LIFE:POWERED

AN ASSET OR A LIABILITY:

How Data Center Policy Will Shape ERCOT's Future

WRITTEN BY **Brent Bennett, Ph.D., Joshua Loucks, and Kiersten Saathoff**

November 2026

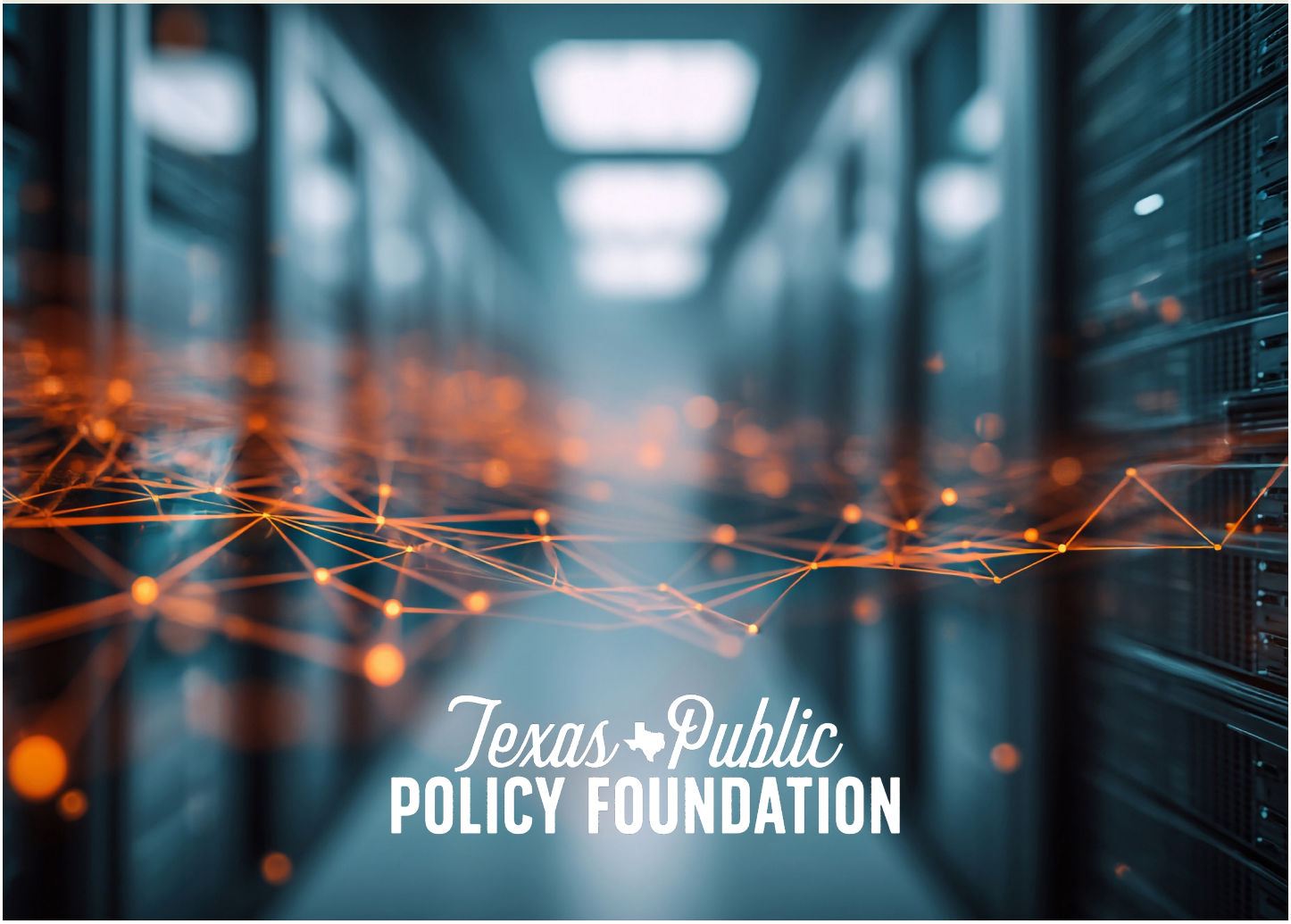


TABLE OF CONTENTS

Executive Summary | Page 3
Introduction | Page 4
An Economic Case for Data Centers: Too Important to Get Wrong | Page 7
Challenges with Data Center Development | Page 10
Responsible Growth of AI Data Centers | Page 12
The Impact of Data Centers on ERCOT | Page 15
The Economic and Reliability Needs of AI Data Centers | Page 20
Policy Recommendations | Page 23
Conclusion | Page 27
References | Page 29

AN ASSET OR A LIABILITY:

How Data Center Policy Will Shape ERCOT's Future

WRITTEN BY **Brent Bennett, Ph.D., Joshua Loucks, and Kiersten Saathoff**

KEY POINTS

- **Texas is facing unprecedented electricity demand** from data centers . ERCOT is tracking over 474 GW of large-load requests, nearly 90% from data centers, on a grid that currently peaks over 90 GW.
- **Data centers are not the cause** of rising electricity bills. Renewable energy mandates and retirements of thermal generation have driven most price increases, while the states adding the most demand have seen some of the smallest rate increases.
- **Texas is positioned to lead the world** in AI infrastructure the way it has led previous energy booms. Getting the integration right turns a decade of private capital into durable grid capacity and economic opportunity.
- **Data centers can strengthen the ERCOT grid** by financing new reliable generation and transmission, but not without changes to current policies that socialize reliability and transmission costs.
- **The window for getting the design right** is closing. Interconnection decisions and transmission commitments made in the next few years will lock in outcomes for decades.

EXECUTIVE SUMMARY

Texas is on track to host the largest concentration of data centers in the world. The Electric Reliability Council of Texas (ERCOT) is now tracking more than 474 GW of large load interconnection requests, 90% of it from data centers (ERCOT, 2026b). Only a fraction of that queue will ever be built, but the fraction under contract is already substantial: Cushman & Wakefield estimates 3,396 MW operating in Texas today with 4,730 MW under construction, and 28,763 MW planned—more than eight times what is in operation today (Cushman & Wakefield, 2026). Whether that growth becomes an asset or a liability to the ERCOT grid is not a matter of chance. It is a matter of policy design, and the window for getting the design right is now.

Data center growth is increasingly being blamed for rising electric bills, but the evidence does not support that claim. California and New York have posted the largest rate increases while losing demand, while Texas and Virginia have added the most data center load in the country without steep price increases. Recent research has found that doubling data center capacity at a fixed level of residential demand lowers residential retail prices by 3.5% (Watten et al., 2026). The drivers of rising bills are elsewhere: mandated retirement of reliable generation, transmission built for distant wind and solar generation, and aging distribution infrastructure (Pyle et al., 2025).

The risk to Texas consumers is not the data centers themselves but the electric grid that will serve them. Less than 20% of the 456,578 MW in ERCOT's generation queue is reliable thermal generation (ERCOT, 2026c). In fact, Texas has not added significant firm dispatchable capacity in over a decade, and wind and solar have been overbuilt to a point

that significant amounts are being curtailed (Modo Energy, 2025). Both facts point to a market design that produces a reliability deficit and promotes inefficiency, which ratepayers pay for. Meeting new load with still more intermittent generation would erode reliability and push hyperscalers off the grid—a shift already underway, with more than 40 GW of private grid development announced in Texas (Thomas, 2026a). Islanded capacity enables speed and innovation, but it does not ensure the on-grid reliability residential consumers depend on.

Handled well, data centers become the anchor load that funds firm generation and spreads fixed costs across a broader base. Handled poorly, Texas adds inflexible load to an increasingly intermittent fleet. That is the choice facing Texas today, and Texas needs to make the right policy choices to responsibly accommodate data center growth to the benefit of all its citizens.

INTRODUCTION

Texas will likely be home to more data centers than any other state by the end of this decade. While the size and pace of the buildout may seem like a daunting reality, in practice it offers a unique opportunity to improve ERCOT's ability to handle extreme weather and supply volatility while lowering electricity costs for Texas consumers.

Throughout history, technological and industrial advancements that have drastically altered society have also impacted policy, economics, and our environment. From the invention of the cotton gin to the dawn of the internet, American society, culture, and economics have been drastically altered at each turn. Typically, advancements have been tangible or relatively narrow in scope, but artificial intelligence (AI) is bringing rapid and sweeping changes to almost every business sector as well as in our personal lives. AI, particularly the means by which it is powered, presents a historically unique intersection between industry and consumers.

The power, water, and even land required to make advanced computing possible is not particularly unique to new industrial advancements . Whether it was textile mills powered through hydropower in Upstate New York in the 1800s or coal-fired steel mills in Pennsylvania, American energy and ingenuity have been the catalyst of growth and opportunity. The challenges data centers present in the 21st century may be more abstract, but in reality, are quite similar to challenges other power-intensive industries have posed in the past. These challenges, if addressed promptly and cohesively through responsible water, land, and electricity production and use agreements, can lead America and the globe into a new era of prosperity and growth.

This paper will address the various power and electrical infrastructure needs to provide the computing necessary for the "AI Revolution," in addition to the economic opportunity these facilities can bring. We will also briefly cover several of the externalities of AI (such as water and land use), while reserving more extensive discussion of those issues for future work.

Texas and America cannot afford to get AI and hyperscale industrial capacity wrong, as global adversaries, most notably China, present an ever-looming threat—especially if they hold greater computing and industrial capacity. Fostering a reliable and secure grid through the increased capital investments from hyperscalers will allow Texas and other states to modernize their grids in a few years rather than over a few decades. The policy decisions made by elected officials and utility commissions today will determine how beneficial the technology will be for everyone from consumers to businesses to our nation's military for decades to come.

HOW DATA CENTERS CONSUME AND MANAGE ENERGY

The Different Varieties of Data Centers

The phrase "data center" is a catch-all that obscures meaningful differences between different types of computing and how they interact with the electric grid. In practice, there are three main types of

data centers: cloud computing (traditional) facilities, bitcoin mining operations, and AI data centers (hyperscalers). Each is distinct in function, scale, and how they affect grid operations.

Cloud Computing (Traditional Data Centers/ Storage)

Traditional data centers power the digital services of the modern world, enabling almost every online service, including banking, healthcare, e-commerce, and social media. This category encompasses both enterprise-owned facilities and colocation/cloud providers that rent computing capacity to multiple businesses. They represent a marginal amount of future load growth, but they are a significant part of the computing mix today and demonstrate the techniques all types of data centers use to provide high reliability and uptime (Smith et al., 2026, p. 21).

From a grid-interaction standpoint, these facilities are relatively predictable. Although very large compared to almost every other type of load, typically between 10–50 MW, they are smaller than the newer classes of data centers and less variable in their power use. They commonly prioritize uptime and maintain significant redundancy, including Uninterruptible Power Supply (UPS) systems and backup generation. This means they can and often do island (or disconnect) themselves from the grid during disturbances, but their load profile is stable enough that grid operators can easily plan around them.

Bitcoin Mining (Cryptocurrency)

Bitcoin mining facilities represent the most grid-flexible category of large computational load, and paradoxically, they are sometimes treated as a nuisance in policy discourse. Unlike data centers, which ramp up or down in response to customer demands or training cycles, cryptocurrency mining can be fully ramped up and down on command by the facility operator and is very sensitive to the price of electricity. Therefore, these facilities can provide a release valve when the system gets tight and prices are high, helping to prevent emergency conditions by reducing demand during those periods.

The challenge with bitcoin mining is that many facilities share the same price threshold, and the simultaneous ramping down of their demand can cause voltage disturbances across the grid. On the other hand, this simultaneous disconnection from large load users frees up significant capacity to accommodate demand surges. These facilities will self-curtail at relatively low margins compared to almost every other large industrial user.

AI Data Centers (Hyperscalers)

AI data centers are the centerpiece of the current historic demand growth and require the most specialized policy decisions to accommodate. Data centers alone may account for up to 12% of all U.S. electricity consumption by 2030, up from 4.4% in 2023 (Shehabi et al., 2024, p. 5; Smith et al., 2026, p. 9). What distinguishes AI facilities from traditional data centers is both their scale, often gigawatt-sized campuses, and the variability of their demand. They are also highly sensitive to power outages, both because of physics—the chips can easily melt or short out when hit with power fluctuations—and economics—the chips are so expensive that every minute they are idle, they lose significant revenue. Even a brief outage can cost operators millions of dollars, making reliable power not a preference but a hard constraint on where AI infrastructure gets built.

This dependence is a policy lever. AI data centers are among the few load categories that will actively seek out and contract for firm, always-on power, creating a natural offtake market for the baseload generation that ERCOT's increasingly weather-dependent grid urgently needs. Transmission planning, interconnection rules, and energy management policies can be designed to channel that demand toward firm resources, turning what is broadly described as a grid stressor into a grid stabilizer. How policymakers address hyperscalers, particularly AI data centers, will make or break electricity policy over the next decade from an affordability and reliability perspective as well as broader economic growth and national security concerns.

The Scale of Hyperscaler Investment

The financial scale of the AI boom is without precedent in modern American history, and the electrical capacity needed to accommodate the largest companies and capital investors in the world cannot be overstated. If current projections hold, combined capital expenditure (capex) from the five largest hyperscalers alone will exceed the GDP of entire countries such as Switzerland, Norway, Sweden, and Vietnam (World Bank Group, 2026). Understanding the size and scale of global capital investment into these hyperscale facilities is essential context for grid planners and regulators who must design systems capable of accommodating this demand.

Various firms, including Bank of America and Evercore, have estimated that 2027 capital expenditures (capex) of hyperscalers will be more than $1 trillion (Burns, 2026). According to Bessemer Venture Partners, "By next year, capex from the big five (Google, Amazon, Meta, Microsoft, and Oracle) are expected to cross $1 trillion per year. Cumulatively from 2025 to 2031, roughly $8 trillion will go into the ground, running parallel with the entire yearly U.S. Department of War budget over the same period" (Deeter & Bondy, 2026). An August 2026 Wall Street Journal analysis put the true figure higher still, identifying roughly $3 trillion in off-balance-sheet AI-related commitments across nine leading technology firms, chiefly uncommenced data-center leases and chip purchase obligations, about triple the debt and leases carried on their books (Rudegeair & Santilli, 2026).

As shown in **Figure 1**, the five largest investors in the AI buildout each saw their cloud computing and AI revenue expand between 20% and 48% year over year, with the cloud computing divisions of Google and Microsoft posting the steepest gains (Alphabet, 2026; Microsoft, 2025). That growth is being realized now rather than just in projections, indicating that the capital committed to data centers is supported by demand that is already showing up.

NVIDIA, the world's leading AI chip producer, surpassed $5 trillion in market capitalization last year. NVIDIA's rise is emblematic of the meteoric race for AI computing power, which is quickly reshaping the global economy. The company reported fiscal year 2026 revenue of $215.9 billion, with data center segment revenue of $193.7 billion, representing 90% of total revenue and a 68% increase from the prior year (NVIDIA, 2026).

The ERCOT interconnection queue alone illustrates the scale: as of July 2026, ERCOT is tracking more than 474 GW of large load interconnection requests, 90% of it from data centers (ERCOT, 2026b). Even at a 10% realization rate applied to the data-center share of the queue, that represents roughly 42 GW of new load and on the order of $2 trillion in capital investment to build out, including the GPUs. The capital expenditures flowing into AI infrastructure have reached a scale that defies historical comparison.

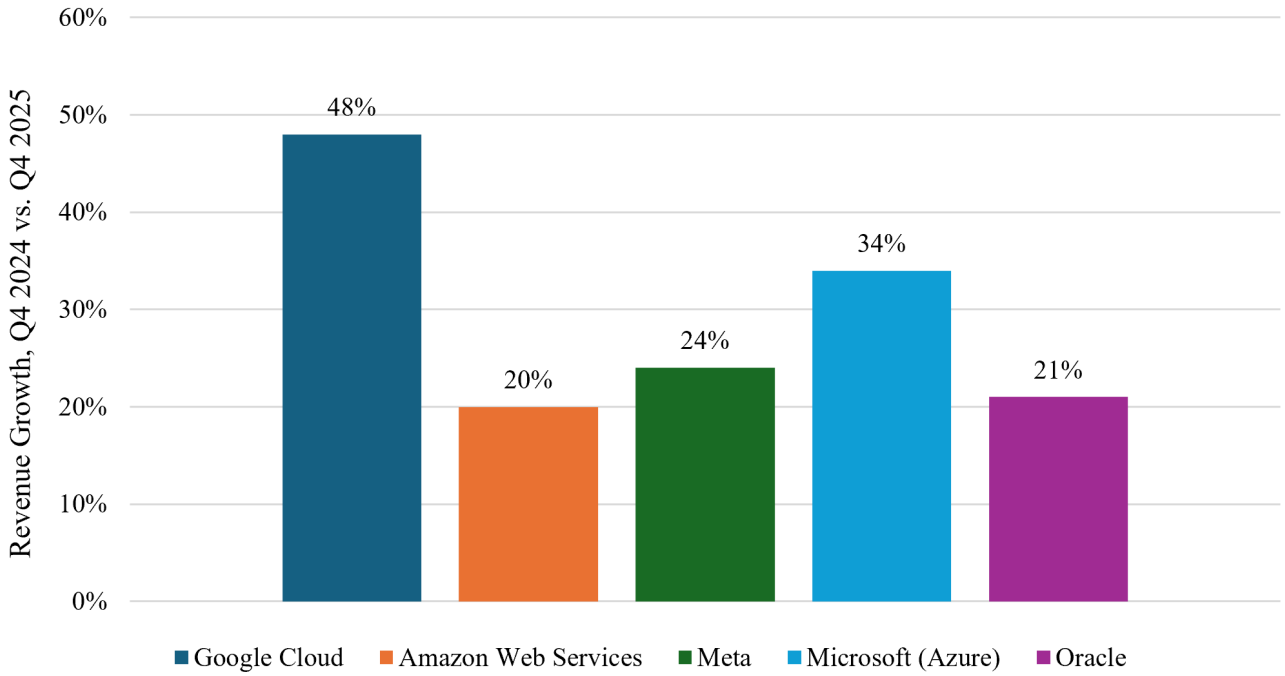
National Projections for AI Data Center Energy Consumption

While U.S. power consumption rose steadily throughout most of the 20th century, growth was relatively stagnant from the turn of the century until around 2023 (Schipper & Hodge, 2025). The rise of reindustrialization, particularly large industrial data companies, and electrification efforts has grid operators and electricity policymakers set to face, in absolute terms, the largest increase in electricity consumption the U.S. has ever seen (NEMA, 2026).

Due to the economic and policy impacts of such drastic load growth, dozens of institutions, including the Department of Energy (DOE), have attempted to model the projected growth and its effects. The DOE's Lawrence Berkeley National Laboratory (LBNL) is what we primarily rely on for many of our assumptions in this paper, but financial and consulting groups like Bloomberg, Bain & Co., and McKinsey also help provide alternative estimates and important perspectives on the capital investment projections that are necessary to facilitate an industrial and energy boom. The uncertainties and

Figure 1

The Five Largest AI Infrastructure Investors Saw Revenue Grow by at Least 20% from Q4 2024 to Q4 2025



Note: Data from Alphabet. (2026). *Alphabet announces fourth quarter and fiscal year 2025 results*. (https://s206.q4cdn.com/479360582/files/doc_financials/2025/q4/2025q4-alphabet-earnings-release.pdf), Amazon.com, Inc. (2026). *Amazon.com announces fourth quarter results*. (https://ir.aboutamazon.com/news-release/news-release-details/2026/Amazon-com-Announces-Fourth-Quarter-Results/), Meta Platforms, Inc. (2026). *Meta reports fourth quarter and full year 2025 results*. (https://investor.atmeta.com/investor-news/press-release-details/2026/Meta-Reports-Fourth-Quarter-and-Full-Year-2025-Results/default.aspx), Microsoft. (2025). *Annual report 2025*. (https://www.microsoft.com/investor/reports/ar25/index.html), Oracle Corporation. (2026). *Oracle announces record Q4 and FY 2026 results driven by cloud infrastructure & cloud applications*. (https://investor.oracle.com/investor-news/news-details/2026/Oracle-Announces-Record-Q4-and-FY-2026-Results-Driven-by-Cloud-Infrastructure--Cloud-Applications/default.aspx), and author's calculations.

challenges in such forecasting are plainly evident in all of these reports. LBNL's own scenarios carry wide uncertainty—on the order of -11% to +21% around its reference case, widening to roughly -20% to +30% once compounding factors are included (p. 26).

Across these models, U.S. data center demand roughly doubles from most recent estimates of roughly 40 GW in 2024–25 to approximately 80 GW by 2028. Beyond 2028, the estimates diverge sharply, reflecting genuine uncertainty about grid capacity, economic growth, and the pace of AI adoption. But the disagreement is about magnitude, not direction. Every credible projection points the same way: sustained, substantial growth in load that policymakers and grid operators must account for.

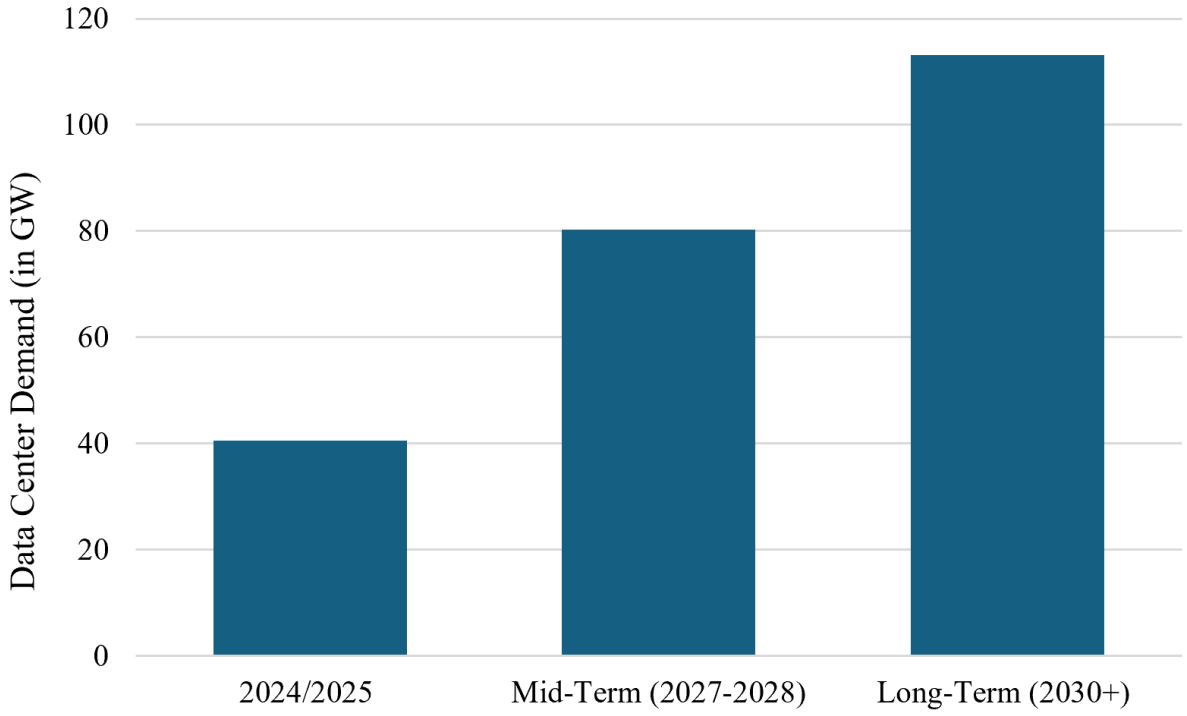
AI data centers represent only one significant driver of load growth among several categories. Broader attempts to electrify transportation, industrial reshoring, and the digitization of commerce (including cryptocurrency mining) all contribute to rising electricity demand independent of AI. Overweighting AI demand while ignoring these parallel drivers risks both overbuilding of infrastructure for AI and underbuilding for other needs.

AN ECONOMIC CASE FOR DATA CENTERS: TOO IMPORTANT TO GET WRONG

The stakes of data center policy extend far beyond the grid. Artificial intelligence and the infrastructure that supports it represent one of the most significant drivers of long-term economic growth in modern

Figure 2

U.S. Data Center Electricity Demand Doubles by 2028 and Nearly Triples by 2030



Note: Data from BloombergNEF. (2025). *AI and the power grid: Where the rubber meets the road*. (https://about.bnef.com/insights/clean-energy/ai-and-the-power-grid-where-the-rubber-meets-the-road/), Morgan Stanley. (2026). *Powering AI: Markets race to invest in AI energy solutions*. (https://www.morganstanley.com/insights/articles/powering-ai-energy-market-outlook-2026), Kok, E., Rauer, J., Sachdeva, P., Pikul, P., Sutton, D., & Hasrouni, R. (2025). *Scaling bigger, faster, cheaper data centers with smarter designs*. McKinsey & Company. (https://www.mckinsey.com/industries/private-capital/our-insights/scaling-bigger-faster-cheaper-data-centers-with-smarter-designs), Shehabi, A., Smith, S. J., Hubbard, A., Newkirk, A., Lei, N., Siddik, M. A. B., Holecek, B., Koomey, J., Masanet, E., & Sartor, D. (2024). *2024 United States data center energy usage report*. Lawrence Berkeley National Laboratory. (https://escholarship.org/uc/item/32d6m0d1), Hering, G., & Dlin, S. (2025). Data center grid-power demand to rise 22% in 2025, nearly triple by 2030. *S&P Global*. (https://www.spglobal.com/energy/en/news-research/latest-news/electric-power/101425-data-center-grid-power-demand-to-rise-22-in-2025-nearly-triple-by-2030), and the author's calculations.

history. Texas, more than any other state, is positioned to capture a disproportionate share of that growth. This growth can either happen in Texas, responsibly and to the benefit of wage-earners and businesses across the state, or it can happen elsewhere. The policy choices Texas leaders make will determine whether Texas leads that transformation or cedes it to other states (or even countries).

The macroeconomic case is compelling. The Penn Wharton Budget Model projects that AI will produce a permanent increase in the level of economic activity, with GDP and total factor productivity levels 1.5% higher by 2035, nearly 3% higher by 2055, and 3.7% higher by 2075 (Arnon, 2025). Penn Wharton also estimates that "40% of current GDP could be substantially affected by generative AI," permanently changing the economic impact of key industries in Texas and abroad (Arnon, 2025). These are not speculative projections about a distant future; they describe an economic transformation already underway, and one that will reward the states that build the infrastructure to support it. AI, just like the internet and portable devices (i.e., smartphones and laptops), will not only impact productivity but also usher in new companies, business models, and economic growth.

There are also immediate and tangible construction and labor benefits. Every billion dollars of construction spending generates approximately 3,450 jobs, a multiplier that ripples through local supply chains, trades, and service economies. The construction industry will need to absorb 349,000 new workers in 2026 alone just to keep labor markets in equilibrium, and data center megaprojects are among the primary drivers of that demand (Associated Builders and Contractors, 2026). As the Associated Builders and Contractors has noted, more than one in every five dollars spent on non-residential construction currently flows toward manufacturing and data center projects, and those projects are absorbing a significant share of regional labor forces wherever they land (Associated Builders and Contractors, 2026).

For Texas, which has led the country in job creation and population growth for over a decade, this represents a natural extension of an existing competitive advantage and further opportunity for workers and businesses.

Beyond direct economic output, data center investment generates property tax revenues that fund local schools and infrastructure. In Loudoun County, Virginia, data centers occupy roughly 3% of the county's land but will provide 45% of its nearly $2.9 billion in tax revenue (Glock, 2026). That revenue density has allowed the county to cut its real property tax rate every year for the past decade, to the lowest in Northern Virginia. In Texas, Microsoft alone has paid more than $48 million in property taxes across Bexar and Medina counties, prior to even beginning construction of its new multibillion-dollar campus in Reeves County (Walsh, 2026). While the fiscal case alone is compelling, AI stands to play an outsized role in the future of America's economic and national security.

Building AI infrastructure in America is one of the most consequential national security and economic competitiveness decisions of this generation. While a full discussion of this topic falls outside the scope of this paper, no honest accounting of data center policy can ignore it: AI developed in jurisdictions with strict rule of law, robust IP protections, and democratic accountability is categorically different from AI developed under the influence of authoritarian regimes, most notably China. The economic stakes alone are enormous, but the national security dimension is what elevates this from a technology policy question to a societal imperative. Militaries with properly integrated AI will hold decisive advantages in intelligence, logistics, autonomous systems, and weapons development. That is an edge the United States cannot afford to cede, and every policy decision that slows domestic AI infrastructure buildout moves the needle in the wrong direction.

States like Virginia, Pennsylvania, and Michigan are considering banning data centers outright or making them economically unfeasible (National Conference of State Legislatures, 2026). New York has already enacted a moratorium on new data centers over 50 MW (National Conference of State Legislatures, 2026). For states that create a fair and open place to do business, these policy mistakes by other states create a tremendous opportunity. Providing reliable power and accelerated permitting will be a stronger attraction force for jobs and economic investment than any subsidy scheme could ever be.

Like all policy decisions, large-load integration involves trade-offs: between speed and cost assignment, load growth and grid stability, short-term infrastructure investment and long-term reliability. The asymmetry of those trade-offs is important: the cost of getting this right is manageable, and the cost of getting it wrong or of driving investment to other states (or countries) through regulatory overreach is permanently damaging to the U.S.'s ability to compete and protect itself on an international level.

Over the past decade, America has continued to split into places in which businesses and their employees are allowed to grow and flourish, and ones in which the government, first and foremost, decides what is in its best interest. Texas did not become the leading

state for Fortune 500 headquarters, energy production, and population growth by accident, and data centers present yet another opportunity to be a leader in entrepreneurship, innovation, and societal growth rather than decay (Munis, 2026).

CHALLENGES WITH DATA CENTER DEVELOPMENT

Energy Management for Data Centers

Energy management is one of the most important aspects of data centers from both an economic and policy perspective. The standard metric for data center energy efficiency is power usage effectiveness (PUE), calculated as the ratio of total facility energy consumption to the energy consumed by computing equipment. A PUE of 1.0 represents perfect efficiency; any value above 1.0 reflects energy used for cooling, power delivery, and other systems. The global average PUE stood at 1.56 in 2024, meaning that for every unit of energy powering IT equipment, an additional 56% is consumed by cooling, power delivery, and other overhead (Devasia, 2026).

This average, however, masks significant variation: while current industry averages range from 1.5 to 1.7, modern facilities are targeting PUEs as low as 1.1, and leading hyperscalers are already well below that threshold. For example, Google averages 1.09 across its fleet, with best-in-class sites reaching 1.04, while other hyperscale operators range from 1.08 to 1.46 (Kok et al., 2025). The long-term average of AI-optimized facilities will likely range from 1.145 to 1.136 PUE, according to LBNL (Smith et al., 2026, p. 24).

Liquid cooling generally achieves a PUE improvement versus traditional air-cooling (Shehabi et al., 2024, p. 47). However, improved cooling efficiency does not typically reduce absolute energy consumption at any given site because it enables higher rack densities. NVIDIA Blackwell-class GPUs push racks past 100 kW versus prior generations at 30 kW and below (Milyutin, 2026). Hyperscalers are deploying liquid cooling to enable next-generation silicon, not to consume less energy in aggregate. Grid planners should therefore assume that contracted load will be fully utilized and likely exceeded as facilities scale, regardless of cooling technology.

What is largely lost in popular media coverage of data centers is what these facilities actually do. According to LBNL, AI-related use cases currently account for roughly 15-20% of all data center electricity consumption in 2024 (Smith et al., p. 25). As AI integration grows among both individual and corporate users, this share is projected to rise to 55% of total data center energy use by 2030 (Smith et al., p. 25). Even at 55%, it is worth remembering that hyperscalers, including AI-focused facilities from companies like Meta, still use a significant amount of their electricity for non-AI workloads like cloud storage, enterprise computing, and content delivery.

Within that AI slice, the mix of workloads matters for grid planning. Contrary to popular perception, the dominant and fastest-growing component of AI energy demand is not the building of models but the use of them. AI training, or the process of constructing and refining models, is projected to account for roughly 10-20% of long-run AI-specific load, with inference (responding to user queries in real time) already accounting for the rest and growing rapidly. Training is highly schedulable and can be shifted to off-peak hours, making it a more grid-friendly workload than it first appears. Inference, however, cannot be similarly deferred as it must be able to respond immediately to consumer demand and scales directly with user adoption. By most estimates, inference will account for 75-85% of total AI compute energy by 2028 to 2030, meaning long-term grid planning must be built around 24/7 demand growth rather than the more flexible training loads (Avelar et al., 2023).

This asymmetry has important policy implications: training load is an ideal candidate for demand response and curtailment programs, while inference load requires the kind of firm, always-available power supply that AI facilities are already willing to pay a premium to secure. One meaningful way to manage inference (or user-driven queries) is to

establish portable demand agreements with other data centers owned by the same firm. Portable demand agreements with facilities in other states would enable workload shifting during periods of grid stress (or high prices). This would also give data centers the comfort of avoiding user inference shutdowns in the face of curtailment or grid emergencies. However, portable demand is difficult to implement due to latency-sensitive workloads, cross-ISO jurisdictional friction, data transfer costs, and correlated energy supply patterns. On-site power and energy management remain the most viable tools for ensuring the data center meets its reliability needs in times when the broader grid is under stress.

Water and Land Use

Hyperscalers, particularly AI-focused firms, have faced tremendous scrutiny surrounding their water use both at the local level and the national level (Tan, 2025). Undoubtedly, there are real concerns from residents and practical adaptations that must be made by data centers to help minimize externalities imposed on local communities. However, those realities can be missed among the noise of highly dubious headlines such as "Amazon Data Center Linked to Cluster of Rare Cancers" or "'I can't drink the water' - life next to a US data centre" (Wilkins, 2025; Fleury & Jimenez, 2025). Separating fact from fiction and outlining the truth regarding the very real, yet often overstated, externalities of data centers is paramount to creating policy to reasonably regulate data centers as well as ensuring communities are not misled into actively fighting against fundamental economic and societal growth.

Most individuals concerned with data centers are concerned first about water use. Nationally, American data centers consumed roughly 17.4 billion gallons in 2023, about 48 million gallons per day, with hyperscale demand projected to reach up to 89 million gallons per day by 2028 (Shehabi et al., 2024, pp. 55-56). But those volumes are unremarkable compared to the industries Texas already hosts. Hydraulic fracturing in the Permian Basin alone draws 113.5 million gallons of freshwater per day—more than twice the consumption of every data center operating in the United States—in the most water-constrained region of the state (Texas Produced Water Consortium, 2022). Steel tells the same story from a different sector: U.S. Steel Corporation consumed approximately 110 million gallons per day in 2023, meaning one company exceeds the entire national data center fleet (U.S. Steel, 2025). Neither industry required emergency legislation to prevent drought in their regions.

Semiconductor manufacturing presents an even more direct comparison. Unlike data centers, which use water of varying quality for cooling, chip fabrication requires ultrapure water, a resource so refined that it takes 1,400 to 1,600 gallons of municipal water to produce just 1,000 gallons of usable supply (James, 2024). An average chip fabrication facility can consume 10 million gallons of ultrapure water per day, and Taiwan Semiconductor Manufacturing Company (TSMC), the world's largest chipmaker, reported consuming 101 million cubic meters (approximately 26.7 billion gallons) in 2023 alone (Mozumdar, 2025; Bithell, 2025). In comparison, Colossus 2, the largest data center currently in operation and a relatively inefficient user of water, uses approximately 1 million gallons of water per day (Bontigui & Patel, 2026). Matching TSMC's annual water consumption would require roughly 73 facilities of that scale.

These comparisons are not offered to diminish the water management responsibilities of the data centers, but to situate them accurately within the broader industrial landscape that policymakers are already regulating without equivalent alarm. AI's marginal impact on water supply, which can be exacerbated if not managed properly at the local level, could also be offset by the advancements AI can offer local infrastructure. Currently, "Nearly one in five gallons—19.5%—of treated drinking water in the U.S. is lost before it reaches customers or is improperly billed" (Bluefield Research, 2025). Additionally, the U.S. water system collectively loses approximately 6.75 billion gallons of treated drinking

water on a daily basis (Bluefield Research, 2025). Put into perspective, our current water system loses 140 times more treated drinking water than what data centers use in total water consumption—and most of what data centers use is non-potable. If AI tools can help reduce those losses even marginally, they would more than offset the water consumption from the computing needed to support AI. In upcoming work, we will address this question, as well as provide an in-depth analysis of industrial and agricultural water use.

The physical land footprint of AI data centers is also frequently overstated and used to mislead the general public due to a lack of understanding of land requirements for other industries. Siting decisions carry real consequences for communities, but measured in acreage alone, data centers are not a meaningful claim on Texas land. Proper community planning and zoning regulations to encourage data centers to build within existing infrastructure corridors will not only reduce socialized transmission costs but also limit noise and aesthetic concerns, which have resulted in major backlash across the country.

Industry siting practice bears this out. Systematic, parcel-level site screening across ERCOT consistently finds that the strongest candidate sites are rarely open greenfield; they are parcels adjacent to existing high-capacity infrastructure—138 kV and 345 kV corridors, substations with available headroom or room to expand, and industrial land that already carries the roads, water, fiber, and rights-of-way a large facility requires. Screening at the parcel level for substation capacity, corridor adjacency, and load-serving capability routinely surfaces sites where a data center can energize faster, and with materially less new transmission, than a greenfield alternative. The policy implication is straightforward: when data centers site within existing infrastructure corridors, the transmission costs that would otherwise be socialized across ratepayers shrink, land disturbance is concentrated where heavy industry already operates, and the noise and aesthetic frictions that have derailed projects elsewhere are substantially muted. Zoning and permitting frameworks that steer large loads toward these corridors convert a contentious land-use question into an infrastructure-reuse strategy.

RESPONSIBLE GROWTH OF AI DATA CENTERS

Beyond water and land concerns, the claim most frequently repeated in popular media is that data centers are driving up the price of electricity. The reality is more nuanced: while data centers can contribute to short-term rate increases—particularly through transmission buildout costs—when properly integrated and anticipated in grid planning, they improve the economics of public electricity markets, ultimately lowering rates over the long term.

Planning early and organizing load growth in a responsible manner can ultimately result in a large industrial base, subsidizing the cost of electricity for residential customers. Here are a few ways that proper integration can prevent rate hikes to consumers:

- Locate data centers near existing transmission infrastructure.
- Require data centers to pay for new transmission build-out.
- Require data centers to build on-site power and/or to be able to offer extra power to the grid during emergencies.

The political narrative casting data centers as the primary villain behind rising electricity bills is, on the whole, not supported by the available evidence. Some analyses have found that data centers generate surplus utility revenues by paying more than the cost to serve them—effectively cross-subsidizing other ratepayers. An examination of Amazon data centers across four utility territories found that each site generated an average of $3.4 million in net surplus revenue that could help offset costs for other customers (Energy and Environmental Economics, 2026). Far from extracting value from the rate base,

well-structured large-load tariffs can make electricity cheaper for everyone else.

This dynamic is not without precedent in Texas. The state's refining and petrochemical industry—among the most energy-intensive in the world—has long served as an anchor load that strengthened ERCOT's financial foundation and helped socialize fixed grid costs across a broader base of consumers. Industrial growth, in other words, has historically been a feature of grid economics in Texas, not a bug. The data center boom represents an opportunity to replicate that dynamic at scale, but only if the revenue it generates is directed toward the lowest-cost reliable generation rather than simply absorbed into the expansion of intermittent resources while reliability and transmission costs are socialized to other ratepayers. If new data center load is met primarily with additional wind and solar, reliability will decline, and hyperscalers will look to move off-grid, thereby raising long-term costs for ratepayers who are tied to the grid. Texas has not seen a significant addition of firm, dispatchable generation capacity in over fifteen years (ERCOT, 2026a); the data center boom represents a unique opportunity to fix that problem.

State-level evidence reinforces the fact that data centers can be added at scale while keeping prices stable. Texas and Virginia, which have seen the largest increases in data center load of any states in the country, recorded among the smallest electricity rate increases. California and New York, by contrast, saw the largest rate increases despite a reduction in total load. A study by researchers at the Electric Power Research Institute (EPRI) found that "a doubling of data center capacity causes residential retail prices to fall by 3.5% for a fixed level of residential demand" (Watten et al., 2026, p. 21). Additionally, they found that "Data centers and average price changes in percentage terms are negatively correlated," further highlighting how recent increases in electricity costs are associated with other factors, and not data centers (Watten et al., 2026, p.13).

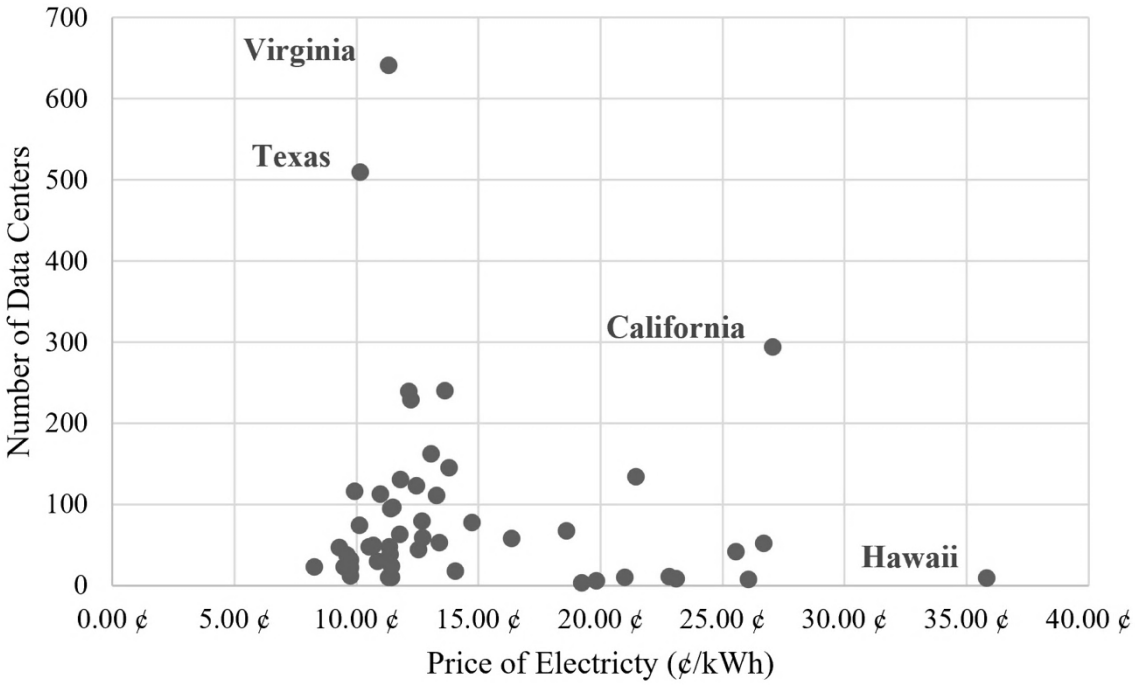
The implication of these studies is clear: load growth driven by data centers is not what is driving bills higher. The culprits lie in mandated retirement of reliable generation, expensive transmission buildout to support intermittent resources, and aging distribution infrastructure. Conflating those drivers with data center growth obscures both the diagnosis and the cure.

Virginia offers perhaps the clearest case study. Despite hosting the largest concentration of data centers in the world, the state has maintained relatively stable residential rates—demonstrating that data centers alone do not drive higher prices. Electricity rates are starting to rise, but the root cause is policy decisions, not data centers. Governor Spanberger's decision to rejoin the Regional Greenhouse Gas Initiative (RGGI) reversed Governor Youngkin's earlier repeal of the program, which had been projected to save Virginia ratepayers $1–1.2 billion over four years. Instead, Dominion Energy estimates that RGGI will increase the average residential electricity bill by $156 each year (Virginia Department of Environmental Quality, 2022). When those bills arrive, green energy and net zero advocates will blame data centers, but Dominion's rate adjustment letter to the state regulators will tell a different story (Virginia State Corporation Commission, 2026). Higher electricity bills are the result of reductive and restrictive energy policies, not data centers.

These case studies are not isolated or mere coincidence; several studies, including from the Institute for Energy Research, have shown that green energy mandates and energy subtraction policies have been the primary driver of high rates, not data centers or thermal energy (Pyle et al., 2025).

The lesson for Texas is that rate hikes due to data centers are not inevitable. They are a function of policy design, not data center presence itself. Texas is, in fact, uniquely positioned to capture these benefits. Abundant natural gas supply and distribution infrastructure, flexible industrial siting, private land ownership, streamlined environmental and zoning

Figure 3
Number of Data Centers and Average Electricity Prices by State, 2025



Note: Data from Pyle, T. J., Stein, K., & Stevens, A. (2025). *Blue states, high rates*. Institute for Energy Research (https://www.instituteforenergyresearch.org/the-grid/blue-states-high-rates/), and Data Center Map. (2026). USA data centers (https://www.datacentermap.com/usa/), and the author's calculations.

regulation, and the independence of the ERCOT grid all create conditions that no other state can fully replicate.

The central policy challenge is synchronizing new data center load with transmission capacity and critically, new reliable generation. HB 3970, which was introduced (and passed the Texas House) during the 89th Texas Legislature, proposed allowing large loads that integrate a minimum amount of firm capacity into the ERCOT grid, or that agree to reduce their demand accordingly, to receive priority in the permitting process (HB 3970, 2025). This approach would help transform data centers from passive load into active participants in grid buildout, aligning private investment with public reliability needs. Managed correctly, the data center boom is not merely an opportunity to strengthen ERCOT's financial foundation—it is the catalyst for a renaissance of reliable generation capacity that Texas consumers have not seen in a generation.

To make this renaissance possible, the state legislature must address the current market structure, which is overvaluing variable generation and undervaluing reliability by paying all generators the same market price. HB 1500, which passed the 88th Texas Legislature in 2023, led to the creation of a firming program that could mitigate this problem (pp. 22-23). However, the program only applies to generators that enter into service after 2026, and the PUC promulgated a rule with a very weak penalty structure that is not sufficient to offset the current reliability deficit in the ERCOT market (PUC, 2025). The rule needs to be strengthened, and new legislation needs to be passed that applies the rule to both existing and new generators.

THE IMPACT OF DATA CENTERS ON ERCOT

Demand Growth

Texas has quickly emerged as one of the most coveted places for data centers to locate, and projected growth in data centers in the state would make it the undisputed national leader within this decade (Cushman & Wakefield, 2026).

Notable developments in Texas include Chevron's announcement of plans to develop up to 2.5 GW of off-grid generation capacity in West Texas in partnership with Microsoft (Walsh, 2026). This is part of a broader trend towards on-site natural-gas-fueled generators and private substations that enable hyperscalers to secure reliable, scalable electricity independent of the ERCOT grid.

Off-grid generation, however, is not a full substitute for interconnection. Even as hyperscalers build private capacity in West Texas and elsewhere across the state, they are seeking grid service in unprecedented volume. The scale and pace of data center interconnection requests have fundamentally altered ERCOT's planning environment. In 2024, approximately 6.5 GW of large load was approved to operate in ERCOT (Energy and Environmental Economics, 2026). By 2025, ERCOT's large-load interconnection queue had surged nearly 300%, with data centers dominating the pipeline. Projections indicate Texas will become the top U.S. data center market imminently; ERCOT's interconnection queue now tracks more than 474 GW of large load seeking connection, with 90% of it from data centers (ERCOT, 2026b). Private-sector projections suggest Texas could exceed 40 GW of data center capacity or nearly 30% of total U.S. demand, representing a 142% increase in market share relative to today (Bloom Energy, 2026).

Taken together, these figures show that Texas is not merely competing for data center growth; it is becoming the center of gravity for it. The policy question is whether ERCOT can accommodate that growth in a way that strengthens reliability and lowers system costs, rather than forcing consumers to absorb the consequences of socialized reliability and transmission costs. Policymakers and legislators in Texas, and across the country, need to seriously plan and manage data center interconnection to ensure neither reliability nor consumer affordability suffers.

Table 1

Planned Texas Data Center Capacity Could Exceed 28 GW, More Than Eight Times What Operates Today

Market	Operational	Under Construction	Planned
Texas (Total)	3,396 MW	4,730 MW	28,763 MW
Dallas-Fort Worth	1,715 MW	1,178 MW	11,891 MW
Austin / San Antonio	1,154 MW	615 MW	7,823 MW
West Texas	213 MW	2,886 MW	8,005 MW

Note: Data from Cushman & Wakefield. (2026). *Americas data center update.* (https://www.cushmanwakefield.com/en/insights/americas-data-center-update).

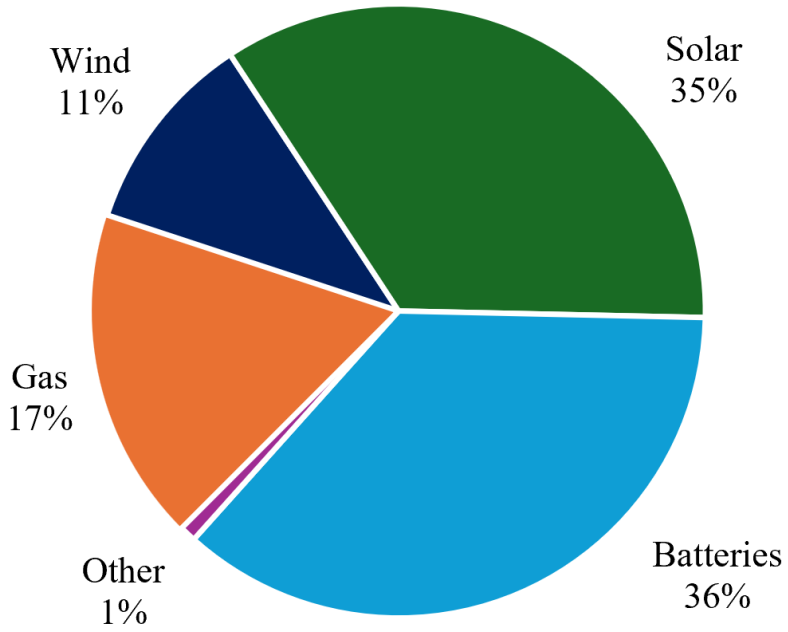
Future Generation Growth

The good news for Texans is there are an estimated 1,906 active generation interconnection requests in the ERCOT region totaling 456,578 MW as of July 31, 2026 (ERCOT, 2026c). The bad news is that gas generation is less than 20% of those interconnection requests (ERCOT, 2026c), and the region has only added 4 GW of natural gas since 2021, compared to over 30 GW of solar, 20 GW of energy storage, and 10 GW of wind (ERCOT, n.d.). No matter how much generation comes into the system, if the mix of generation keeps trending toward more variable resources, the system will become more volatile, and ratepayers will bear the cost of that volatility.

Too much dependency on intermittent wind and solar (let alone limited-duration battery storage) will increase transmission costs and decrease reliability for both residential and industrial consumers. As of January 2026, wind and solar comprised over 40% of the installed generating capacity in ERCOT, about 76 GW (ERCOT, 2025a), yet those resources produced only 9% of that, 7 GW, during the peak demand hour of Winter Storm Fern in January 2026 (Reed & Bennett, 2026, p. 5). Wind and solar produced on average about 18% of their installed capacity for the duration of the storm, compared to 87% availability for the combined gas, coal, and nuclear fleet (p. 5). If the current trends in ERCOT continue, a moderate winter storm—similar to Fern but with higher demand—could result in a multi-day outage similar to what happened during Winter Storm Uri in 2021 (p. 12).

These imbalances are also visible in the form of wasted wind and solar energy. ERCOT curtailed more than 8 TWh of wind and solar output in 2024 (Modo Energy, 2025), and instantaneous curtailment peaked above 8.4 GW in June 2025 (ERCOT, 2025b)—energy whose delivery infrastructure ratepayers helped finance, discarded because generation and

Figure 4
Thermal Generation is less than 20% of ERCOT's Current Interconnection Queue



Note: Data from Electric Reliability Council of Texas. (2026). *Generator Interconnection Status Report, July 2026*. (https://www.ercot.com/mp/data-products/data-product-details?id=PG7-200-ER) and the author's calculations.

demand sit in different places at different hours. Much of this curtailment stems from renewable generation concentrated in export-constrained zones like West Texas and the Panhandle, far from load—a geographic mismatch that transmission relieves only slowly and partially (Modo Energy, 2025). Stranded output at this scale is a ratepayer liability today; as Recommendation 4 develops, well-sited large loads are the one instrument that can convert it into productive use without adding a dollar of socialized cost.

The historic amount of new load brought by data centers can either be a gift or a curse. Policymakers can choose to welcome data centers responsibly by allowing them to interconnect and bring reliable thermal generation sources into ERCOT to help strengthen the grid and reduce costs to consumers. Alternatively, failing to do so will result in additional nonflexible load and intermittent energy that is not well-matched to that load and is vulnerable to the weather.

Beyond overreliance on intermittent resources, a second concern is infrastructure build-out: the risk that transmission and distribution investments triggered by specific large loads will be socialized across all ratepayers rather than borne by the facilities that necessitated them. Implemented incorrectly, consumers will be forced to pay for the transmission necessary for hyperscalers, just as they continue to pay for the costs of new transmission for wind and solar (Bennett, 2026).

Current ERCOT procedures assign dedicated facilities and load-triggered upgrades directly to the connecting customer, limiting cost socialization, but meaningful risk remains in how transmission and distribution utilities manage network-level costs and whether retail electric providers can hedge adequately against price volatility during periods of rapid demand growth (Energy and Environmental Economics, 2026). Cost assignment and interconnection reform are a good start to help ensure those who need this new flux of bulk power are the ones paying for it.

The Texas Legislature's most significant recent response to large-load interconnection challenges was SB 6 (2025), which establishes a new framework for how data centers and other large loads access the ERCOT grid. Key provisions include:

- Requiring PUCT to ensure that large load customers contribute to network upgrades necessitated by their interconnection, establishing interconnection standards designed to minimize stranded costs and maintain system reliability.

- Mandating disclosure of duplicative interconnection requests within Texas.

- Requiring demonstration of site control and financial commitment before allowing a large load to enter the interconnection queue.

- Mandating disclosure of on-site backup generation and requiring large load customers that can serve at least 50% of on-site demand to be able to either deploy the customer's on-site backup generating facilities or curtail load during emergency conditions.

The policy direction advanced in this paper reflects a growing consensus that stretches from Austin to Washington, D.C. In March 2026, the White House secured commitments from Amazon, Google, Meta, Microsoft, OpenAI, Oracle, and xAI under the Ratepayer Protection Pledge. Each company agreed to build, bring, or buy the generation needed to power their facilities, cover the full cost of grid infrastructure upgrades, and negotiate separate rate structures with utilities and state governments rather than passing those costs to households (The White House, 2026). In June 2026, Governor Abbott directed the PUC and ERCOT to enforce the same principles in Texas, requiring data centers to fully fund their own infrastructure costs and reducing residential transmission costs. Additionally, he pledged to work with the legislature to codify and expand the framework established by SB 6 (Office of the Governor of Texas, 2026a).

The throughline across all three is the same argument this paper makes: data center growth should strengthen the grid and lower costs for ordinary ratepayers, and the policy tools to make that happen are available, tested, and long overdue.

Grid Integration: Interconnection Incentives and Load Response

The growing viability of private grids and islanded power solutions represents perhaps the most consequential development in ERCOT's data center policy environment. Private grid solutions offer hyperscalers something the current grid cannot always guarantee: the ability to innovate and scale on their own timeline, free from interconnection queues, socialized cost assignments, and reliability constraints imposed by a grid whose policy framework has not kept pace with the speed of AI infrastructure buildout. For grids that are underprepared for this level of growth, private grids also function as a release valve, helping to absorb and mitigate demand that would otherwise overwhelm an already strained interconnection process.

Structured well, both interconnected systems and private grids can leave Texas consumers better off than the status quo. Interconnection strengthens the grid directly if hyperscalers bring the firm capacity they require for their operations. Private grids strengthen it indirectly, by pulling the largest and least price-sensitive loads out of a queue that is already oversubscribed. Every gigawatt that self-supplies is a gigawatt of studied capacity, transmission headroom, and interconnection capacity left available to loads that have no alternative to the queue, such as manufacturers, petrochemical operations, military industrial base infrastructure, and mid-sized industrial customers who cannot finance a private power plant. Against a queue tracking more than 474 GW of requests, that relief is not incidental—it determines whether the other growth Texas is attracting can still get connected on a commercially reasonable timeline.

Speed to power is the more immediate advantage of private grids. A developer building its own generation controls the critical path forward, allowing greater flexibility and continuity in timelines. Delayed energization is delayed revenue and uptime, which for hyperscalers can cost hundreds of millions of dollars for even just a few months of delays. The same speed advantage can benefit the broader system: generation sited, permitted, and built quickly on site is generation that already exists, and interconnection rules that preserve the option to connect it later capture capacity that would otherwise be locked away permanently or never built in the first place.

Private grids also concentrate risk and reward on the same balance sheet. Whether the bet is a long-term commitment to small modular reactors (SMRs) or a renewable-heavy generation mix built to satisfy corporate emissions targets, the developer absorbs the consequences. SMRs demand capital on a payback timeline only a buyer valuing firm power as an input to a far more valuable product can rationally accept—which is why hyperscalers, not utilities, are underwriting first-of-a-kind reactor commitments (Martucci, 2025). A firm facing a profit-and-loss signal takes first-of-a-kind risk for the reward; a utility earning a regulated return does not. These first-of-a-kind costs are paid once, and the licensed designs, supply chains, and trained crews that follow are what make the next unit cheaper—for ratepayers who never financed the first. Private grids let hyperscalers pursue solutions on a risk tolerance that regulated cost recovery cannot replicate, and Texas ratepayers stand to inherit the long-term benefits without having to underwrite the experiment.

These advantages are real, but they are not the whole story. When hyperscalers opt to build private, islanded facilities rather than interconnect with the ERCOT grid, they avoid imposing socialized costs on the surrounding electric grid, but they also represent a lost opportunity to contribute more to the system and fall entirely outside the cost-assignment framework that SB 6 was designed to enforce. Every facility that opts out of grid interconnection is a missed

opportunity to increase the utilization of existing grid assets and provide revenue to help the grid spread out its costs among more consumers. Due to this viable alternative, it is important to strike a balance to ensure interconnection helps both hyperscalers and other consumers. The goal of Texas data center policy should be to allow private grid development while ensuring that grid interconnection remains an attractive and economically rational alternative to islanding.

The scale of off-grid development already underway makes this balance urgent. According to Cleanview (Thomas, 2026a), data centers have already announced over 40 GW of behind-the-meter capacity in Texas, part of what they expect to be 59 sites with a total of approximately 90 GW of announced capacity.[1] If policymakers are unable to make interconnection in ERCOT affordable, efficient, and reliable, data centers will choose to build off-grid, rather than interconnect. The audit requested by Gov. Abbott in August 2026 (Office of the Governor of Texas, 2026b) sharpens this distinction.[2] The verification only applies to projects in ERCOT's interconnection queue, not the fully islanded builds that sit outside it. If the review slows grid-connected development too much, it will lead toward the very outcome Texas should least want—capacity that exits the system rather than supporting it. Data quality is essential in the planning process for Batch Zero and must not be sacrificed, but that scrutiny must be paired with a faster, more certain path to interconnection.

The extra capex of dedicated power is real but modest compared to the overall capex of most data centers. Depending on the generation mix and reliability target, the added on-site generation-and-storage layer runs roughly $1.5–7 million per MW—the low-end being gas turbines or engines with a battery buffer, the high end a fully islanded renewable build, where solar and wind must be overbuilt several-fold and paired with large storage, based on published capital cost ranges for these technologies (Lazard, 2025). Fully islanded, renewable-heavy configurations account for the upper end of that range; the gas-anchored builds hyperscalers favor sit at the low end of the range. Against the whole project, that cost is modest: a large AI data center typically runs on the order of $35–40 million per MW all-in, dominated by the GPUs and cooling (Michael & Cottier, 2026). Data center operators are already internalizing this calculation, which is why announced off-grid capacity nationwide continues to grow despite the cost premium (Thomas, 2026a).

This vast cost difference for hyperscalers shows how meaningful policy changes to increase reliability would allow for greater interconnection. This interconnection will benefit all parties involved, rather than forcing data centers to island themselves from the broader grid. Hyperscalers are not inherently opposed to interconnection—grid connection offers meaningful advantages, including backup reliability during maintenance and cost-leveling during low utilization periods. The policy challenge is ensuring those advantages are preserved so that interconnection remains the rational default and that hyperscalers are incentivized to bring reliable thermal generation to the broader ERCOT grid rather than reserving that generation only for themselves.

A critical distinction runs through this discussion, one that Batch Zero briefly wrote into ERCOT's rules before the August 2026 audit paused the process: on-site generation is not the same thing as islanding. A fully islanded facility builds its own generation and never touches the grid. A grid-connected facility

1 "Behind-the-meter" is the term Cleanview uses, and it is unclear how much they are referring to private grid capacity and how much is simply on-site capacity at a grid interconnected facility. The authors prefer more precise terms, such as private grid, hybrid, or interconnected, but retain Cleanview's terminology here for consistency with the source.

2 It should be noted that the governor also directed the Texas Commission on Environmental Quality (TCEQ) to halt permits for data centers "until the Electric Reliability Council of Texas (ERCOT) completes its audit" (Office of the Governor of Texas, 2026c).

with on-site generation—the posture formalized in Batch Zero's Withdrawal-Limited Private Use Network ("bring your own generation") pathway—interconnects, nets that on-site generation against its draw, and remains part of the system. The policy questions the two raise are nearly opposite. Islanding is a rational response to grid constraints and should remain available, but grid-connected generation is the behavior policy should actively enable, because it enables private capital to increase system capacity and economies of scale. Much of the announced pipeline of private generation is better understood as developers hedging interconnection timelines than as a permanent exit from the grid—which means the right rules can still capture it.

Once connected, ensuring data centers function as assets rather than liabilities requires contractual and technical mechanisms that translate computing flexibility into real grid services. Throttling and curtailment agreements, formal demand response participation, portable demand arrangements enabling cross-regional workload shifting, and cooling optimization strategies all offer meaningful tools for grid operators during periods of stress—though how to structure and enforce these arrangements presents a significant policy challenge addressed later in this paper.

Realizing these benefits depends on electricity price signals that accurately reflect real-time grid conditions; if data centers are insulated from true system costs, the economic incentive to flex simply does not exist. Any large-load integration policy must ensure that grid interconnection is faster and more affordable than privatization. ERCOT's PGRR 134 for dispatchable loads offers, though withdrawn in May 2026, a useful model: by combining flexibility requirements with expedited queue treatment, it demonstrated that grid interconnection and operational agility are not mutually exclusive goals (Sharma Frank, 2025). Future policy should preserve and reinforce those advantages so that bringing reliable generation to the broader ERCOT grid remains the rational default.

Yet flexibility alone is not sufficient. For data centers, grid reliability is not merely a policy preference; it is a core business requirement. Private grids meet that requirement on their own terms, and Texas benefits when they do. But the largest gains for ratepayers come when that same generation lands inside the system and pays for system upgrades that all ratepayers benefit from, which means interconnection has to remain the economically rational choice rather than the residual one. Making it so requires attracting investment in reliable generation and the infrastructure needed for a cost-effective and reliable grid.

THE ECONOMIC AND RELIABILITY NEEDS OF AI DATA CENTERS

As established earlier, realizing the grid benefits of data center flexibility depends on curtailment and demand response mechanisms that remain genuinely challenging to design and enforce. The reason is simple: AI load is not like other flexible industrial loads. The economics of AI computing create reliability requirements so stringent that conventional curtailment frameworks, traditionally designed around revenue margins per MWh, must be fundamentally rethought and reconsidered for hyperscalers.

Traditional energy-intensive industries like aluminum smelting and steel production generate between $60 and $160 of revenue per megawatt-hour of electricity consumed, with curtailment thresholds within the same $60–$160 range (Royal, 2026, p. 19). Frontier AI inference, by contrast, generates approximately $74,000 per MWh, with a blended average across AI workloads of roughly $12,500 per MWh. Its blended curtailment threshold—the price at which powering down becomes the economically rational choice—sits near $6,350 per MWh, or 127 times the roughly $50 per MWh marginal cost of a conventional gas-fired unit (Royal, 2026, p. 8). Put plainly: a smelter shuts down once power consumes nearly any of its hourly revenue; an AI data center keeps running until power consumes half.

Table 2
When It Comes to Curtailment, AI Is in a Category of Its Own

Sector	Revenue per MWh of electricity consumed	Curtailment Price
Aluminum smelting	~$60–$80/MWh	$60–$80/MWh
Steel production	~$80–$120/MWh	$80–$120/MWh
Chemical manufacturing	~$100–$160/MWh	$100–$160/MWh
Frontier AI inference	**$74,000/MWh**	**$53,650/MWh**
Blended AI workloads	**$12,500/MWh**	**$6,350/MWh**

Note: Data from Royal, H. (2026). *The compute heat rate: Quantifying AI-driven electricity price tolerance and its implications for wholesale market repricing*. SSRN. (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6322318).

The cost of an outage reflects this gulf: a single megawatt of lost capacity costs approximately $1.78 million per day at frontier-inference rates, or roughly $300,000 per day for blended workloads. An hour-long outage at a gigawatt-scale facility would cost the operator between $12.5 million and $74 million in forgone revenue, depending on whether the load is running blended or frontier inference workloads.

These economics translate directly into reliability requirements that have no precedent in ERCOT's existing framework. Tier IV data centers (the highest reliability tier) are usually designed to achieve 99.995% uptime, allowing just 26 minutes of downtime per year (Hewlett Packard Enterprise, n.d.). Meeting these thresholds requires firm, guaranteed power with predictable pricing, proximity to generation, and the ability to scale: conditions that intermittent resources alone cannot reliably provide and that a grid under stress cannot guarantee without deliberate policy design.

As previously outlined, due to AI hyperscalers such as Meta, Google, and Microsoft investing such large amounts of capital into these data centers, they expect (and shareholders require) them to achieve these revenue targets at a constant rate. This leaves these firms two options: interconnect into a grid that can and will achieve these high reliability standards (without risk of curtailment) or build their own power off-grid. While the economics of building enough generation with the necessary redundancy for maintenance and unscheduled downtime is expensive compared to on-grid power, in comparison to other capital expenditures of hyperscalers, it is marginal.

If ERCOT can maintain reliability without threat of curtailment, and provide power quickly enough, it can offer a much better deal for both hyperscalers and consumers than islanded data centers. Data centers that bring revenue for reliable generation on the grid will lower costs in the long run for all consumers (due to economies of scale) while also ensuring that the proper investments are made in resource adequacy, transmission and distribution, and grid resiliency.

The best outcome, for the facility, the grid, and ratepayers alike, is not a choice between these poles but a hybrid of them: data centers that interconnect and bring firm generation and storage with them. The facility gets the reliability and speed the queue alone cannot yet deliver; ERCOT gains capacity it can call on during scarcity rather than losing it behind a fence; and ratepayers get grid reinforcement financed by private capital instead of socialized across the rate

The policy goal should be to make grid-connected development the path of least resistance, so that data center demand leads to more reliable generation on the grid, rather than permanently off it, while also facilitating better management of existing variable generation.

base. Texas policy should treat this hybrid posture as the default to enable—not a concession to tolerate—and measure every interconnection rule against whether it makes that posture easier or harder to choose.

ERCOT's Batch Zero process, established through PGRR 145 (ERCOT, 2026e) and NPRR 1325 (ERCOT, 2026f) and approved by the PUCT in June 2026, was designed as a one-time, transitional framework that replaces ERCOT's old case-by-case study with a single, system-wide batch study—a triage mechanism for an interconnection queue that has swelled to more than 470 GW. It also carries forward the flexible interconnection principles of PGRR 134, giving dispatchable and self-supplied loads a faster path to energization (Sharma Frank, 2025). These are real improvements, but the process asks developers for substantial commitments amid substantial uncertainty: a financial security deposit of $50,000 per MW, capacity allocations that may come in below requested levels, and energization timelines that can vary by years and remain subject to curtailment obligations.

The implementation of these changes is also more uncertain due to the ongoing data center audit. ERCOT suspended the Batch Zero study before issuing project classifications, and the PUC granted the good-cause exceptions needed to depart from the original schedule. ERCOT now expects to verify roughly 250 to 300 projects by December 2026 and has said it will not meet its April 2027 deadline to deliver the batch study results (ERCOT, 2026d). More fundamentally, Batch Zero can allocate capacity on paper, but it cannot by itself build the generation and transmission needed to serve the load. It is a procedural reform—a necessary one—but it is not a substitute for the physical buildout that SB 6's reliability goals ultimately depend on, nor for permanent interconnection rules that can outlast a single batch.

It is precisely this reliability imperative that makes the interconnection question so consequential for the rest of ERCOT's customer base. A recent Bloomberg New Energy Finance analysis found that 100 GW of on-site gas-burning capacity—equivalent to 18% of all existing U.S. natural gas generation—is planned to power data centers across the country (Rissman and Gimon, 2026). Hyperscalers are not pursuing private generation out of indifference to the grid; they are pursuing it because the grid has not kept pace with the speed and reliability their operations require.

Residential electricity users in Texas would benefit tremendously if this massive thermal generation buildout was integrated into the broader ERCOT grid. Not only would reliability increase to the extremely high standards data centers require, but consumers would also benefit from reliable power and transmission brought to their grid at the expense of hyperscalers. The lesson is not that off-grid gas should be banned or discouraged; it is that when interconnection cannot match the speed and certainty hyperscalers require, the generation gets built anyway—but outside the framework where it could support ERCOT during scarcity. The policy goal should be to make grid-connected development the path of least resistance, so that data center demand leads to more reliable generation on the grid, rather than permanently off it, while also facilitating better management of existing variable generation.

POLICY RECOMMENDATIONS

The preceding analysis establishes that data centers can be either significant assets or significant liabilities to the ERCOT grid depending on how they are integrated. This section will translate that analysis into specific, actionable policy recommendations.

Recommendation 1: Reform ERCOT's Load Forecasting Methodology and Plan for Long-term Demand

ERCOT's planning challenge is a two-sided one: avoid underbuilding for facilities that will consume every megawatt they contract and likely more, while avoiding overbuilding for projects that never break ground.

An estimated 10–20% of proposed data centers ultimately get built on the schedules reflected in interconnection queue filings. ERCOT already models scenarios that count only 50%—or none—of unsigned "officer letter" load, and reduced large-load projections by roughly 50–55% in its 2025 adjusted forecast (ERCOT, 2025c; Watson, 2025). In that forecast, ERCOT reduced data-center load requests to 49.8% of submitted value and officer-letter load to 55.4% (ERCOT, 2025d). While applying a discount factor based on historical experience is a good start, we recommend an evolving, probability-weighted queue forecast rather than a static discount factor.

Here is how this refinement could work. Each project in the interconnection queue would carry a realization probability that updates as it clears observable, verifiable milestones: executed site control (ownership or lease rather than options), interconnection agreement execution with financial security posted, air and water permits, long-lead equipment procurement (transformers, turbines, and switchgear are the binding constraints), committed equity and debt financing, and construction notice-to-proceed. An early-stage filing with none of these commitments might carry a 5–10% weight; a project with an executed interconnection agreement, posted security, and equipment on order approaches certainty.

ERCOT is already beginning to collect much of this information through the Batch Zero eligibility criteria, which distinguish projects by maturity and commitment, so the incremental data burden of maintaining such a forecast is modest. Publishing the result as a probability-weighted band, updated quarterly, would give transmission planners, generators, and policymakers a shared, defensible view of demand that neither overreacts to speculative filings nor underestimates committed load.

Equally important, ERCOT should not assume that efficiency improvements will meaningfully reduce the electricity consumption of data centers over time. While individual chips and cooling systems are becoming more efficient, the history of computing suggests these gains will not lead to lower per-site demand. As power-per-rack falls, operators increase rack density, deploy more capable next-generation chips, and expand total computing capacity to chase increasing demand.

Efficiency gains have historically moderated projected consumption growth without reducing absolute consumption, and there is no reason to assume this pattern will change with the current wave of new load. ERCOT's planning models should reflect this reality by assuming that contracted load will be fully utilized, and in many cases exceeded in the future, as facilities scale into their buildout and adapt to power-intensive new chips to remain commercially viable. Striking this balance is critical to avoid unnecessary costs and ensure long-term reliability.

Recommendation 2: Market Reform to Support Firm Generation and Load Flexibility

The current ERCOT market structure inadequately compensates firm, dispatchable generation, a structural flaw that predates the data center boom but is acutely exposed by it. The solution is not to force data centers into curtailment frameworks designed for aluminum smelters or other flexible loads; it is to reform the market so that the thermal generation data centers require is financially viable to build and operate within ERCOT.

ERCOT and the Texas Legislature should pursue two parallel tracks. First, market reforms should ensure that firm, dispatchable generation—primarily natural gas—receives adequate compensation for the capacity and reliability services it provides. To accomplish this goal, either the energy market must be modified to end the practice of paying all resources, whether dispatchable or variable, the same market price, or a program must be developed to ensure all resources in the energy market meet the same reliability standard. Texas has roughly the same amount of combined gas, coal, and nuclear capacity that it did 10 years ago, despite substantial demand growth (ERCOT, n.d.). Data center load is the demand signal that can justify the upfront cost of that investment, but only if the market structure rewards it.

Natural gas is what makes the first track practical on data center timelines. It is the only firm resource that can be developed at the speed and scale this load requires, and Texas's in-state supply, pipeline network, and services workforce give it an advantage no other market can replicate. The binding constraint is not fuel but equipment: gas turbine order books now stretch years into the future, which means the projects that will serve load late this decade must reach commitment now—and developers will not commit amid regulatory uncertainty. Policy certainty around interconnection treatment, netting rules, and cost allocation is, in effect, generation policy. Structured to interconnect rather than island, this gas fleet also does double duty: serving its facility day to day while standing as capacity ERCOT can call on during emergency conditions, exactly as SB 6 contemplates.

Second, large load customers should be permitted and encouraged through price signals to self-manage grid stress through resources such as battery storage and additional on-site generation as an alternative to traditional curtailment obligations. A data center that can ride through a two-hour peak event on batteries or internal workload shifting via portable agreements will impose less stress on ERCOT than one that either curtails inference or demands uninterrupted grid power with no flexibility at all. Facilities with these redundancies in place should be given priority over those who are unable to manage load during peaks or outages.

Energy storage is the hinge that makes both tracks work and is a necessary component of the data center energy stack given the rapid demand fluctuations and high reliability needs that AI data centers present. On the generation side, batteries bridge the seconds-to-minutes gap between a grid disturbance and turbine response, allowing a gas-anchored facility to meet Tier IV reliability without overbuilding turbines for redundancy alone. On the load side, storage is what turns flexibility commitments from a workload risk into an engineering specification: a facility with hours of on-site storage can ride through a curtailment window or a peak event without touching inference workloads, which is precisely why hyperscalers will accept flexible interconnection terms they would otherwise refuse. Storage also supplies fast frequency response and other stability services the grid needs more of as thermal inertia declines. Every element of the framework this paper recommends—flexible interconnection, bring-your-own-power, demand volatility management—becomes more credible and cheaper to deliver when storage is part of the stack, and qualification and cost-allocation rules should treat co-located storage accordingly.

This dual approach of firming the grid through market reform while allowing data centers to contribute flexibility on their own operational terms aligns private investment incentives with public reliability needs more effectively than any mandatory curtailment regime could. Hyperscalers will build firm generation when the market rewards it and when doing so is cheaper than the alternative of full islanding. ERCOT's job is to make sure that calculus points toward the grid.

Recommendation 3: Direct Cost Assignment for Transmission Triggered by Large Load

Cost-causation principles are already reflected in ERCOT's interconnection procedures for dedicated facilities, but those rules should be extended and clarified so that any transmission upgrade whose primary driver is a large-load interconnection is directly assigned to that customer rather than socialized across the rate base. Financial security requirements should be structured to make that assignment enforceable even if a data center subsequently reduces its load or exits the market. Transmission and distribution utilities should also receive clear regulatory guidance on stranded asset recovery procedures for network infrastructure associated with large-load customers, thus reducing the uncertainty that currently produces conservative, inefficient planning decisions.

The PUC has moved substantially toward this framework already. Proposed 16 TAC §25.194 in Project No. 58481 requires large load customers to fund all direct interconnection costs through a contribution in aid of construction, bars recovery of those costs under any regulated rate, and applies forfeited security as a rate base offset when a customer withdraws (PUC, 2026b). The pending amendments in Project No. 58000 exclude post-effective-date interconnection costs from a TSP's cost of service and establish a twenty-year minimum billing demand for large load customers (PUC, 2026a).

What remains unresolved is the treatment of network upgrades beyond the interconnection facilities themselves. The Foundation's comments in Project No. 58481 support the non-refundable interconnection fee but raise two concerns with its structure: the $50,000 per megawatt value appears arbitrary and untethered to any clear cost-causation framework, and routing those payments straight into the rate base dilutes the fee's purpose of recovering project-specific interconnection costs and protecting against stranded infrastructure risk (PUC, 2026c). The Commission should extend direct assignment to the network upgrades large loads actually trigger, with a per-megawatt cap to prevent excessive allocation in low-interconnection years.

Recommendation 4: Incentivize Grid-Connected Data Centers Near Existing Infrastructure

Hyperscalers seeking expedited grid interconnection should be required to integrate a minimum quantity of firm capacity into the ERCOT system as a condition of that expedited interconnection (a "bring your own capacity" requirement) that benefits all ratepayers rather than just the interconnecting facility. The minimum is best expressed as a performance standard—e.g., firm capacity, verifiable by ERCOT and deliverable during scarcity events—rather than a fixed percentage of nameplate load, which invites gaming through paper capacity and penalizes configurations that deliver equivalent reliability through different mixes of generation and storage. This priority status should also be focused on data centers that site within existing high-capacity transmission corridors rather than requiring new greenfield transmission builds. Where facilities locate should remain a private decision, but policymakers should make the grid-beneficial choice the more economically attractive decision.

This is not a novel proposal. PGRR134, filed in November 2025, would have allowed a large load electing Controllable Load Resource status to be authorized for earlier energization where constraints could be mitigated by dispatching down to its Low Power Consumption level, while still being studied for full capacity on its Load Commissioning Plan timeline and subject to any reliability limits ERCOT specified (Sharma Frank, 2025). It was tabled and withdrawn in May 2026, weeks before the Commission approved the Batch Zero framework, which absorbed part of the same logic by permitting a Provisional Controllable Load Resource election and allowing on-site generation to offset a project's transmission capacity requirement. But Batch Zero is a one-time transitional process, and its accommodations expire with it; the question PGRR134 raised should be settled as standing policy, not left to the

design of each successive batch. The August 2026 suspension of Batch Zero makes the point concrete: the accommodations that let on-site generation offset a project's transmission requirement are now paused along with the batch that created them.

The model here is Texas's own refining industry: large industrial loads strengthened the grid for everyone by creating dense, efficient load centers that justified and financed transmission investment. No other state can match Texas's combination of locally accessible natural gas supply, flexible industrial siting, and an independent market structure purpose-built for large industrial customers, making a broader BYOC framework not just sound policy, but commercially viable in a way it simply would not be elsewhere.

Making this workable requires preserving the regulatory plumbing that now exists. Specifically, ERCOT and the PUC should preserve and expand the Withdrawal-Limited Private Use Network framework so that co-located generation can net against facility load; extend expedited study treatment to bring-your-own-generation configurations; and ensure co-located generation is not penalized in transmission cost allocation relative to standalone plants. The refining-industry model only works if the rules make participation cheaper than exit.

The same siting logic unlocks a second stranded asset: generation Texas has already built but cannot use. As documented earlier, ERCOT curtailed more than 8 TWh of wind and solar in 2024—roughly the annual consumption of a gigawatt-scale data center campus. A substantial amount of wind and, particularly, solar is also forcing the curtailment of gas generation that could be operating more efficiently, while suppressing prices and preventing the financing of new gas generation that is sorely needed for reliability. Critically, much of that curtailment reflects renewable generation concentrated in export-constrained zones far from load, a geographic mismatch that transmission projects—including the $33 billion 765-kV STEP—relieve only slowly and partially, but that well-sited demand can absorb directly, where the surplus already is. Ratepayers financed the transmission serving these resources; today the energy is a sunk cost producing nothing.

A data center sited at these pockets of stranded capacity completes the grid-asset loop this paper describes. It absorbs output that would otherwise be curtailed through schedulable workloads and storage, anchors its reliability with on-site gas, balances the entire stack behind a single interconnection, and returns surplus firm capacity to the grid during scarcity. None of this requires building more intermittent generation; it requires putting demand where the existing overbuild already sits. Financing should follow the user-pays principle of **Recommendation 3**. Interconnection and network upgrades tying stranded generation to large loads can be funded through revenue bonds serviced by the facilities themselves—at no cost to ratepayers or taxpayers.

Recommendation 5: Demand Volatility Management

The ERCOT grid's ability to maintain reliability depends on matching supply and demand in real time. Data centers—if properly integrated—can reduce demand volatility rather than exacerbate it. By shifting interruptible computing tasks, primarily AI training, to periods of high renewable generation and low grid stress, data centers can absorb supply volatility rather than amplify it. They can reduce the need for residential conservation calls during peak summer hours by providing a large, dispatchable demand resource that operators can draw on before residential demand becomes a reliability concern. With appropriate technical agreements, data centers can also provide ancillary services such as frequency response, functions currently supplied almost exclusively by thermal generation and energy storage. A larger grid with more diverse, flexible load lowers per-unit costs for all consumers—a direct benefit to residential ratepayers that grows as data center load scales.

What converts these capabilities from contractual promises into operating reality is orchestration: software-driven, physics-aware coordination of a facility's on-site generation, battery storage, cooling systems, and schedulable workloads as a single dispatchable resource. An orchestrated campus can shift training load into hours of renewable surplus, discharge storage into evening ramps, and shape its draw against real-time grid conditions automatically and verifiably—rather than through phone calls during an emergency.

The technology exists today; what lags is the market plumbing. ERCOT's demand response and ancillary service qualification frameworks were built around industrial loads that curtail slowly and manually, and they should be updated to recognize and compensate large loads that can respond in seconds. ERCOT already has the building blocks: Controllable Load Resources can provide ancillary services today, and the Real-Time Co-optimization market design launched in December 2025 makes fast, verifiable response more valuable than ever—qualification simply needs to scale to gigawatt campuses. A gigawatt-scale campus operated this way is not merely a smaller burden on the grid; it is one of the largest fast-responding resources on the system—serving its own load, absorbing surplus the grid cannot otherwise use, and returning firm capacity to it when scarcity hits.

CONCLUSION

The question before policymakers is not whether AI data centers will have a large impact on our grids. They are already doing so at a scale and pace without precedent in modern American history. The question is whether Texas will capture the full benefits of this investment surge, or whether regulatory missteps will saddle the grid with socialized transmission and reliability costs, leaving ordinary ratepayers exposed to the financial and reliability risks. Under the right policy conditions—conditions Texas is uniquely positioned to create—data centers will become assets to the ERCOT grid: sources of firm load, catalysts for reliable generation investment, and a long-term subsidy for residential electricity costs.

The evidence assembled here cuts sharply against the popular narrative that data centers drive up electricity bills and strain the grid. Texas and Virginia have seen the steepest increases in data center load yet have recorded minimal electricity rate increases in recent years, while California and New York, which added minimal large industrial load, saw the largest residential price hikes. The culprits behind rising bills are elsewhere: mandated retirements of reliable generation, expensive transmission buildout to support intermittent wind and solar, aging distribution infrastructure, and broader inflationary trends.

Conflating those cost drivers with data center growth obscures the diagnosis and the cure. The environmental picture is similarly distorted. Data centers' water and land footprint, while real, is modest relative to agriculture, steel manufacturing, and semiconductor fabrication—industries that face nothing approaching the level of public alarm directed at hyperscalers. Managing these externalities is achievable within existing regulatory frameworks, applied evenhandedly and without double standards. Management of far more intensive industries and consumers has existed for decades, if not centuries.

The central grid challenge is not only the data center load itself but where the power to serve it comes from and if it can provide it reliably. ERCOT's interconnection queue now tracks more than 474 GW of large load seeking connection, with 90% of it from data centers (ERCOT, 2026b). Meanwhile, less than 20% of new generation requests are firm thermal generation, including natural gas (ERCOT, 2026c). If that imbalance persists, Texas will have invited the largest industrial load growth in its history onto a grid increasingly dependent on weather-driven generation that will be especially vulnerable to unpredictable winter storms like Winter Storm Uri.

The alternative is available and within reach. Hyperscalers usually prefer grid interconnection as it offers reliability and cost advantages over building private power plants. As discussed above, the on-site generation-and-storage layer that a grid-connected facility adds runs roughly $1.5–7 million per MW, depending on generation mix and reliability target—a modest increment against the roughly $35–40 million per MW all-in cost of a large AI data center, which is dominated by GPUs and cooling. Fully islanded, renewable-heavy configurations sit at the high end of that range; the gas-anchored, grid-connected builds hyperscalers actually favor sit at the low end.

The economics of interconnection already favor the grid, but grid planners, transmission utilities, and policymakers are not keeping up with the speed of development that hyperscalers need. Hence the need for the private grid option. Policy needs to preserve and reinforce the advantage of grid economics, not undermine it through slow queues, uncertain cost assignment, and curtailment frameworks that do not account for the $74,000 per MWh economics of AI inference load (Royal, 2026, p. 10).

Texas has the tools to lead, and the five recommendations in this paper can help extend Texas's leadership.

1. ERCOT should refine its load forecasting methodology using project-specific feasibility indicators to avoid overbuilding for projects that never materialize, while planning for full utilization of facilities that do get built.

2. The PUC and the Texas Legislature should work together to reform the wholesale market to adequately compensate firm, dispatchable thermal generation while allowing data centers to self-manage grid stress through on-site generation and storage; aligning private investment incentives with public reliability needs is more effective than curtailment.

3. Transmission costs triggered by large-load interconnection should be directly assigned to the customers that cause them, with financial security requirements that keep that assignment enforceable if a facility reduces its load or exits, preventing cost shifting onto ratepayers.

4. Hyperscalers seeking grid interconnection who agree to co-develop a minimum quantity of firm generation should receive expedited permitting and queue priority. A bring-your-own-power incentive like this is commercially viable in Texas in a way that it is not in many other places. Siting incentives should point these facilities toward existing corridors and the state's stranded generation, funded on a user-pays basis.

5. Data centers should be integrated as active grid participants that shift interruptible workloads to periods of low grid stress, provide ancillary services such as frequency response, and reduce demand volatility. Orchestrating on-site generation, storage, and workloads as a single resource lets these facilities absorb surplus energy and return firm capacity to the grid, ultimately lowering per-unit costs for all consumers in the process.

The companies building AI are making trillion-dollar bets on where the power will be, and Texas has a unique opportunity to be a global leader in the most consequential technology of the 21st century. The question is whether this surge of investment can be done in such a way that all Texans will benefit from it, instead of benefiting only a few while imposing costs on many. Setting the right policies, particularly regarding the ERCOT grid, is essential to enable Texas entrepreneurs, workers, and taxpayers to all flourish from this wave of innovation and investment. ■

REFERENCES

Alphabet. (2026). *Alphabet announces fourth quarter and fiscal year 2025 results.* https://s206.q4cdn.com/479360582/files/doc_financials/2025/q4/2025q4-alphabet-earnings-release.pdf

Amazon.com, Inc. (2026). *Amazon.com announces fourth quarter results.* https://ir.aboutamazon.com/news-release/news-release-details/2026/Amazon-com-Announces-Fourth-Quarter-Results/

Arnon, A. (2025). *The projected impact of generative AI on future productivity growth.* Penn Wharton Budget Model. https://budgetmodel.wharton.upenn.edu/p/2025-09-08-the-projected-impact-of-generative-ai-on-future-productivity-growth/

Associated Builders and Contractors. (2026). *ABC: Construction industry must attract 349,000 workers in 2026 despite macroeconomic headwinds.* https://www.abc.org/News-Media/News-Releases/abc-construction-industry-must-attract-349000-workers-in-2026-despite-macroeconomic-headwinds

Avelar, V., Donovan, P., Lin, P., Torell, W., & Torres Arango, M. A. (2023). *The AI disruption: Challenges and guidance for data center design* (White Paper 110). Schneider Electric. https://media.datacenterdynamics.com/media/documents/Schneider_Electric_-_Modernization_WP110_V1.1_EN.pdf

Beer, L., Chan, N., Patel, K., Picciano, P., Ramirez, S., Riu, I., & Shah, A. (2026). *Understanding the drivers of rising electricity rates and the role of data centers*. Energy and Environmental Economics. https://www.ethree.com/wp-content/uploads/2026/05/Understanding-the-Drivers-of-Rising-Electricity-Rates-and-the-Role-of-Data-Centers_E3-2026.pdf

Bennett, B. (2026). *An economic assessment of the 765-kV Strategic Transmission Expansion Plan.* Texas Public Policy Foundation. https://www.texaspolicy.com/an-economic-assessment-of-the-765-kv-strategic-transmission-expansion-plan/

Bithell, C. (2025). *Water usage in semiconductor manufacturing to double by 2035.* IDTechEx. https://www.idtechex.com/en/research-article/water-usage-in-semiconductor-manufacturing-to-double-by-2035/32746

Bloom Energy. (2026). *2026 data center power report: When power defines growth.* https://www.bloomenergy.com/wp-content/uploads/2026-power-report.pdf

BloombergNEF. (2025). *AI and the power grid: Where the rubber meets the road.* https://about.bnef.com/insights/clean-energy/ai-and-the-power-grid-where-the-rubber-meets-the-road/

Bluefield Research. (2025). *Water losses cost U.S. utilities US$6.4 billion annually.* https://www.bluefieldresearch.com/ns/water-losses-cost-u-s-utilities-us6-4-billion-annually/

Bontigui, J., & Patel, D. (2026). *From tokens to burgers: A water footprint.* SemiAnalysis. https://newsletter.semianalysis.com/p/from-tokens-to-burgers-a-water-footprint

Burns, T. (2026). *AI boom: Big Tech capital expenditures now seen topping $1 trillion in 2027.* CNBC. https://www.cnbc.com/2026/04/30/ai-boom-big-tech-capital-expenditures-now-seen-topping-1-trillion-in-2027-.html

Cobler, P. (2026, August 14). *Texas will audit up to 300 projects, mostly data center proposals.* The Texas Tribune. https://www.texastribune.org/2026/08/14/texas-data-center-approval-pause-ercot-power-grid/

Cushman & Wakefield. (2026). *Americas data center update.* https://www.cushmanwakefield.com/en/insights/americas-data-center-update

Data Center Map. (2026). *USA data centers.* https://www.datacentermap.com/usa/

Deeter, B., & Bondy, S. (2026). *Inside the biggest bet in corporate history.* Bessemer Venture Partners. https://www.bvp.com/atlas/inside-the-biggest-bet-in-corporate-history

Devasia, A. (2026*). Data center energy consumption statistics & data (2026).* The Network Installers. https://thenetworkinstallers.com/blog/data-center-energy-consumption-statistics/

Electric Reliability Council of Texas (ERCOT). (n.d.). *Capacity changes by fuel type charts July 2026 in Resource adequacy.* Retrieved August 26, 2026, from https://www.ercot.com/gridinfo/resource

Electric Reliability Council of Texas (ERCOT). (2025a, November 7). *January 2026 Monthly Outlook for Resource Adequacy (MORA).* https://www.ercot.com/files/docs/2025/11/07/MORA_January2026.pdf

Electric Reliability Council of Texas (ERCOT). (2025b, July 17). *New renewable generation record: June 21, 2025* [Analysis]. https://www.ercot.com/files/docs/2025/07/17/062125_RenewableRecord_Public.pdf

Electric Reliability Council of Texas (ERCOT). (2025c, February 12). *Capacity, demand and reserves report: December 2024*. https://www.ercot.com/files/docs/2025/02/12/CapacityDemandandReservesReport_December2024.pdf

Electric Reliability Council of Texas (ERCOT). (2025d, April 8). *2025 long-term load forecast report*. https://www.ercot.com/files/docs/2025/04/08/ERCOT-2025-Long-Term-Load-Forecast-Report.pdf

Electric Reliability Council of Texas (ERCOT). (2026a). *Capacity changes by fuel type charts September 2026* in *Resource adequacy* (webpage). Retrieved October 22, 2026, from https://www.ercot.com/gridinfo/resource

Electric Reliability Council of Texas (ERCOT). (2026b, July 29). *ERCOT Update: Senate Committee on Business & Commerce.* https://www.ercot.com/files/docs/2026/07/29/ERCOT-Senate-July-29-Panel-1-Assessing-The-Grid.pdf

Electric Reliability Council of Texas (ERCOT). (2026c, August 3). *GIS_Report_Jul2026.* https://www.ercot.com/mp/data-products/data-product-details?id=PG7-200-ER

Electric Reliability Council of Texas (ERCOT). (2026d, August 3). *Update regarding Batch Zero timelines* (Market Notice M-A080326-01). https://www.ercot.com/services/comm/mkt_notices/M-A080326-01

Electric Reliability Council of Texas (ERCOT). (2026e, June 18).

Batch Zero Process for Large Load Interconnections. https://www.ercot.com/mktrules/issues/PGRR145#summary

Electric Reliability Council of Texas (ERCOT). (2026f, June 18). *Related to PGRR145, Batch Zero Process for Large Load Interconnections.* https://www.ercot.com/mktrules/issues/NPRR1325#summary

Fleury, M., & Jimenez, N. (2025). *'I can't drink the water' – life next to a US data centre.* BBC News. https://www.bbc.com/news/articles/cy8gy7lv448o

Glock, J. (2026). *The surprising heart of the data-center boom*. City Journal. https://www.city-journal.org/article/loudoun-county-virginia-data-centers-construction

HB 1500. Enrolled. 88th Texas Legislature. Regular. (2023). https://capitol.texas.gov/tlodocs/88R/billtext/pdf/HB01500F.pdf

HB 3970. Engrossed. 89th Texas Legislature. Regular. (2025). https://capitol.texas.gov/tlodocs/89R/billtext/pdf/HB03970E.pdf

Hering, G., & Dlin, S. (2025). *Data center grid-power demand to rise 22% in 2025, nearly triple by 2030.* S&P Global. https://www.spglobal.com/energy/en/news-research/latest-news/electric-power/101425-data-center-grid-power-demand-to-rise-22-in-2025-nearly-triple-by-2030

Hewlett Packard Enterprise. (n.d.). *What are data center tiers?* Retrieved October 22, 2026 from https://www.hpe.com/us/en/what-is/data-center-tiers.html

James, K. (2024). *The water challenge for semiconductor manufacturing: What needs to be done?* World Economic Forum. https://www.weforum.org/stories/2024/07/the-water-challenge-for-semiconductor-manufacturing-and-big-tech-what-needs-to-be-done/

Kok, E., Rauer, J., Sachdeva, P., Pikul, P., Sutton, D., & Hasrouni, R. (2025). *Scaling bigger, faster, cheaper data centers with smarter designs.* McKinsey & Company. https://www.mckinsey.com/industries/private-capital/our-insights/scaling-bigger-faster-cheaper-data-centers-with-smarter-designs

Lazard. (2025). *Levelized cost of energy+ (LCOE+)* (18th ed.). https://www.lazard.com/research-insights/levelized-cost-of-energyplus-lcoeplus/

Martucci, B. (2025). *Nuclear specialist Oklo inks data center partnerships with Vertiv, Liberty Energy*. Facilities Dive. https://www.facilitiesdive.com/news/nuclear-specialist-oklo-inks-data-center-partnerships-vertiv-liberty/756159/

Michael, A., & Cottier, B. (2026). *Servers account for 60% of the total cost of ownership of a one-gigawatt AI data center*. Epoch AI. https://epoch.ai/data-insights/ai-datacenter-cost-breakdown

Modo Energy. (2025). *The curtailment crisis: Saving wind and solar investments in ERCOT.* https://modoenergy.com/research/en/ercot-curtailment-crisis-solar-wind-data-battery-colocated-trends-maps-texas

Mozumdar, A. (2025). *Water-tech nexus: Technology's growing water footprint, part 1.* Medium. https://arvindmozumdar.medium.com/water-tech-nexus-technologys-growing-water-footprint-part-1-82a70031cf71

Meta Platforms, Inc. (2026). *Meta reports fourth quarter and full year 2025 results.* https://investor.atmeta.com/investor-news/press-release-details/2026/Meta-Reports-Fourth-Quarter-and-Full-Year-2025-Results/default.aspx

Mey, A. (2026, August 3). *Hourly peak load in ERCOT set a new record, exceeding 91 GW on July 22*. U.S. Energy Information Administration (EIA). https://www.eia.gov/todayinenergy/detail.php?id=67906

Microsoft. (2025). *Annual report 2025.* https://www.microsoft.com/investor/reports/ar25/index.html

Milyutin, Y. (2026). *Nvidia Blackwell Explained: What It Actually Demands From Your Data Center*. Moduledge. https://www.moduledge.com/blog/nvidia-blackwell

Morgan Stanley Institute. (2026). *Energy Markets Race to Solve the AI Power Bottleneck.* Morgan Stanley. https://www.morganstanley.com/insights/articles/powering-ai-energy-market-outlook-2026

Munis, J. (2026). *Texas is the new capital of the Fortune 500—taking California's crown.* Fortune. https://fortune.com/2026/06/03/which-state-has-the-most-fortune-500-companies-texas-2026/

National Conference of State Legislatures. (2026). *Which states are banning data centers?* https://www.ncsl.org/fiscal/which-states-are-banning-data-centers

National Electrical Manufacturers Association. (2026). *New NEMA study shows widening gap between surging power demand and grid capacity.* Make It Electric. https://www.makeitelectric.org/newsroom/news/new-nema-study-shows-widening-gap-between-surging-power-demand-and-grid-capacity/

NVIDIA. (2026). *NVIDIA announces financial results for fourth quarter and fiscal 2026.* https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-fourth-quarter-and-fiscal-2026

Office of the Texas Governor. (2026a). *Governor Abbott directs PUC and ERCOT to shield Texans from data center infrastructure costs.* https://gov.texas.gov/news/post/governor-abbott-directs-puc-and-ercot-to-shield-texans-from-data-center-infrastructure-costs

Office of the Texas Governor. (2026b, August 3). *Governor Abbott directs comprehensive data center audit*. https://gov.texas.gov/news/post/governor-abbott-directs-comprehensive-data-center-audit

Office of the Texas Governor. (2026c, September 21). *Governor Abbott Directs TCEQ To Halt Data Center Permits. https://gov.texas.gov/news/post/governor-abbott-directs-tceq-to-halt-data-center-permits*

Oracle Corporation. (2026). *Oracle announces record Q4 and FY 2026 results driven by cloud infrastructure & cloud applications.* https://investor.oracle.com/investor-news/news-details/2026/Oracle-Announces-Record-Q4-and-FY-2026-Results-Driven-by-Cloud-Infrastructure--Cloud-Applications/default.aspx

Pyle, T. J., Stein, K., & Stevens, A. (2025). *Blue states, high rates*. Institute for Energy Research. https://www.instituteforenergyresearch.org/the-grid/blue-states-high-rates/

Public Utility Commission of Texas (PUC). (2025, December 18). *Order adopting new 16 TAC §25.65 in Project No. 58198.* https://interchange.puc.texas.gov/Documents/58198_77_1567651.PDF

Public Utility Commission of Texas (PUC). (2026a). *Project No. 58000 filing.* https://interchange.puc.texas.gov/Documents/58000_6_1662424.PDF

Public Utility Commission of Texas (PUC). (2026b). *Project No. 58481 filing.* https://interchange.puc.texas.gov/Documents/58481_122_1600475.PDF

Public Utility Commission of Texas (PUC). (2026c). *Project No. 58481 comments.* https://interchange.puc.texas.gov/Documents/58481_190_1624462.PDF

Reed, M., & Bennett, B. (2026). *Forecasting the Future of the ERCOT Grid, Part 1: The Ongoing Winter Reliability Challenge.* Texas Public Policy Foundation. https://www.texaspolicy.com/wp-content/uploads/2026/06/2026-05-LP-Future-ERCOT-Grid-Part1-ReedBennett-1.pdf

Rissman, J. & Gimon, E. (2026). *Behind-the-meter data center gas plants will raise US energy bills*. Utility Dive. https://www.utilitydive.com/news/data-centers-raise-energy-bills-not-for-reason-you-think/822205/

Royal, H. (2026). *The compute heat rate: Quantifying AI-driven electricity price tolerance and its implications for wholesale market repricing.* SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6322318

Rudegeair, P., & Santilli, P. (2026, August 17). Why Big Tech's AI spending is $3 trillion higher than it seems. *The Wall Street Journal*. https://on.wsj.com/4qEIFxL

SB 6. Enrolled. 89th Texas Legislature. Regular. (2025). https://capitol.texas.gov/tlodocs/89R/billtext/pdf/SB00006F.pdf

Schipper, M., & Hodge, T. (2025). *After more than a decade of little change, U.S. electricity consumption is rising again.* U.S. Energy Information Administration (EIA). https://www.eia.gov/todayinenergy/detail.php?id=65264

Sharma Frank, A. (2025). *134PGRR-01 Interconnection Studies Reform for Dispatchable Loads 110125* in PGRR 134: Key Documents (webpage). Electric Reliability Council of Texas. https://www.ercot.com/mktrules/issues/PGRR134#keydocs

Shehabi, A., Smith, S. J., Hubbard, A., Newkirk, A., Lei, N., Siddik, M. A. B., Holecek, B., Koomey, J., Masanet, E., & Sartor, D. (2024). *2024 United States data center energy usage report.* Lawrence Berkeley National Laboratory. https://escholarship.org/uc/item/32d6m0d1

Smith, S., Hubbard, A., Newkirk, A., Ganeshalingam, M., Holecek, B., Sartor, D., et al. (2026) *United States Data Center Energy Usage Report: 2025 Update.* Lawrence Berkeley National Laboratory. https://escholarship.org/uc/item/33m6w3x0

Tan, E. (2025). *Their water taps ran dry when Meta built next door.* The New York Times. https://www.nytimes.com/2025/07/14/technology/meta-data-center-water.html

Texas Produced Water Consortium. (2022). *Beneficial Use of Produced Water in Texas: Challenges, Opportunities and the Path Forward* Texas Tech University. https://www.depts.ttu.edu/research/tx-water-consortium/downloads/22-TXPWC-Report-Texas-Legislature.pdf

The White House. (2026). *Ratepayer protection pledge.* https://www.whitehouse.gov/releases/2026/03/ratepayer-protection-pledge/

Thomas, M. (2026a). *Bypassing the Grid: How Data Center Developers Are Building Their Own Power Plants.* Cleanview. https://cleanview.co/reports/behind-the-meter-data-centers

Thomas, M. (2026b). *OpenAI's Stargate data centers are taking longer and costing more than its competitors'*. Distilled. https://www.distilled.earth/p/openais-stargate-data-centers-are

United States Steel Corporation. (2025). *Nature: Managing Nature at U.S. Steel.* https://sustainability.ussteel.com/planet/nature

Utility Dive. (2026, August). *ERCOT aims to complete Texas governor's data center audit by December.* https://www.utilitydive.com/news/ercot-texas-puc-data-center-audit/828472/

Virginia Department of Environmental Quality. (2022). *Agency background document: Proposed repeal of Virginia's participation in the Regional Greenhouse Gas Initiative (9VAC5-140).* Virginia Regulatory Town Hall. https://townhall.virginia.gov/L/GetFile.cfm?File=1%5C6082%5C9879%5CAgencyStatement_DEQ_9879_v1.pdf

Virginia State Corporation Commission. (2026). *Petition of Virginia Electric and Power Company, For approval to reinstate and update the rate adjustment clause designated Rider RGGI, pursuant to § 56-585.1 A 5 e of the Code of Virginia.* https://www.scc.virginia.gov/docketsearch/DOCS/8cn%2401!.PDF

Walsh, N. (2026). *Powering the next wave of AI: Expanding capacity with our new datacenter in Pecos.* Microsoft.https://blogs.microsoft.com/blog/2026/06/22/powering-the-next-wave-of-ai-expanding-capacity-with-our-new-datacenter-in-pecos/

Watson, M. (2025, June 5). *Texas regulators OK 'good cause exception' for cutting power grid's load forecast.* S&P Global Commodity Insights. https://www.spglobal.com/commodity-insights/en/news-research/latest-news/electric-power/060525-texas-regulators-ok-good-cause-exception-for-cutting-power-grids-load-forecast

Watten, A., Bistline, J., & Blanford, G. (2026). *Have data centers raised your electric bill? Causal evidence from the United States.* arXiv. https://arxiv.org/pdf/2606.19777

Weaver, L. E., Henderson, R., & Blieszner, J. (2016). *Potential vulnerability of US petroleum refineries to increasing water temperature and/or reduced water availability*. U.S. Department of Energy. https://www.energy.gov/sites/prod/files/2016/03/f30/US%20DOE%20Refinery%20Water%20Study.pdf

Wilkins, J. (2025). *Amazon data center linked to cluster of rare cancers.* Futurism. https://futurism.com/artificial-intelligence/amazon-data-center-oregon

World Bank Group. (2026). *GDP (current US$).* https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

ABOUT THE AUTHORS



Brent Bennett, Ph.D., is the policy director for Life:Powered, an initiative of the Texas Public Policy Foundation to raise America's energy IQ and promote human flourishing through energy freedom. Dr. Bennett is responsible for Life:Powered's research and policy development, leading efforts to roll back electricity subsidies, end electric vehicle subsidies and mandates, stop discrimination against responsible energy producers, and promote grid reliability.

Dr. Bennett has an M.S.E. and Ph.D. in materials science and engineering from the University of Texas at Austin and a B.S. in physics from the University of Tulsa. His graduate research focused on advanced chemistries for utility-scale energy storage systems. Prior to joining the Foundation, Dr. Bennett worked for a startup company selling carbon nanotubes to battery manufacturers.



Joshua L. Loucks is a policy analyst for Life:Powered, an initiative of the Texas Public Policy Foundation to raise America's energy IQ and to promote policies that ensure access to responsible, reliable, and affordable energy. As part of the Life:Powered team, Joshua is responsible for research and policy development on data centers, grid reliability, and energy economics, and for advocating those policies with the Texas Legislature, the Public Utility Commission of Texas, and ERCOT.

Joshua has an M.A. in Economics from George Mason University and a B.A. in Political Science with a concentration in Legal Studies from Virginia Tech. He most recently served as a political appointee at the U.S. Department of Energy, where he was special assistant to the Deputy Secretary and Under Secretary of Energy. He has also worked for several other organizations in energy and economic policy, including the Cato Institute, the Mercatus Center at George Mason University, and the Heritage Foundation. At Cato, he was a research associate in energy and tax policy where he co-authored reports on the Inflation Reduction Act, federal permitting, and government overreach. He was a Mercatus MA Fellow during his graduate studies, and his thesis examined the economics of nuclear energy.



Kiersten Saathoff is the CEO and co-founder of HyperBase, an Austin-based company whose mission is to make reliable power a given for the reindustrialization of America and the technologies that come next. HyperBase leads site validation, campus design, development, and energy orchestration for data center projects. It works with a network of industry partners to bring AI infrastructure online in ways that strengthen the grid and the communities that host it. Kiersten leads the company's strategy and business development, aligning data center and behind-the-meter power developers and operators, grid operators and regulators, and institutional capital.

Born and raised in Austin, Kiersten has a B.B.A. and an M.P.A. from the McCombs School of Business at the University of Texas at Austin. A private equity-trained operator, she has deep expertise in utility-scale renewables and distributed energy resources. She has served as an executive and advisor to portfolio companies of Vista Equity Partners and Goldman Sachs.

Texas Public
POLICY FOUNDATION

901 Congress Avenue | Austin, Texas 78701 | (512) 472-2700 | www.TexasPolicy.com