenzie, S&P Global, Lawrence Berkeley National Laboratory, Center for Strategic and International Studies, and Wells Fargo.

The leading AI infrastructure developers that are scaling data center networks globally are known as hyperscalers. Each of the top three hyperscalers' largest US data centers currently draw less than 500 megawatts (MW) of power but the largest data centers they are constructing or planning to build are more than double to quadruple the capacities of completed projects. The largest of these are expected to require up to 2,000 MW—that is, 2 gigawatts (GW) (figure 2).

Figure 2

New data centers will need gigawatts of power

Size and location of largest data center in each status category for each of the top hyperscalers with the largest data center capacities in the United States



Notes: Under construction project capacities may also include capacities in completed, planned, and early-stage phases. Planned project capacities may also include capacities in the early-stage phase. Bubble size represents 1 GW of IT capacity—that is, the net power capacity that can be provided to computer servers and their supporting mechanical and electrical equipment.

Source: Deloitte analysis of DC Byte data as of June 2025.

Deloitte Insights | deloitteinsights.com

However, even these capacities are modest compared to what is on its way. There are 50,000-acre data center campuses in early-stage phases, which could consume 5 GW³—the amount of power needed for five million residential homes, and more than the capacity of the largest existing nuclear or gas plants in the United States.⁴

AI data centers uniquely challenge grid operations because they can create large, concentrated clusters of 24/7 power demand. Some of the leading data center growth regions have experienced harmonic distortions, load relief warnings and near-miss incidents, and generation shutdowns.⁵

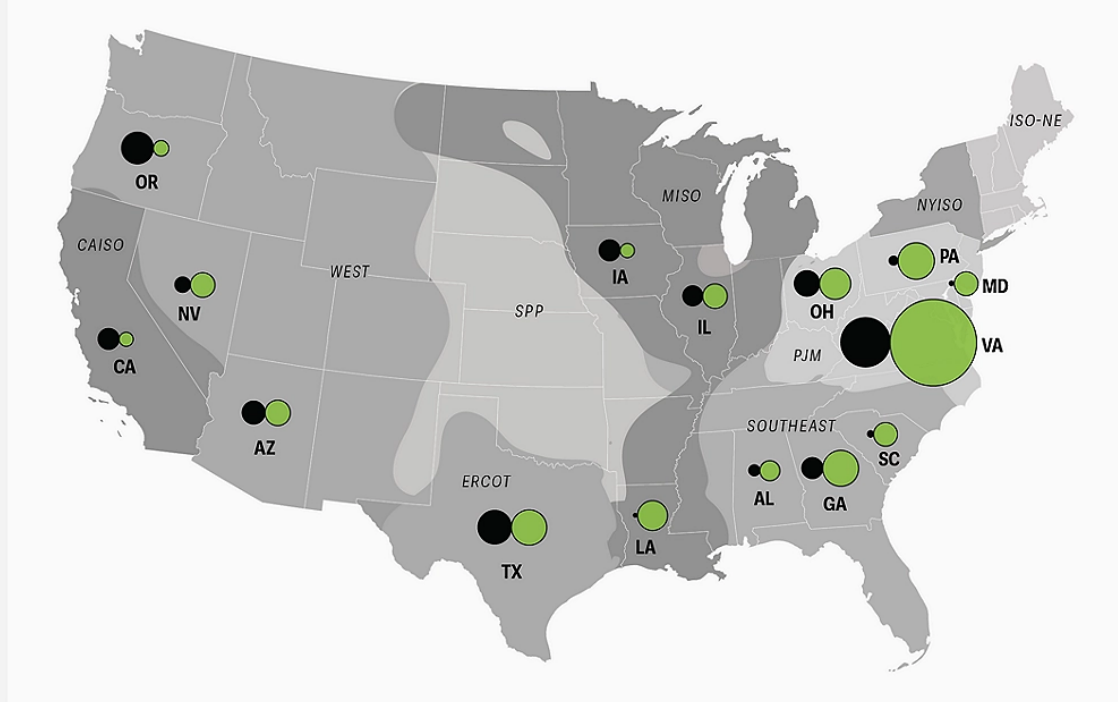
At the same time, the constellation of data center markets is changing. AI infrastructure development is spreading to more states and decentralizing as AI models are deployed closer to users to provide faster responses (figure 3).

Figure 3

Data center markets with existing or expected gigawatt capacity

Data center capacity (GW) and RTO/ISO regions

	VA	TX	OR	OH	AZ	CA	GA	IA	IL	NV	AL	PA	SC	MD	LA
Current capacity	6.0	2.9	2.3	1.6	1.3	1.1	1.1	1.1	1.0	0.6	0.2	0.2	0.1	<0.1	<0.1
Capacity under construction and planned	16.8	3.1	0.6	2.4	1.6	0.5	3.2	0.5	1.5	1.5	1.0	3.2	1.4	1.5	2.2



Notes: The map shows data center markets with at least a gigawatt in existing or planned capacity; circle size represents 1 GW of IT capacity—that is, the net power capacity that can be provided to computer servers and their supporting mechanical and electrical support equipment. CAISO: California Independent System Operator; ERCOT: Electric Reliability Council of Texas; ISO-NE: Independent System Operator New England; MISO: Midcontinent Independent System Operator; NYISO: New York Independent System Operator; PJM: PJM Interconnection; SPP: Southwest Power Pool.

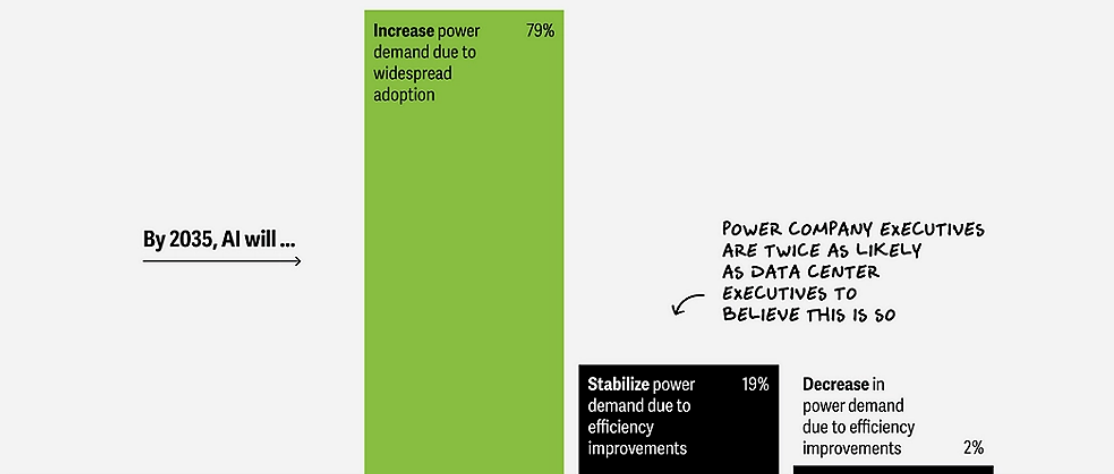
Source: Deloitte analysis of DC Byte data as of June 2025.

Deloitte
Insights | deloitteinsights.com

Indeed, an April 2025 Deloitte survey of 120 US-based power company and data center executives shows that grid stress was the leading challenge for data center infrastructure development (see [methodology](#)). Seventy-nine percent of respondents said that AI will increase power demand through 2035 due to widespread adoption (figure 4).⁶

Figure 4

How power company and data center executives think AI will impact 2035 power demand



Notes: n = 98. This reflects only those respondents who selected that AI will impact power demand growth.

Source: Deloitte 2025 AI Infrastructure Survey.

Deloitte Insights | deloitteinsights.com

This is where the AI ambitions of the US government and industry come up against the grid's capacity to power or to even interconnect data centers: There's currently a seven-year wait on some requests for connection to the grid.⁷

There are other challenges, too.

The forecast ranges for AI data center demand vary widely, complicating efforts to build new generation while insulating residential customers from rate increases. Supply chain issues are also constraining energy companies and hyperscalers as manufacturers face sharp increases in demand for key infrastructure components. The overall uncertainty increases the risk of overbuilding or missing growth opportunities due to the inability to serve new customers.

To explore challenges such as these, and to identify opportunities and strategic approaches to building data center infrastructure, the Deloitte Center for Energy and Industrials conducted an AI infrastructure survey of executives from US-based data center and power companies. The results also support a set of strategic recommendations that can work together to close these gaps in the nation's critical AI infrastructure.

7 gaps to powering AI

Deloitte's 2025 AI Infrastructure Survey included questions on infrastructure build-out challenges, resource mix to meet future energy consumption, workforce issues, AI workload planning, drivers of load growth, and investment priorities. The results were then divided based on respondent group to elucidate the differences between power companies and data center operators. The findings were further supplemented with secondary research and interviews with industry experts.

The results demonstrate that the primary challenge for data center infrastructure build-out is power and grid capacity, which 72% of all respondents consider to be very or extremely challenging (figure 5). In addition, companies surveyed express concern about supply chain disruptions (65%) and security (64%).

Figure 5

Top challenges among respondents for data center infrastructure development

Question: What does your organization believe are the greatest challenges to data center infrastructure build-out?

	Data centers	Power companies
1	Power and grid capacity constraints in meeting load growth	
2	Data center and grid build-out timeline mismatches	Supply chain disruptions affecting equipment and materials
3	Cyber and/or physical security concerns	
4	Supply chain disruptions affecting equipment and materials	Difficulty in securing permits
5	Difficulty in securing permits	Data center and grid build-out timeline mismatches
6	Workforce constraints	Gas (un)deliverability
7	Gas (un)deliverability	Workforce constraints

Note: Ranking is based on the number of respondents who selected "very" or "extremely" challenging.

Source: Deloitte 2025 AI Infrastructure Survey.

Deloitte Insights | deloitteinsights.com

These constraints cited by survey respondents highlight seven key gaps in the build-out of data center infrastructure.

Gap 1: Peak demand is spiking as base load generation capacity contracts

At the same time as data centers are pushing up peak demand, baseload generation is contracting, while new generation projects are stuck in increasingly long interconnection queues, 95% of which consist of renewables and storage (figure 6).⁸

Figure 6

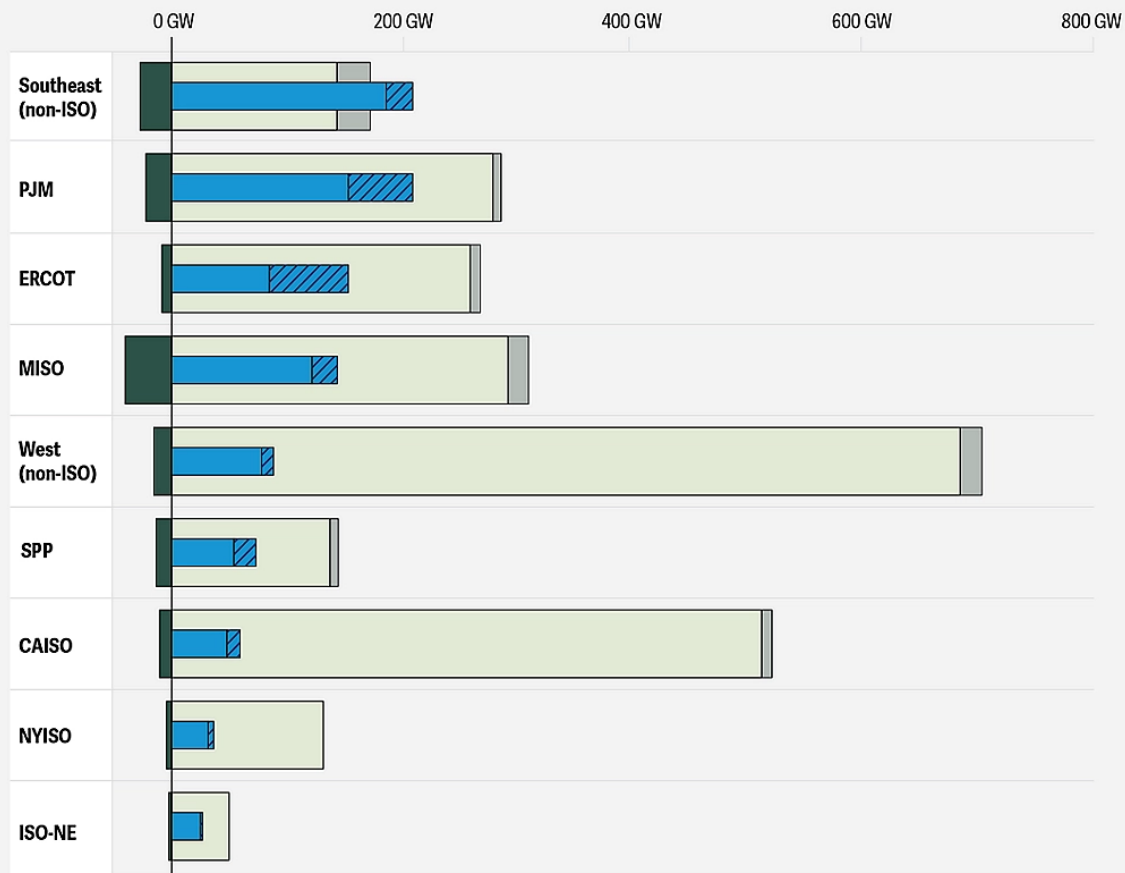
Supply struggles to keep up with rising demand

Demand

● Peak demand (2024) ● Estimated growth in peak demand by 2035

Supply

● Coal and gas retirements (2025–2035) ● Carbon-free energy in interconnection queue (2024) ● Fossil energy in interconnection queue (2024)



Notes: Interconnection queue estimates reflect the queue as of the end of 2023; PJM: PJM Interconnection; MISO: Midcontinent Independent System Operator; ERCOT: Electric Reliability Council of Texas; SPP: Southwest Power Pool; CAISO: California Independent System Operator; NYISO: New York Independent System Operator; ISO-NE: Independent System Operator New England.

Source: Deloitte analysis of data from Energy Markets & Policy Berkeley Lab, "Queued up: 2024 edition, characteristics of power plants seeking transmission interconnection as of the end of 2023," April 2024; The American Clean Power Association, "US national power demand study," March 7, 2025; and ISO websites and latest forecasts as of June 2025.

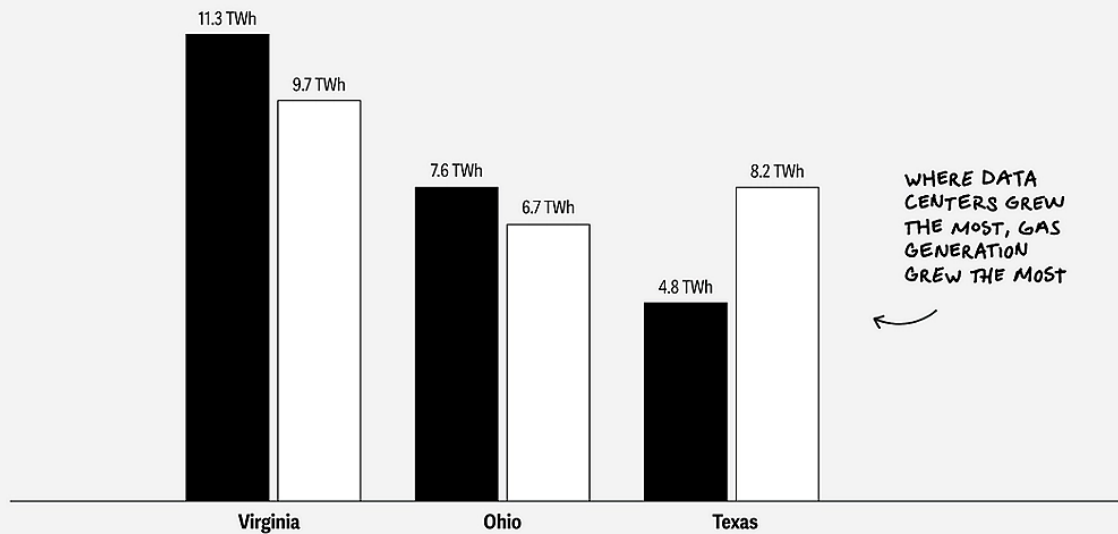
Deloitte Insights | deloitteinsights.com

These stuck projects are key to meeting the demand of hyperscalers, which have set clean energy goals and are the leading buyers of renewable power purchase agreements.⁹ Despite these targets, load growth over the past year in the top markets for data center growth has primarily been met with increased gas generation (figure 7).

Figure 7

Despite clean energy goals, AI growth is fueled by gas

- Growth in data center power consumption among the top 3 states with the largest increases, 2023 to 2024
- Growth in gas generation among the top 3 states with the largest increases, 2023 to 2024



Source: Deloitte analysis of DC Byte, Ember, and Energy Information Administration data.

Deloitte Insights | deloitteinsights.com

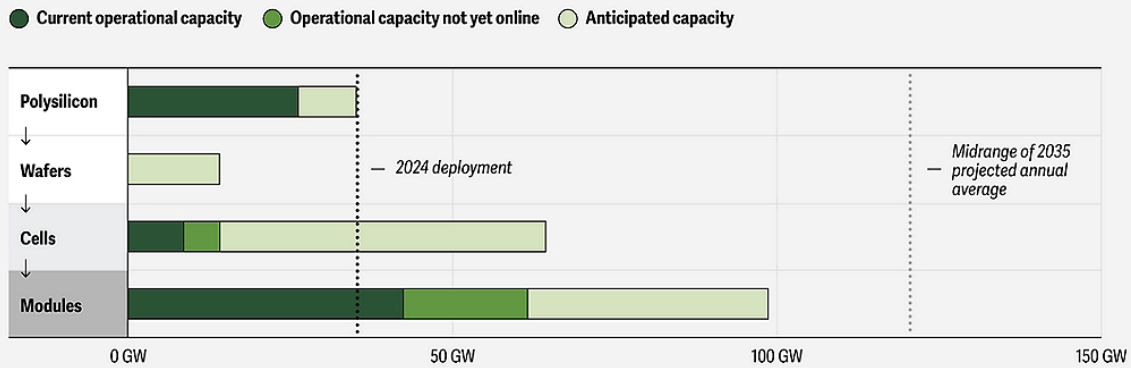
Gap 2: Supply chain disruptions are complicating project plans

The second greatest challenge among power company respondents is supply chain disruptions. Although the United States has enough manufacturing capacity to meet domestic demand for solar modules, demand is rising—and critical components are still imported and subject to tariffs (figure 8). Potential cost increases of steel, aluminum, copper, timber, and cement could also impact the build-out of power and gas infrastructure in the context of rising construction material costs, which have already jumped 40% over the past five years.¹⁰

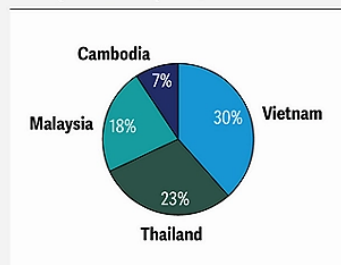
Figure 8

To meet future demand, the solar industry should ramp up domestic production

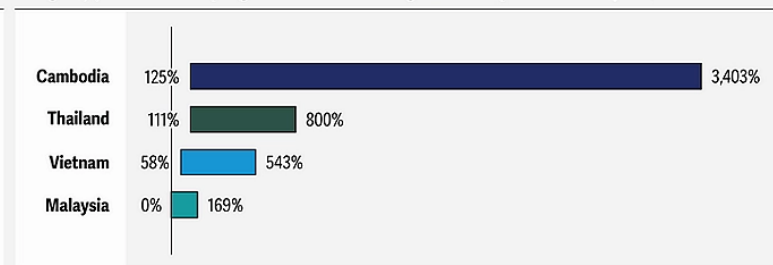
Domestic solar manufacturing capacity



Solar product imports, 2024



Range of final antidumping and countervailing duties imposed on May 20, 2025



Source: Deloitte analysis of data from Rhodium Group and MIT CEEPR, Clean Investment Monitor, "The state of US clean energy supply chains in 2025"; and US International Trade Commission and Department of Commerce.

Deloitte Insights | deloitteinsights.com

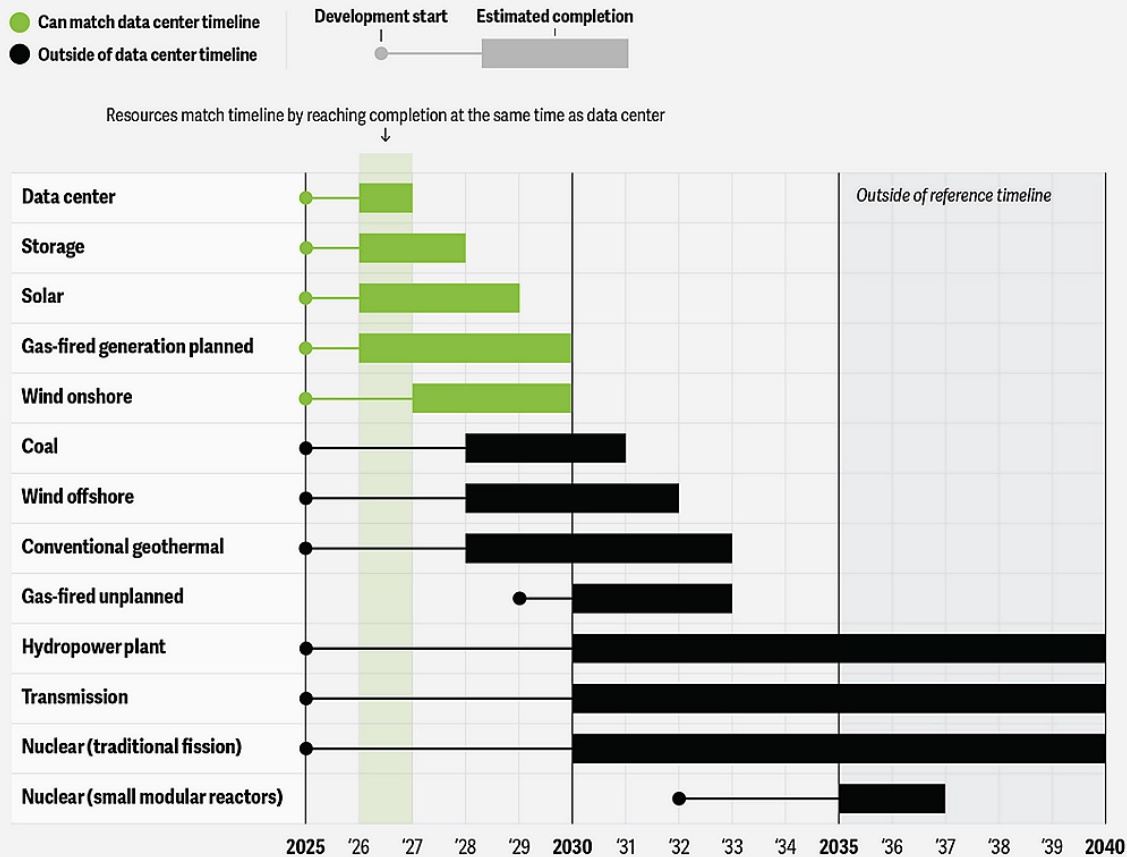
Gap 3: Long—and growing—grid build-out timelines

Supply chain issues intersect with a top-3 challenge for the surveyed data center companies: Most power capacity development can take longer than data center build-outs, which can be completed in a year or two (figure 9). Gas power plant projects that have not already contracted equipment, for example, are not expected to become available until the 2030s.¹¹

Renewables and battery storage can match this timeline and account for 92% of all planned generating capacity additions to the grid in 2025.¹² But in some cases, the transmission needed to bring renewable capacity to load can take over a decade to build, and space constraints can limit the build-out of sufficient capacity next to data centers.

Figure 9

Few energy resources align with data center timelines



Source: Deloitte analysis of data from International Energy Agency, Brattle, and Energy Innovation reports.

Deloitte
Insights | deloitteinsights.com

Gap 4: Cyber and power security are growing concerns

As AI capabilities grow, data centers should be secured against hackers who could subvert an AI model by gaining access to its weights, which govern model training and output. AI data centers can be especially vulnerable to supply chain attacks.¹³ Points of entry for cyber infringement can include servers, storage systems, cooling equipment, network gear, and other digital equipment that are shipped from multiple countries.¹⁴ Security of power supply is another concern for data center operators, as back-up generators have limited capacity.

Gap 5: The permitting process can be long and unpredictable

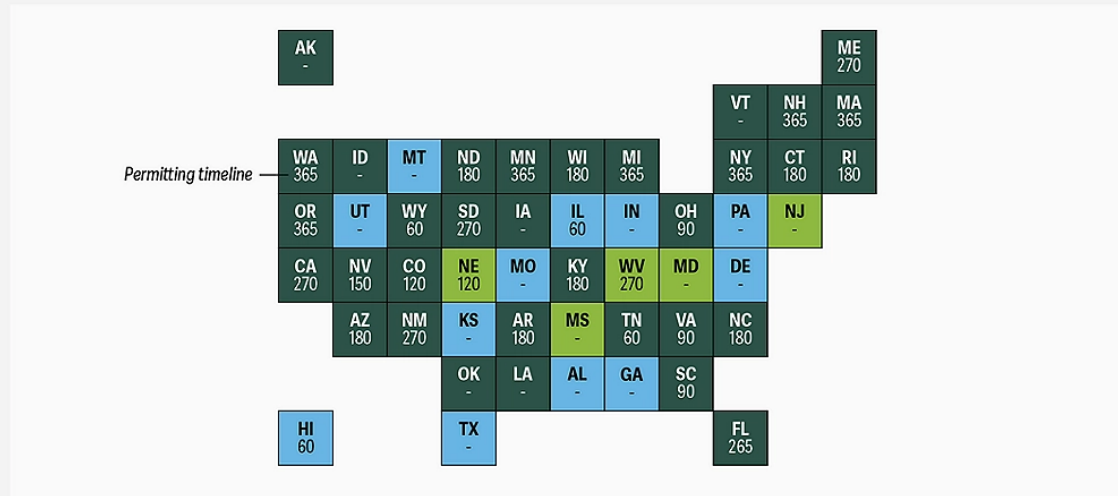
The difficulty of securing permits and their varying timelines—ranging from months to years—can impact project schedules and inflate costs. It still takes more than two years to complete an environmental impact statement.¹⁵ The locus of authority for permitting mostly sits with the states, and over the past year, state-level restrictions have more than doubled overall (figure 10).¹⁶ Contested projects have grown 29%, and local restrictions on renewable projects have grown by 73%.¹⁷

Figure 10

Local restrictions and contested projects compound long permitting timelines

Principal authority for siting and permitting process and states with permitting timelines for renewable projects (in days)

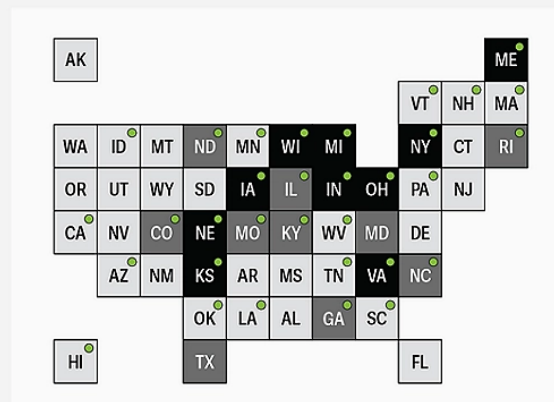
Local State Mixed



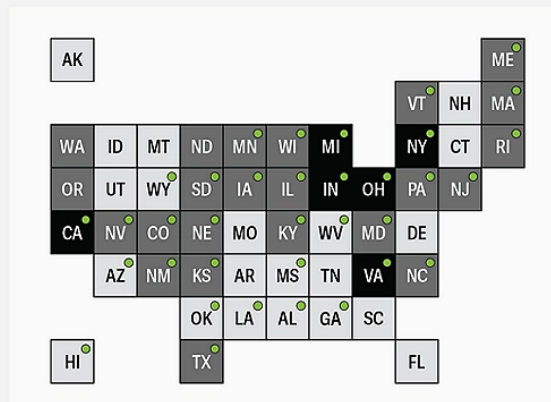
Number of local restrictions on renewables and contested renewable projects

Low (<5) Moderate (5–15) High (>15) Trending upward (2023 to 2024)

Local and state restrictions



Contested projects



Notes: "Mixed" means that either both the state and local levels have some authority, or one has principal authority—depending on the size of the project in nearly all cases; In states with multiple permitting processes, the timeline reflects the longest process.

Source: Deloitte analysis of "Laws in order 2024" report by Regulatory Assistance Project and "Opposition to renewable energy facilities in the United States," 2023 and 2024 editions, by Columbia Law School Sabin Center for Climate Change Law.

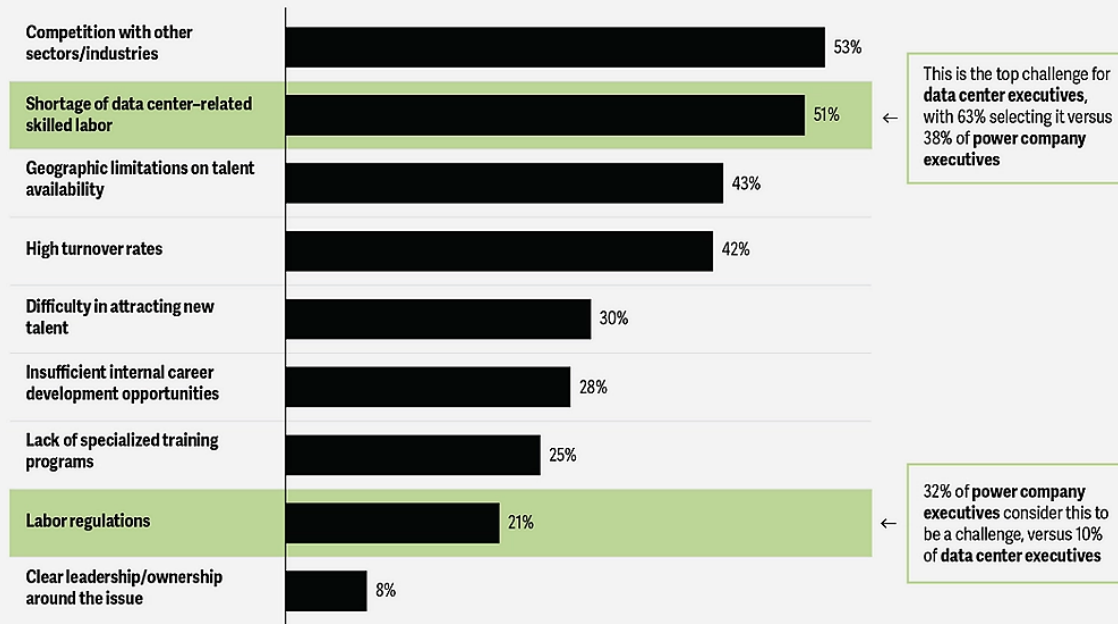
Deloitte Insights | deloitteinsights.com

Gap 6: The industry needs more skilled workers

Most respondents face the related challenges of competition with other industries and a shortage of skilled labor (figure 11). The latter is particularly acute for data center respondents, 63% of whom view it as the top challenge. For respondents from power companies, high turnover rates are the second greatest challenge.

Figure 11

Competition for skilled labor is a core challenge



Note: Question: Which of the following workforce challenges is your organization facing specific to data center and energy infrastructure expansion?

Source: Deloitte 2025 AI Infrastructure Survey.

Deloitte Insights | deloitteinsights.com

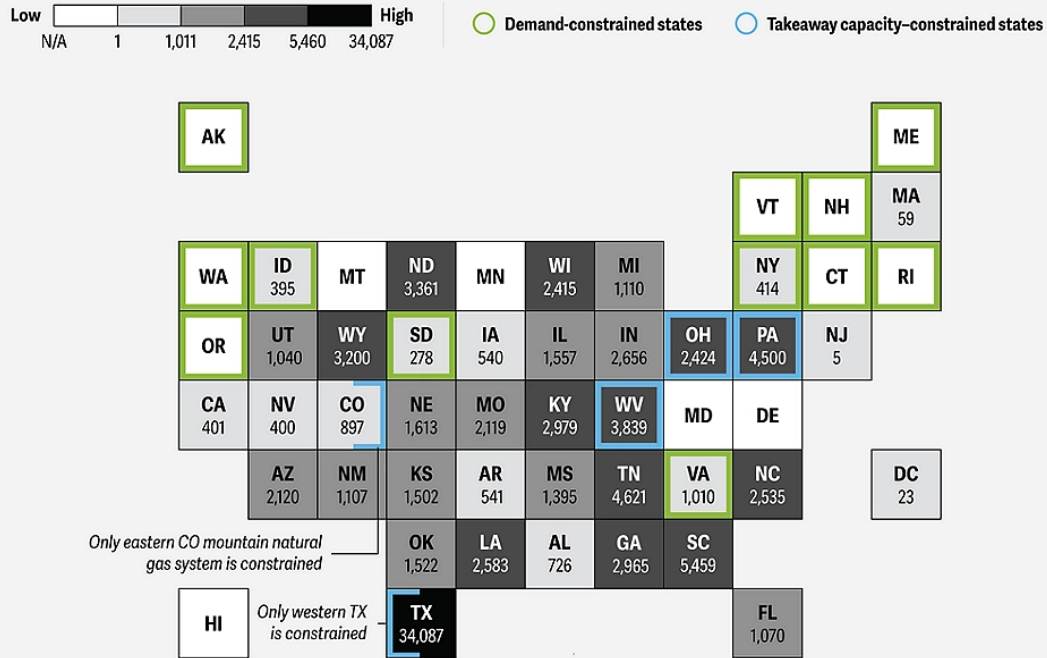
Gap 7: Limited pipeline capacity stymies natural gas delivery

Industry estimates of data center gas demand through 2030 range from 3 billion to 12 billion cubic feet per day¹⁸ amid a growing pipeline of planned natural gas plants. More than 99 GW of gas-fired capacity is planned across 38 states; yet, many top data center markets have constrained gas pipeline takeaway and transportation capacity (figure 12).

Figure 12

Gas infrastructure build-out and constraints

Natural gas power plants planned through 2035 (in MW)



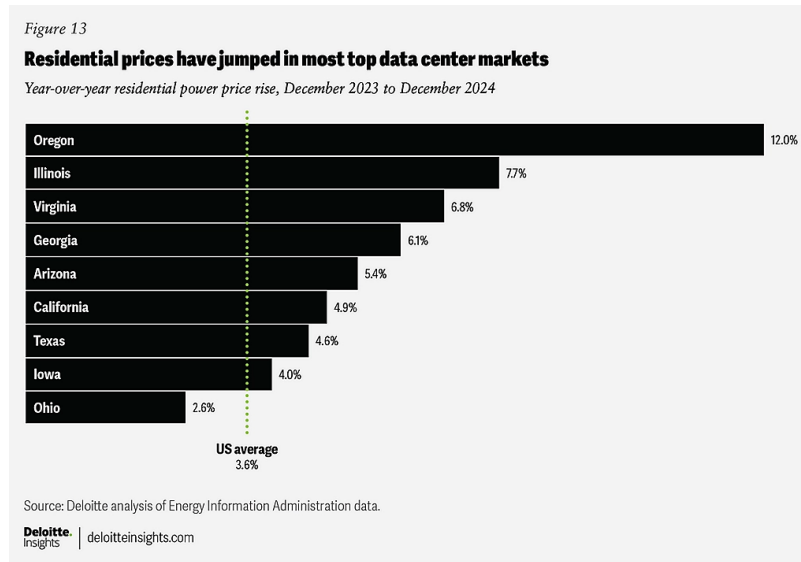
Note: Pipeline constraints are often very regional.

Source: Deloitte analysis of data from S&P Global, the US Energy Information Administration, and Congressional Research Service.

Different experiences of shared challenges

Despite shared industry challenges, data center and power companies surveyed diverge on how they experience these gaps and the relative importance of other issues.

- **Affordability crunch:** In theory, growth in electricity demand and an expansion in the system could spread the cost of grid investment across more customers, putting downward pressure on rates. However, the scale of the upfront grid investments required to interconnect AI data center loads has pushed some of the cost burden onto residential customers.¹⁹ Over the past year, residential power prices increased more than the national average in 8 of the 9 top data center markets (figure 13).



While 60% of power companies surveyed believe residential affordability concerns will not constrain data center development, only 47% of data center operators share this view.

- **Competition for resources:** Most respondents are experiencing resource competition, but the issue is more acute for data centers (80%) than power companies (57%). Moreover, 92% of data centers see power capacity as a locus of resource competition, versus 71% of power companies.
- **Operational inflexibility:** Data center operators and power companies are misaligned on the potential for flexibility that could help ease grid stress. More than half of data centers (57%) anticipate AI training and inference computing workloads will be managed differently, versus only 38% of power companies.

Strategies for closing the AI infrastructure gaps

To close these gaps, the United States should develop additive infrastructure that can deliver efficiency, capacity, and flexibility. According to survey respondents, the most important strategies to overcome these challenges are technological innovation, regulatory changes, and more funding (figure 14).

Figure 14

Top strategies among respondents for overcoming infrastructure gaps

Question: What do you think can help in overcoming data center infrastructure build-out challenges?

	Data centers	Power companies
1	Technological innovation	Regulatory changes
2	Regulatory changes	Technological innovation
3	More funding	
4	Computational task mobility	New business models
5	Legislative enablement	Computational task mobility
6	New business models	Legislative enablement
7	Data center resource reallocation	

Note: Ranking is based on the number of respondents who selected "very" or "extremely" challenging.

Source: Deloitte 2025 AI Infrastructure Survey.

Deloitte Insights | deloitteinsights.com

Strategy 1: Technological innovation can support efficiency

According to 82% of those surveyed by Deloitte, a highly impactful means of closing the gaps is technological innovation, which can help build additive infrastructure by driving efficiency. This is also the top investment priority for 90% of data center respondents, followed by making infrastructure more intelligent (81%) and integrating renewable energy sources (66%).

Innovation in AI data center, chip, and grid efficiencies includes:

1. **Tapping new resources to cool data centers:** Cooling can account for 40% of data center electricity demand, and AI data centers are especially heat-intensive.²⁰ The use of liquid cooling systems can reduce power demand, but they require a lot of water—which can be a scarce resource in some of the current and prospective gigawatt data center markets. To help overcome this, data centers could install rooftop rainwater collection on-site, which could meet up to a third of cooling needs.²¹ Another opportunity is the use of geothermal energy, which can offer direct data center cooling, thus reducing both energy and water demand. If data centers are located at the most suitable sites to develop enhanced geothermal systems, these baseload resources could also meet significant AI data center demand growth by 2035.²²
2. **Transforming power delivery in chips:** As chip manufacturers reach the physical limits of efficiency gains from shrinking transistors and sizing up chips, they are looking to increase transistor count per chip by reconfiguring the chip's power. Moving power delivery from the front to the backside of chips could reduce power losses by 30%.²³ Encoding data in light instead of wires could enable optical data transmission at 10% the energy cost of electronic transmission.²⁴ And 3D stacking of chips using carbon-nanotube circuits could be 10 times more efficient.²⁵

Another innovation is integrating memory with processing within a transistor to more closely mimic simulations of electrical currents in the brain, which requires only 20 watts, compared to the billions required to power an AI

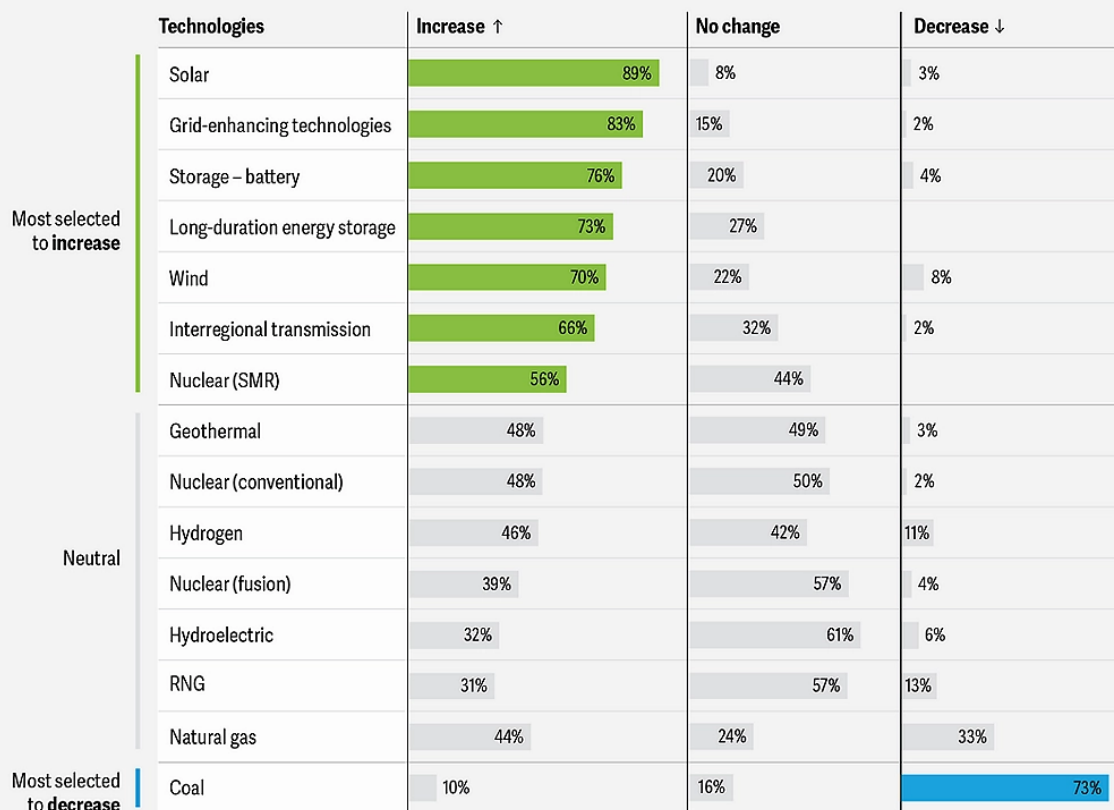
model.²⁶ Transistor efficiencies that can reduce the energy needs of a generative pre-trained transformer could help bend the rising demand curve for grid transformers.

3. **Transmitting power on the grid in new ways:** Power companies can move power more efficiently through transformer and transmission technologies that could also help address supply chain, timeline, and grid challenges. Solid-state transformers are more compact and enable greater standardization and communication with the grid, driving efficiency.²⁷ Some low-cost advanced transmission technologies can also increase the capacity of the existing transmission network within a timeline that matches that of data center development. Grid-enhancing technologies (GETs), including dynamic line rating, can be developed in three months and increase transmission capacity by 10% to 30%.²⁸ Flexible alternating current transmission systems and reconductoring can be developed in 8 to 18 months and increase capacity as much as 50% to 100%. According to the survey results, 83% of respondents expect GETs to play an increasing role in meeting data center energy demands through 2035, second only to solar (89%) (figure 15).

Figure 15

Solar and grid-enhancing technologies will be needed to meet 2035 data center demand

Question: How do you believe your resource mix and grid will evolve to meet data center energy consumption through 2035?



Notes: SMR: Small modular reactor; RNG: Renewable natural gas.

Source: Deloitte 2025 AI Infrastructure Survey.

Deloitte Insights | deloitteinsights.com

Divergent perspectives on innovation in the grid resource mix

The perspectives of the surveyed executives from data center and power companies diverge the most with respect to the prospects for conventional and fusion nuclear energy. Most data center respondents believe conventional nuclear will play a greater role in meeting data center power consumption through 2035, versus a third of power companies. And 38% of data center respondents see a greater role for nuclear fusion, versus 17% of power companies. Similarly, half of data centers anticipate natural gas will play a greater role, compared to 37% of power companies. The same share of power companies sees a greater role for renewable natural gas, versus only 17% of data centers.

Strategy 2: Innovate regulatory frameworks to build capacity

According to 76% of respondents in Deloitte's survey, regulatory change is highly impactful in addressing challenges to data center build-out. To add new capacity, the industry needs regulatory certainty built around system integration, stakeholder collaboration, and residential ratepayer protection.

1. **System integration:** The interconnections of generation and large loads could benefit from regulatory reform that can increase transparency, remove speculative projects, and advance priorities (figure 16).

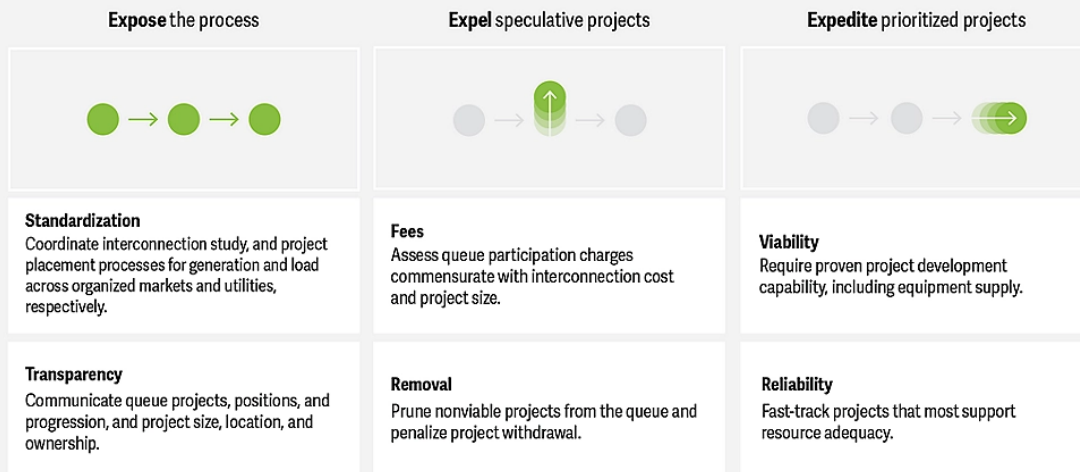
To help expedite generation interconnection, the Federal Energy Regulatory Commission's (FERC) Order 2023 kick-started reform by transitioning grid operators to "first-ready, first-served" cluster studies.²⁹ While the order's multiyear implementation is underway, several operators have taken more immediate action to address the risk of capacity shortfalls over the next year (see "[Initiatives to speed interconnection](#)")

The focus on the load side should be on flexibility. To help expedite data center interconnection, regulatory reform could prioritize data centers that temporarily curtail load during grid emergencies or participate in demand response. Electric Reliability Council of Texas (ERCOT) has a Controllable Load Resource program that can interconnect large loads within two years.³⁰ Grid operators could also offer expedited or joint interconnection to data centers that build their own on-site generation. New compensation mechanisms could incentivize these colocated project developers to invest in generation sized to exceed load and supply the grid during peak demand.

Load interconnection challenges have contributed to interest in bypassing interconnection queues and reducing transmission costs by colocating data centers with existing power plants. This type of colocation has raised broader regulatory questions about transmission service use, transmission cost calculation, and implications for reliability and customer rates. FERC and ERCOT rulemaking underway may help create more regulatory certainty to enable colocation-based business model innovation.

Figure 16

Deloitte framework for streamlining generation and load interconnection



Source: Deloitte analysis.

Deloitte Insights | deloitteinsights.com

Initiatives to speed interconnection

Grid operators are launching one-time mechanisms to bypass queues. These include PJM Interconnection's emergency Reliability Resource Initiative, which can expedite the interconnection of firm generation by 18 months to meet resource adequacy needs.³¹ Natural gas-fired generation accounted for more than two-thirds of the 11.8 GW of capacity that PJM selected from 94 new and uprated battery, gas, and nuclear applicants.³² PJM also removed Surplus Interconnection Service barriers that could enable an equal capacity to connect to its grid within the next two years.³³

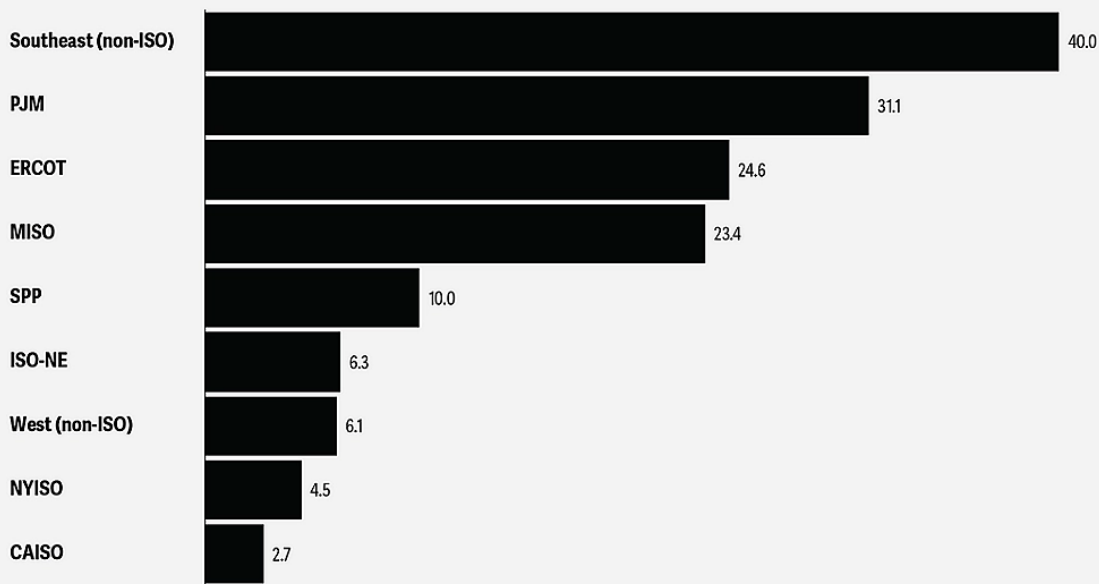
Meanwhile, Midcontinent Independent System Operator (MISO) and Southwest Power Pool (SPP) have developed Expedited Resource Adequacy Studies (ERAS) proposals for projects with commercial operation dates within the next two to three years.³⁴ MISO plans to consider these quarterly and grant interconnection agreements within 90 days.³⁵ SPP aims to further reduce study time and costs through consolidated generation, load, and transmission planning by 2027.³⁶

Beyond interconnection of generation and load, system integration can extend to regulatory certainty around interregional transmission. FERC Order 1920 laid the groundwork for transmission planning to meet load growth, including identification and evaluation of interregional transmission.³⁷ Establishing a minimum interregional transmission capability of up to 30% of a region's peak load could unlock 149 GW in capacity (figure 17).

Figure 17

More interregional transmission could help meet peak demand

Incremental interregional transmission needed to reach import capability up to 30% of peak load (GW)



Notes: PJM: PJM Interconnection; MISO: Midcontinent Independent System Operator; ERCOT: Electric Reliability Council of Texas; SPP: Southwest Power Pool; CAISO: California Independent System Operator; NYISO: New York Independent System Operator; ISO-NE: Independent System Operator New England.

Source: Energy Systems Integration Group, "Interregional transmission for resilience," June 2024.

Deloitte Insights | deloitteinsights.com

2. **Stakeholder cooperation:** Partnership between power companies, data centers, and grid operators can help advance the regulatory process. For example, in May 2025, a public utility commission approved the first clean transition tariff that a hyperscaler, a utility, and a renewable energy developer crafted together to jumpstart the deployment of new clean energy resources.³⁸ The largest US grid operator is also partnering with a hyperscaler in a groundbreaking use of AI to accelerate the interconnection process.³⁹ According to 78% of data center companies and 67% of power companies surveyed by Deloitte, cooperation between data center operators and utilities on infrastructure development and cost allocation is effective, but only 15% of data center executives and 8% of power company executives describe it as highly effective. The data center respondents see partnering with utilities to secure reliable grid power as the top strategy to meet the energy demand for their operations (63%).

3. **Residential ratepayer insulation:** Regulatory mechanisms can help firewall residential customers from costs associated with serving data center loads. These include creating a separate rate class for data centers and implementing tariffs with monthly demand and minimum charges based on contracted capacity, upfront payments, long contract terms, and exit fees.

Strategy 3: Flexibly scheduling computing tasks could reduce peak demand

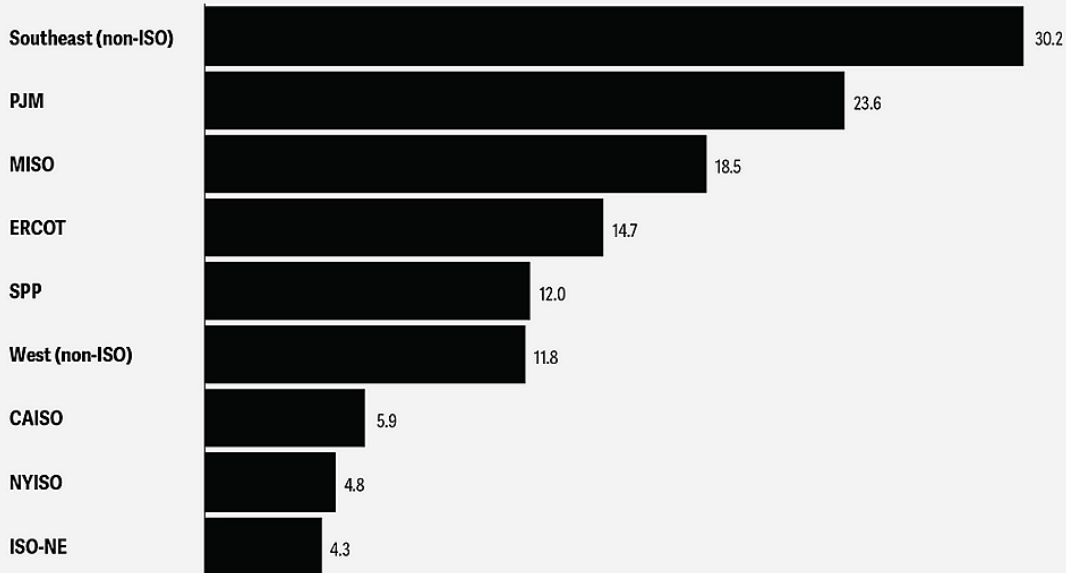
Computational task mobility can unlock a third aspect of additive infrastructure: flexibility. Large, inflexible loads can require major, costly, and time-consuming infrastructure upgrades because they add to the power system's peak demand. However, if data centers could operate with some flexibility during periods of peak demand, they could be considered grid assets and may be able to avoid the need for capacity addition upgrades altogether, thereby enabling rapid connection. Task mobility involves shifting flexible computing tasks to different times of day or data centers in response to grid conditions. If AI data centers agreed to curtail just 1% of their load, they could increase the grid's

ability to interconnect them. A study found that grid operators could add 126 GW in new load with minimal grid capacity expansion if average annual load was curtailed at a rate of 1% (figure 18). Among Deloitte’s survey respondents, 68% of industry executives believe data center demand flexibility will become an acceptable tradeoff to secure speed to market.

Figure 18

AI data centers with flexible operations could reduce peak demand

New load that can be added with minimal grid capacity expansion if average annual load is curtailed by 1% (GW)



Notes: PJM: PJM Interconnection; MISO: Midcontinent Independent System Operator; ERCOT: Electric Reliability Council of Texas; SPP: Southwest Power Pool; CAISO: California Independent System Operator; NYISO: New York Independent System Operator; ISO-NE: Independent System Operator New England.

Source: Tyler H. Norris, Tim Profeta, Dalia Patino-Echeverri, and Adam Cowie-Haskell, “Rethinking load growth: Assessing the potential for integration of large flexible loads in US power systems,” Nicholas Institute for Energy, Environment & Sustainability, February 2025.

Deloitte Insights | deloitteinsights.com

Strategy 4: Funding unlocks as investment stakes hit the trillions

Developing additive infrastructure will require massive funding across all of the industries involved. The impact of AI data center development is already becoming more apparent in investment discussions. Mentions of “data center” in investor earnings calls transcripts in both the manufacturing and energy sectors grew fivefold from 997 in 2023 to a record high 5,402 in 2024.⁴⁰ Data centers have thereby gained importance to long-term strategy and investments, opening the door to new funding opportunities. Among the power company and data center executives surveyed by Deloitte, 67% believe funding to be a key to closing AI infrastructure gaps. This is already well underway.

In 2024, private investment in US data centers reached a record high.⁴¹

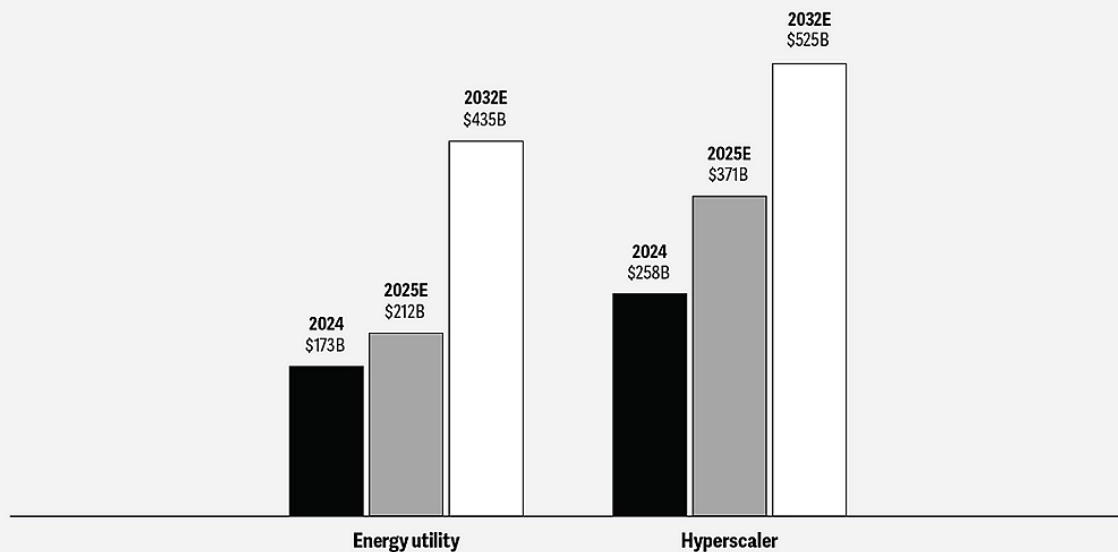
Meanwhile, electric and gas utilities are forecasting a record increase in capex. It is expected to jump 22% year over year to US\$212 billion in 2025 across 47 utilities—a sharp rise from the 7.6% CAGR over the past decade (Figure 19).⁴²

In 2024, capital light hyperscalers surpassed capital-intensive utilities in capex, a trend that is expected to continue.⁴³ Eight hyperscalers expect a 44% year-over-year increase to US\$371 billion in 2025 for AI data centers and computing resources.⁴⁴

Figure 19

US utility and hyperscaler annual capex to reach \$1 trillion by 2032

CAPEX (in US dollars)



Notes: The dollar numbers represented with E are estimates. Energy utility capex is based on 47 electric and gas utilities that form a representative sample of publicly traded, US-based utilities. Hyperscaler capex is based on eight hyperscalers.

Source: Deloitte analysis of S&P and Bloomberg data.

Deloitte Insights | deloitteinsights.com

The changing scope and scale of infrastructure development have likewise increased the investment stakes of building out capacity for data centers, power generation, and manufacturing to trillion-dollar levels.

Electric and gas utility capex is expected to surpass US\$1 trillion cumulatively within the next five years (2025–2029) for the 47 biggest investor-owned utilities. For the hyperscalers, reaching the trillion-dollar threshold is expected in only three years, with spending projections reaching half a trillion dollars annually by the early 2030s.⁴⁵ And for data centers more broadly, spending could reach a trillion dollars within three years.⁴⁶

The tech industry has also announced plans to invest more than a trillion dollars in US manufacturing of AI supercomputers, chips, and servers over the next four years.⁴⁷

The new level of funding reflects confidence that trillion-dollar AI applications and returns will materialize. Fears that cheaper models could undermine hyperscalers' business model have sparked market volatility. But they have not impacted hyperscaler AI chip purchases and capex forecasts.⁴⁸ Rather, hyperscalers have shifted investment from AI training towards inference.⁴⁹

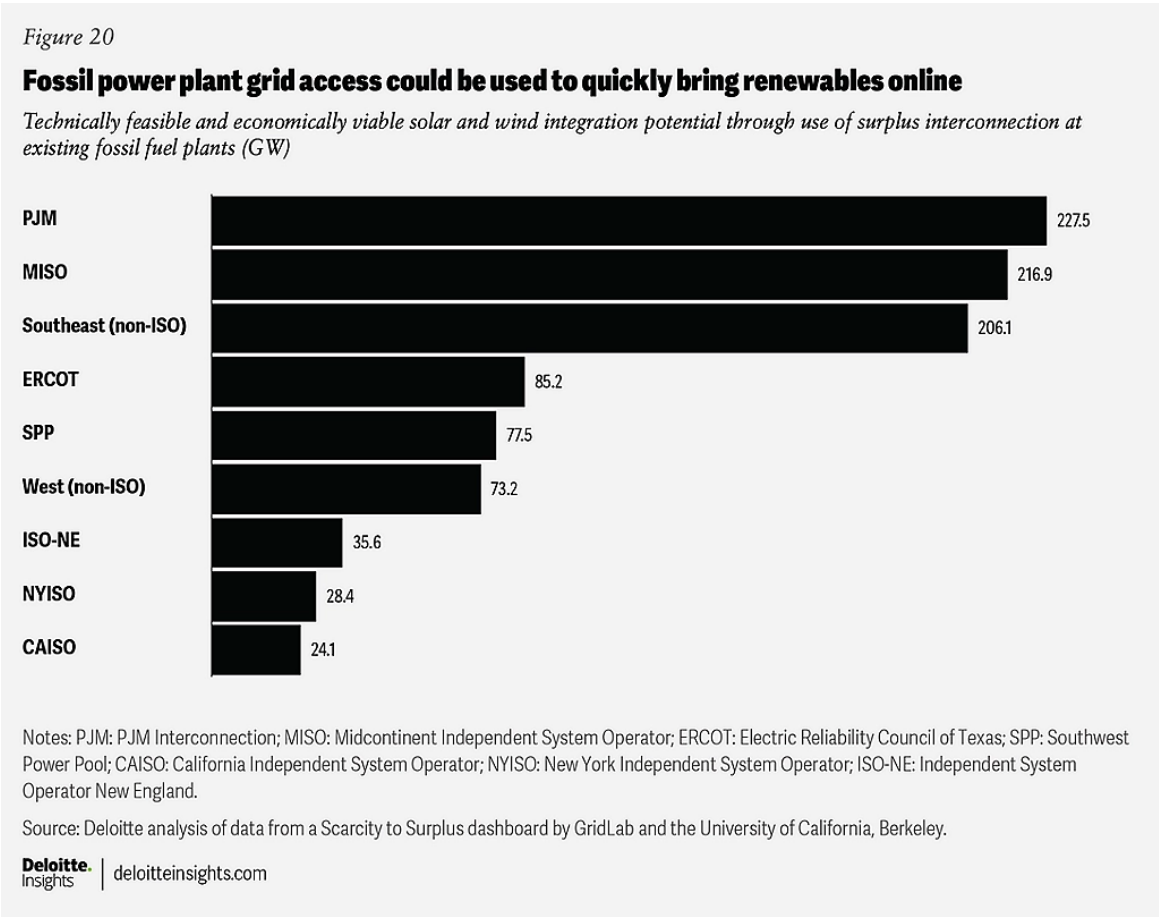
Strategy 5: New business models to drive infrastructural efficiency, capacity, and flexibility

New business models can capitalize on technological, regulatory, computational task mobility, and funding innovation to build additive infrastructure that can unlock efficiency, capacity, and flexibility. Business models fusing AI data center and power infrastructure development could bring capacity online faster while maintaining resource adequacy and affordability. Colocating new load with commensurate new capacity at a single point of interconnection could be grid

supportive, helping to avoid the need to build more transmission, sharing supply-constrained power equipment, and enabling the injection of surplus energy onto the grid during periods of peak demand.

One approach is to redevelop the sites of closed plants into data center campuses powered by higher capacity generation. For example, in Pennsylvania’s largest capital investment (US\$10 billion), the state’s largest retired coal plant is converting into the largest gas plant in the United States. The conversion will more than double the plant’s capacity to meet demand for multiple AI data centers. By leveraging existing grid infrastructure, the data center campus is expected to start operations by 2027.⁵⁰

Another option could be to tap existing underused surplus interconnection, whereby new generation can connect at the site of existing generators in under a year, compared to more than five years for standard interconnection.⁵¹ A recent study found that fossil fuel plants could cost-effectively share grid access with 800 GW of renewables and double US generation capacity (figure 20).⁵² Another study found that 160 peakers and load-following gas plants have sufficient idle capacity to serve as backup generators for data centers colocated with new renewable and storage capacity; combined, these could meet approximately 50 GW of AI load.⁵³



For hyperscalers seeking entirely renewable generation, the fused infrastructure might need to be more fully integrated with the grid, providing surplus low-cost clean energy to the grid while drawing on the grid for backup. One hyperscaler is partnering with a renewable developer and private equity to invest US\$20 billion in developing an energy park with colocated load, generation, and storage to be operational by 2026.⁵⁴

Potential risks of falling short

Achieving the new scope, scale, stakes, and speed of AI infrastructure development can be a complex undertaking. However, technological, regulatory, funding, and business model innovation can help unlock additive infrastructure for artificial intelligence, or “AI for AI.” Additive infrastructure can bring efficiency, capacity, and flexibility to powering

AI. But to achieve these ends, power companies and data centers should consider partnering both with each other and across industries including real estate, investment funds, utilities, power and gas companies, renewable developers, construction companies, and manufacturers of key components.

Given the complexities, partnerships among these industry players are expected to look very different from the ones of the past. If they fail, power and grid capacity constraints could hamstring AI advancement. Power companies could miss an opportunity to expand and modernize the grid. And domestic manufacturing growth in some sectors could stall and lose an edge. These developments could jeopardize US economic and geopolitical leadership. Indeed, staking an infrastructural lead in powering AI may now be a matter of competitiveness and even national security.

About the survey

The Deloitte Center for Energy and Industrials conducted a survey in April 2025 to identify the challenges, opportunities, and strategies of US data centers and power companies, and to benchmark their infrastructure development. The survey sample of 120 respondents included 60 data center executives and 60 power company executives, who responded to questions on infrastructure build-out challenges, resource mix to meet future energy consumption, workforce issues, AI workload planning, drivers of load growth, and investment priorities.

Continue the conversation

Meet the industry leaders

BY

ENDNOTES

1. Deloitte used DC Byte data on live, under-construction, committed, and early-stage data centers' IT capacity and North American forecasts through 2030, as well as pipeline estimates from Wood Mackenzie and studies tracking chip shipments to determine the current total data center capacity and derive operational capacity in 2035. We further validated these pieces of data against AI data center penetration estimates from the Center for Strategic and International Studies and forecasts through 2035 from Wells Fargo, and applied a 24% CAGR through 2030 and an 8% CAGR through 2035 to reflect efficiency gains.

[View in Article](#)

2. Cade Metz, Karen Weise, Marco Hernandez, Mike Isaac, and Anjali Singhvi, "How A.I. is changing the way the world builds computers," *The New York Times*, March 16, 2025.

[View in Article](#)

3. OpenAI, "Infrastructure is destiny: Economic returns on US investment in democratic AI," Sept. 10, 2024; Matthew Gooding, "Up to 5GW data center with behind-the-meter hydrogen power promised for Laredo, Texas," Data Center Dynamics, March 21, 2025; David Meyer, "OpenAI reportedly wants to build 5-gigawatt data centers, and nobody knows who could supply that much power," *Fortune*, Sept. 27, 2024; US Department of Commerce, "UAE and US presidents attend the unveiling of Phase 1 of new 5GW AI campus in Abu Dhabi," press release, May 15, 2025.

[View in Article](#)

4. U.S. Energy Atlas, "Power plants," Energy Information Administration, accessed April 12, 2025.

[View in Article](#)

5. Tim McLaughlin, "Big Tech's data center boom poses new risk to US grid operators," *Reuters*, March 20, 2025; Leonardo Nicoletti, Naureen Malik, and Andre Tartar, "AI needs so much power, it's making yours worse," *Bloomberg*, Dec. 27, 2024.

[View in Article](#)

6. Deloitte 2025 AI Infrastructure Survey. Share of 98 respondents who opined on AI impact on power demand.

[View in Article](#)

7. Zachary Skidmore, "US data centers face grid bottlenecks as regional operators delay upgrades," Data Center Dynamics, Feb. 3, 2024.

[View in Article](#)

8. Increasing the baseload demand pushes the peak higher. This is similar to pushing a mountain higher by lifting it at the base, versus stacking multiple smaller rocks at the top to increase the peak.

[View in Article](#)

9. Deloitte analysis of S&P and BNEF data.

[View in Article](#)

10. Federal Reserve Bank of St. Louis, "Producer Price Index by commodity: Inputs to industries: Net inputs to nonresidential construction, goods," FRED, accessed May 22, 2025.

[View in Article](#)

11. Sonal Patel, "Gas power's boom sparks a turbine supply crunch," *Power Magazine*, April 1, 2025.

[View in Article](#)

12. Deloitte analysis of US Energy Information Administration data.

[View in Article](#)

13. Sella Nevo, Dan Lahav, Ajay Karpur, Yogev Bar-On, Henry Alexander Bradley, and Jeff Alstott, *Securing AI Model Weights* (Santa Monica, California: RAND Corporation, 2024).

[View in Article](#)

14. Christie Simons, Brandon Kulik, Jan Nicholas, Jordan Bish, Duncan Stewart, and Karthik Ramachandran, “2024 Semiconductor Industry Outlook,” Deloitte, 2024.

[View in Article](#)

15. Executive Office of the President—Council on Environmental Quality, “[Environmental impact statement timelines \(2010–2024\)](#),” Jan. 13, 2025.

[View in Article](#)

16. Shawn Enterline and Andrew Valainis, “[Laws in order: An inventory of state renewable energy siting policies](#),” Regulatory Assistance Project, June 2024; Matthew Eisenson, “[Opposition to renewable energy facilities in the United States: June 2024 edition](#),” Sabin Center for Climate Change Law, June 11, 2024; Matthew Eisenson, “[Opposition to renewable energy facilities in the United States: May 2023 edition](#),” Sabin Center for Climate Change Law, May 31, 2023.

[View in Article](#)

17. Eisenson, “[Opposition to renewable energy facilities in the United States: June 2024 edition](#)”; Eisenson, “[Opposition to renewable energy facilities in the United States: May 2023 edition](#).”

[View in Article](#)

18. Zachary Skidmore, “[Welcome to Gas Land - how natural gas is powering the US AI boom](#),” Data Center Dynamics, May 1, 2025.

[View in Article](#)

19. Ari Peskoe and Eliza Martin, “[Extracting profits from the public: How utility ratepayers are paying for big tech’s power](#),” Environmental and Energy Law Program, Harvard Law School, March 5, 2025.

[View in Article](#)

20. Karthik Ramachandran, Duncan Stewart, Kate Hardin, and Gillian Crossan “As generative AI asks for more power, data centers seek more reliable, cleaner energy solutions,” *Deloitte Insights*, Nov. 19, 2025.

[View in Article](#)

21. Justin Talbot Zorn and Bettina Warburg, “[Rainwater could help satisfy AI’s water demands](#),” *Scientific American*, Nov. 7, 2024.

[View in Article](#)

22. Rhodium Group, “[The potential for geothermal energy to meet growing data center electricity demand](#),” March 11, 2025.

[View in Article](#)

23. Semiconductor Engineering, “[Backside power delivery gears up for 2nm devices](#),” Feb. 26, 2024.

[View in Article](#)

24. Katherine Bourzac, “[Fixing AI’s energy crisis](#),” *Nature*, Oct. 17, 2024; Duncan Stewart, Karthik Ramachandran, and Prashant Raman, “TMT Predictions 2025,” *Deloitte Insights*, Nov. 19, 2024.

[View in Article](#)

25. Ibid.

[View in Article](#)

26. Anna Mattson, “[New AI circuitry that mimics human brains makes models smarter](#),” *Scientific American*, Feb. 7, 2024.

[View in Article](#)

27. Sonal Patel, “The solid-state shift: Reinventing the transformer for modern grids,” *Power Magazine*, Jan. 2, 2025.

[View in Article](#)

28. GridLab, “2035 and beyond report: Reconductoring with advanced conductors can accelerate the rapid transmission expansion required for a clean grid,” April 2024.

[View in Article](#)

29. Federal Energy Regulatory Commission, “Order No. 2023,” Nov. 6, 2023.

[View in Article](#)

30. Electric Reliability Council of Texas, “Load resource participation in the ERCOT markets,” accessed May 2025.

[View in Article](#)

31. Federal Energy Regulatory Commission, “Order accepting tariff revisions,” Feb. 11, 2025.

[View in Article](#)

32. PJM Inside Lines, “Reliability Resource Initiative draws 94 applications,” March 21, 2025; “PJM chooses 51 generation resources to address near-term electricity demand growth,” May 2, 2025.

[View in Article](#)

33. Federal Energy Regulatory Commission, “Order accepting tariff revisions.”

[View in Article](#)

34. Midcontinent Independent System Operator, Inc., “Revisions to the Open Access Transmission, Energy and Operating Reserve Tariff expedited resource addition study filing: Docket No. ER25-____-000,” amendment filing, March 17, 2025; Southwest Power Pool, “Joint stakeholder briefing,” Feb. 3, 2025.

[View in Article](#)

35. Ibid.

[View in Article](#)

36. Southwest Power Pool, “Joint stakeholder briefing.”

[View in Article](#)

37. Federal Energy Regulatory Commission, “FERC strengthens Order No. 1920 with expanded state provisions,” press release, Nov. 21, 2024.

[View in Article](#)

38. Public Utilities Commission of Nevada, “Before the Public Utilities Commission of Nevada,” March 11, 2025.

[View in Article](#)

39. Laila Kearney, “Google deploys AI to speed up connections at PJM, largest US power grid,” *Reuters*, April 11, 2025.

[View in Article](#)

40. Deloitte analysis of data from AlphaSense.

[View in Article](#)

41. Deloitte analysis of data from Pitchbook and Infralogic.

[View in Article](#)

42. S&P Global, “Energy utility capex predicted to top \$1 trillion from 2025 through 2029,” April 3, 2025. This is based on the forecast for 47 electric and gas utilities in Regulatory Research Associates’ representative sample of publicly traded, US-based utilities. Capex figures do not isolate or control for spending that was going to occur independent of data centers.
- [View in Article](#)
43. Deloitte analysis of UBS estimates pulled from Alphasense.
- [View in Article](#)
44. Jackie Davalos, “Tech giants expected to ramp up AI spending spree after DeepSeek,” *Bloomberg*, March 17, 2025.
- [View in Article](#)
45. Jason Lehmann, Brian Collins, Dan Lowrey, and Heike Doerr, “Energy utility capex predicted to top \$1 trillion from 2025 through 2029,” S&P Global, April 3, 2025.
- [View in Article](#)
46. NVIDIA, “GTC 2025 – announcements and live updates,” March 25, 2025.
- [View in Article](#)
47. NVIDIA, “NVIDIA to manufacture American-made AI supercomputers in US for first time,” press release, April 14, 2025; Apple, “Apple will spend more than \$500 billion in the U.S. over the next four years,” press release, Feb. 24, 2025; TSMC, “TSMC intends to expand its investment in the United States to US\$165 billion to power the future of AI,” press release, March 4, 2025.
- [View in Article](#)
48. Ben Berkowitz, “Magnificent 7 stocks set for \$1 trillion rout on DeepSeek panic,” *Axios*, Jan. 27, 2025.
- [View in Article](#)
49. Davalos, “Tech giants expected to ramp up AI spending spree after DeepSeek.”
- [View in Article](#)
50. Christa Marshall, “Pennsylvania to host largest-ever US gas plant to fuel data centers,” *PoliticoPro*, April 3, 2025.
- [View in Article](#)
51. GridLab, “Unlocking the power of surplus interconnection: Barriers, opportunities, and strategic solutions,” Feb. 21, 2025.
- [View in Article](#)
52. Umed Paliwal, Emilia Chojkiewicz, Nikit Abhyankar, and Amol Phadke, “Existing fossil fuel plants sharing grid access with renewables can rapidly and cost-effectively double US generation capacity,” Goldman School of Public Policy, November 2024.
- [View in Article](#)
53. Alex Engel, David Posner, and Uday Varadarajan, “How power couples can help the United States win the global AI race,” Rocky Mountain Institute, Feb. 20, 2025.
- [View in Article](#)
54. Brian Martucci, “Intersect Power, Google, TPG partner on up to \$20B renewable energy push for AI,” *Facilities Dive*, Dec. 18, 2024.
- [View in Article](#)
-

ACKNOWLEDGMENTS

The authors would like to thank **Tarun Dronamraju** and **Jaya Nagdeo** for their key contributions to this report, including research, analysis, and writing.

Deloitte Advisory Board:

Tom Keefe, Martin Stansbury, Kelly Marchese, Marlene Motyka, Christian Grant, Duncan Stewart, Adrienne Himmelberger, Diana Kearns-Manolatos, Austin Sincock, and Karthik Ramachandran

The authors would like to acknowledge the support of **Clayton Wilkerson** for orchestrating resources related to the report; **David Levin** and **Oguri Srinivasarao** for their support on survey analysis; **John Forster** for his support on PitchBook analysis; **Tara Treado** for driving the marketing strategy and deploying related assets to bring the story to life; **Alyssa Weir** for her leadership in public relations; and **Elizabeth Payes, Aparna Prusty, Harry Wedel, and Pubali Dey** from the *Deloitte Insights* team for supporting the report's publication.

Cover image by: **Pooja Lnu**
