Powering AI: A Nationwide Review of State Strategies for Sustainable Data Center Integration[1]

Joana M. Abreu, VEIC

ABSTRACT

As hyperscalers and colocation data centers expand to meet growing AI demands, interconnection and power availability have become urgent infrastructure and social challenges. This research presents a nationwide review of state-level strategies for sustainable data center integration, focusing on regulatory frameworks that address grid reliability, cost-sharing, and colocation. Beyond technical interconnection, the paper emphasizes how regulatory design can support long-term alignment between data centers and host communities. It explores how permitting, land use policies, and utility planning address affordability, environmental stewardship, and equitable land access.

The review identifies the following regulatory trends across the country: the restructuring of tax incentives to condition on measurable hosting communities and ratepayer benefit; the emergence of development moratoriums pending electric grid impact, environmental and social and economic impact studies; the migration of environmental performance standards from voluntary commitments into binding regulatory conditions aligned with interconnection reform; the evolution of cost allocation frameworks requiring large loads to bear the incremental system costs their demand imposes; and growing policy support for behind-the-meter and co-located generation models that reduce pressure on shared transmission infrastructure. Together, these trends describe a regulatory arc moving from permissive development toward comprehensive integration standards that treat data center growth as a matter of public infrastructure governance. The paper concludes that while the regulatory tools currently being deployed are necessary, whether the pace of institutional adaptation can match the pace of technological and commercial change remains an open question.

Introduction

Artificial intelligence (AI) is rapidly transforming how businesses operate, altering workforce dynamics. The field has undergone significant shifts driven by algorithmic innovation, data availability, and advances in computing power. AI systems have evolved to learn from massive datasets using neural networks with millions of parameters, bringing applications such as autonomous vehicles, large language models, and real-time translation into everyday life.

This transformation carries a substantial physical footprint: hyperscalers and colocation data centers are the primary infrastructure behind AI training and deployment. They have driven a surge in electricity demand after decades of relatively flat growth (Shehabi et al. 2024). Green et al. 2024 projects that data centers will require more than 50 GW of additional capacity by 2030, growing from 25 GW today to over 80 GW, an infrastructure investment exceeding half a trillion dollars. States have long competed to attract these facilities through tax incentives and streamlined permitting in some cases forfeiting tax revenue to do so (Reali, 2025; Green et al. 2025).

[1] According to research conducted in March, 2026

Policymakers and regulators are increasingly concerned that the infrastructure investments required to serve large-scale data centers will drive up costs for ratepayers (particularly residential customers) and strain systems that were not designed for this pace of growth.

Some regions are already feeling the pressure. For example, CenterPoint Energy in Texas reported a 700% increase in large load interconnection requests in a single year, with its queue jumping from 1 GW to 8 GW between late 2023 and late 2024 (Wolf 2024). In the Pennsylvania-New Jersey-Maryland (PJM) region, which serves 67 million people across 13 states and the District of Columbia, capacity auction prices have risen nearly tenfold in three years, from roughly $29/MW-day in 2024 to $333/MW-day in 2025, reaching FERC's emergency price cap in the last auctions. PJM's market monitor attributes this spike primarily to data center load growth (Kunkel 2025). Critically, the 2027/28 auction (run in 2025) marked the first time PJM failed to procure sufficient capacity to meet its reliability standard, a warning sign that the pace of data center buildout is outrunning the market's ability to respond. States are adapting to the pressure on load growth through interconnection policy, utility planning, land use regulation, and community engagement. This paper examines emerging regulatory strategies and identifies approaches that are proving most effective at integrating large-scale AI infrastructure without compromising grid reliability, affordability, or equity.

Background: The Infrastructure and Impacts of AI Computing

AI workloads are powered by a tiered ecosystem of data center facilities, each distinguished by scale, ownership model, and function.

Hyperscalers and colocation data centers

Highly standardized facilities, ranging from 100 MW to several gigawatts in total load, hyperscalers are operated by major technology companies (e.g. Google, Microsoft, Meta, xAI, Apple, among others) to train and deliver computing power, storage, and network services at massive scale. Some hyperscaler definitions place the lower end at 20 MW, depending on whether one is measuring processing capacity or total facility load. Colocation facilities are operated by third-party providers that supply space, power, cooling, and connectivity while customers place and manage their own servers or lease managed infrastructure. This category encompasses both wholesale colocation datacenters, which serve large enterprise clients, and retail colocation with smaller footprints. DOE and NRL place colocation facilities in the range of 5 to 100 MW (Garg et al. 2025). The power demands of these facilities are not uniform: in AI data centers, the load fluctuates throughout the day in ways that conventional data centers historically did not, because AI data centers have two major phases: training (learning) and inference (querying the model) Training workloads are highly energy intensive, involving long-duration computation with relatively predictable power draw across specialized hardware, requiring approximately 8 to 12 kW per GPU node and 40 to 120 kW per rack, with frontier systems exceeding 100 kW per rack. Inference workloads, which serve end users once a model is deployed, are comparatively lighter at 2 to 10 kW per node and 10 to 40 kW per rack, but present more variable demand across a heterogeneous hardware mix (Garg et al. 2025). Inference accounts for the majority of lifecycle energy use in deployed AI systems, and because it runs continuously across millions of transactions, its cumulative energy needs exceed that of training. Despite the perception that training dominates energy consumption, once a model reaches scale, most energy consumption needs and volatility, is attributable to inference. Overall, it accounts for up to 90% of machine learning energy consumption in deployed systems (Desislavov et al. 2023). At facility scale, training clusters commonly range from 100 MW to multiple gigawatts, while inference clusters range from sub-megawatt installations to over 100 MW hyperscale deployments (Garg et al. 2025).

Speed to Power

The most important operational challenge facing data center developers is speed to power: the ability to obtain reliable grid connection quickly enough to meet operational and competitive timelines. Data centers can typically be developed in one to two years, while the generation and transmission infrastructure required to serve them takes far longer. New generation interconnection takes an average of five years to complete and more than 70% of interconnection projects are withdrawn (NCSL 2025), while major transmission lines average approximately a decade from conception to energization (Sallam 2025).

The timeline from interconnection application to commercial operation in PJM, for example, has risen from an average of less than two years in 2008 to over eight years today (Goldsmith et al. 2025). The tools available to grid planners were built for extending power lines to new neighborhoods or upgrading equipment as communities grow, not for analyzing dozens of simultaneous service requests of several MW each. As a result, planners and engineers are overwhelmed, and most utilities and independent system operators and regional transmission organizations (ISO/RTOs) are still developing formal procedures for interconnecting large loads (Hough 2025). This mismatch between the commercial pace of data center development and the institutional pace of grid planning has become a central friction point in every region of the country: as power transmission becomes constrained in primary markets data center projects are moving to secondary and emergent states (Green et al. 2024). Figure 1, depicts the geographic distribution of hyperscale and colocation projects in 2024.

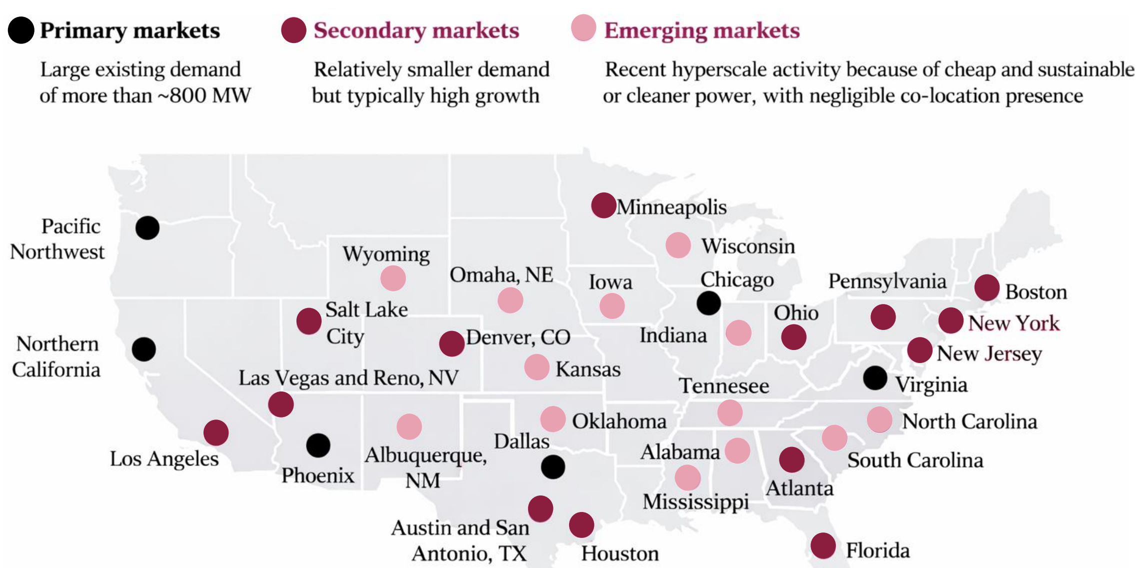

Figure 1: Data center development activity across the United States, showing the geographic distribution of hyperscale and colocation projects. Source: Green et al. (2024) McKinsey & Company

Power Usage Effectiveness

Power Usage Effectiveness (PUE) measures the ratio of total facility energy to the energy consumed by IT equipment alone. A perfect PUE of 1.0 would indicate that all power delivered to a facility goes directly to computing, with zero overhead. In practice, the industry-wide average PUE is approximately 1.56, a significant improvement from 2007 when average PUE exceeded 2.5. Google for example, reported fleet-wide PUE of 1.09 across its global data centers, using approximately 84% less overhead energy per unit of IT equipment energy than the industry average (Google, 2025), but on average, power efficiency in the industry has plateaued in recent years. Small differences in PUE matter at scale[2]; however, improved efficiency has not translated into lower aggregate electricity consumption because as computing becomes more efficient and less expensive per operation, demand for computing grows faster than efficiency gains, resulting in higher total energy use. This is known as the "Jevons Paradox", recognizing that without additional intervention, efficiency gains are overwhelmed by surging demand, and total resource consumption continues to rise even as each individual compute operation becomes more efficient (Kok et al. 2025)[3].

Data Center Flexibility

The same characteristics that make large data centers a challenge for grid operators also make them potentially valuable grid assets when demand is responsive to grid conditions. Data centers do not operate consistently at full capacity. They typically maintain utilization rates of around 80%, leaving meaningful operational headroom. A Duke University study found that small curtailments throughout the year, could unlock nearly 100 GW of extra grid capacity without requiring the multi-year infrastructure buildout for datacenter development (Norris et al. 2025). Training often runs continuously for days or weeks and can tolerate brief performance slowdowns without disrupting end users. Inference workloads serving real-time queries are more sensitive, but processes can be prioritized.

This heterogeneity means that a well-orchestrated datacenter can shed load during periods of grid stress while maintaining quality of service for its most sensitive workloads. There is also an important distinction between load factor and utilization rate that shapes how this flexibility is calculated and regulated. If a customer's peak load is 85% of its rated capacity and its load factor[4] is 90%, its effective utilization rate is 76.5%, not 85%[5].

These distinctions should be considered in flexible interconnection frameworks, but information needs to be available. For example, LBNL (Shehabi et al. 2024) has documented a significant lack of primary operational data from data center operators, who generally do not report IT workload types, hardware configurations, or system utilization to utilities or grid operators. Consequently, this data gap limits the ability of grid planners to model and rely upon data center flexibility as a grid resource.

---

[2] A facility drawing 500 MW with a PUE of 1.56 consumes 280 MW in overhead, while a facility achieving 1.09 consumes only 45 MW in overhead for the same compute output

3 For example, following DeepSeek's release in early 2025, which reduced the cost of AI inference, Microsoft CEO Satya Nadella publicly noted the Jevons Paradox at work, as Meta immediately raised its 2025 AI infrastructure spending in response to the efficiency breakthrough rather than reducing it (Luccioni et al. 2025).

[4] Average load to peak load over a unit of time, usually a month or a year. Data center load factors range between 0.5 to 0.9 depending on usage patterns.

[5] Average Load = "Peak Load" x "Load Factor"

The Phoenix Demonstration and Aurora AI Factory

Field demonstrations have begun to close the gap between theoretical flexibility potential and operational reality. In May 2025, as part of EPRI's DCFlex Initiative at an Oracle data center in Phoenix, Arizona, Emerald AI's Conductor platform orchestrated 256 NVIDIA GPUs across two peak demand events. One of the events occurred on May 1 as a response to an Arizona Public Service (APS) system peak, and the other occurred on May 3 responding to a Salt River Project (SRP) system peak. The pilot achieved 25% power reduction over three-hour periods while meeting utility-set performance requirements, without requiring any hardware retrofits or energy storage systems (Colangelo et al. 2025). The platform classified workloads into three flexibility tiers based on user tolerance for throughput deviations, ranging from up to 10% reduction for the least flexible jobs to up to 50% slowdown for fully flexible batch workloads, while excluding high-priority jobs from flexibility management entirely. Rather than shutting down entire data halls, the platform slowed less critical tasks, paused batch processing that was not time-sensitive, and rescheduled flexible workloads, with control algorithms continuously optimizing based on power-performance trade-offs of individual jobs. In October 2025, a coalition including NVIDIA, Emerald AI, EPRI, Digital Realty, and PJM announced the world's first power-flexible AI Factory: the 96-MW Aurora facility in Manassas, Virginia, serving both as a reference implementation and a certification standard for power-flexible AI infrastructure. The coalition estimates that if adopted nationwide, the reference design could unlock approximately 100 GW of capacity, on the existing system without requiring new transmission infrastructure.

Water Usage Effectiveness

Water Usage Effectiveness (WUE) measures the volume of water consumed by a data center facility divided by the IT energy delivered to computing equipment, expressed in liters per kilowatt-hour (L/kWh). Like PUE, a lower WUE indicates greater efficiency. The industry-wide average WUE ranges between 0.25 L/kWh to 1.8 L/kWh, with best-in-class facilities using closed-loop cooling systems in mild climates achieve below 0.5 L/kWh (Yañez-Barnuevo 2024).

Hyperscalers have reported fleet-average WUE figures approaching or below 1.0 L/kWh at their most efficient locations, though global averages are higher. The water demands of large-scale data centers are substantial[6]: In the United States, data centers consume approximately 1.7 billion liters of water per day, and less than one third of data center operators measure their water consumption at all, creating significant gaps in the data available to regulators and communities making siting and permitting decisions (Mytton 2021).Water consumption depends on siting and location, which affect cooling needs, measured by WUE. But the water footprint of a datacenter can also be measured indirectly, depending on how electricity serving the unit is generated in each region. This impact can vary significantly depending on the power mix and generation technologies in use in the region. A data center sited in a water-stressed region and powered by thermoelectric generation carries a compounded water burden higher than those captured by WUE metrics.

[6] A medium-sized data center can consume up to roughly 110 million gallons of water per year for cooling, equivalent to the annual water usage of approximately 1,000 households, while larger hyperscale facilities may consume up to 5 million gallons daily.

Environmental and Social Impacts

Major Tech Companies have made commitments to achieve 100% renewable energy supply and carbon-neutral or carbon-negative operations by 2030 to 2040 (Google 2018; Microsoft 2024[7]; Amazon 2025). The pace of data center buildout is outrunning the deployment of clean generation in many regions, meaning that marginal load growth is often served by fossil fuels in the near term, even where operators hold long-term renewable energy contracts or purchase renewable energy certificates. Other environmental concerns include noise impacts, loss of amenity for neighboring communities, impacts on natural habitats, land use displacement, and workforce effects, because while data center construction can temporarily employ thousands of workers, standard operations typically require only a few dozen to a few hundred permanent employees per facility, creating a mismatch between the scale of and impact of the infrastructure and the scale of local economic benefit. Surging electricity demand, constrained interconnection queues, water stress, and community impacts have made data center integration one of the most active areas of energy and land use regulation in the country. Nationwide, states introduced more than 200 data center-related bills in 2025, and approximately 300 bills in the first six weeks of 2026 (Shonerd 2026). The volume of legislation introduced in the past two years by region is depicted in Figure 2, showing an almost perfect overlap between primary and secondary data center regions and the volume of bills introduced in those regions.

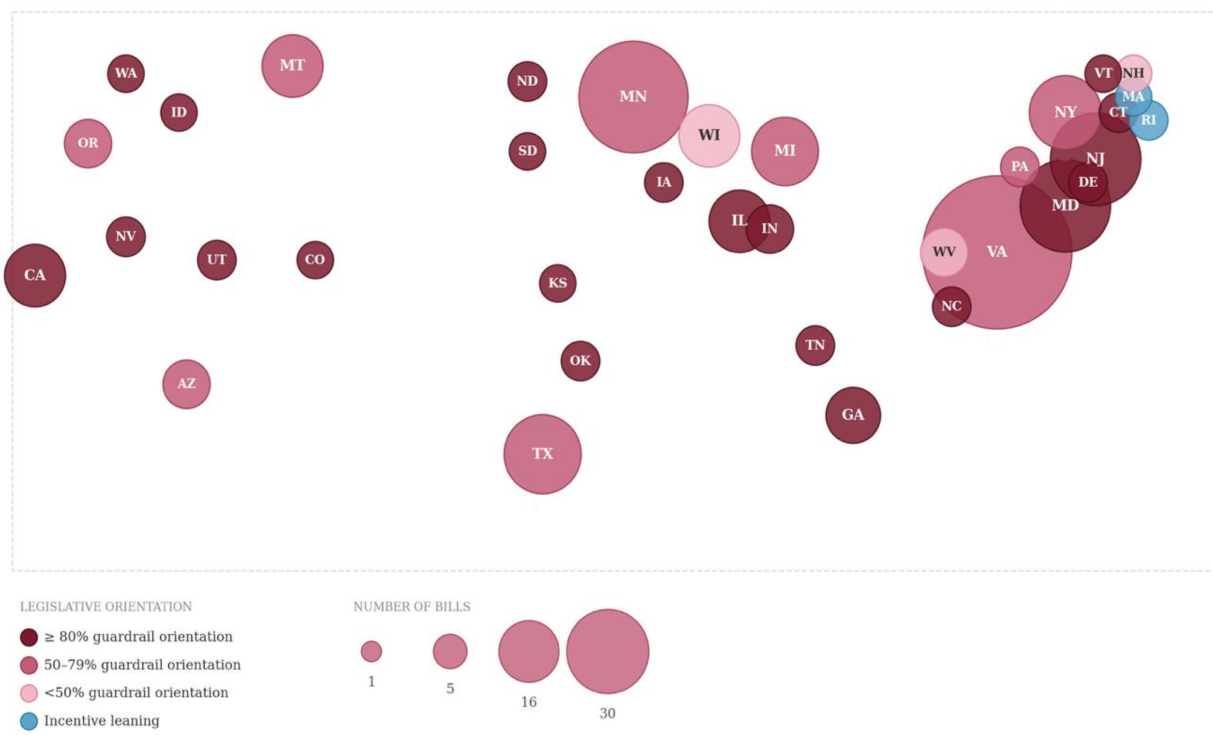

Figure 2: Data center regulation bills by state, showing legislative orientation and volume (2024-2026). Bubble size indicates number of bills; color indicates guardrail versus incentive orientation; N=155 bills (2024-2026) Source: Adapted and expanded from (Wincele 2026 ClimateXChange).

Bills introduced across both years are predominantly "guardrail" in orientation, spanning ratepayer protection, energy-use transparency, renewable energy requirements, siting and land use controls, water management and environmental concerns, and interconnection queue reform (grid reliability). Several states have also conditioned tax credits and incentives on environmental performance standards, including clean energy procurement, clean backup power options, and verified carbon offsets, and a growing number have introduced legislation requiring that a defined percentage of issued tax credits be directed toward the benefit of host communities, linking public subsidies to measurable local return (Baumgarth 2026). Recognizing the significant impact of these facilities, several states have recently introduced moratoriums preventing the development of new facilities, to give legislatures, utilities, and regulators time to adjust to issues of affordability, grid reliability, and social and environmental impact. The following sections examine how states are responding through new regulatory frameworks and contract structures governing the interconnection of hyperscalers and colocation datacenters.

Regulatory Context

Federal Framework

The federal government has approached large-scale data center integration through executive action and regulatory pressure. Executive Order 14318, issued in July 2025, directed federal agencies to accelerate environmental and permitting review for AI data center projects exceeding 100 MW, framing data center development as a matter of national competitiveness. A December 2025 Congressional Research Service report cataloged the full range of federal permits implicated in data center grid connection, covering generation interconnection, onsite backup power, pipeline access, and transmission upgrade approvals, underscoring the complexity that developers and regulators navigate simultaneously (CRS, 2025). At the Federal Energy Regulatory Commission (FERC), two significant actions have shaped the national picture. FERC's 2025 order directed PJM to revise its tariff rules for large loads co-located with generating facilities. In another ruling, FERC approved a PJM's temporary price collar proposal, for capacity auctions, capping prices at approximately $333.44/MW-day cap, while establishing a floor intended to maintain investment signals for new generation and revenue certainty for existing generation ($179.55/MW-day floor). These interventions reflect FERC's recognition that existing market structures were not designed for the pace or scale of data center load growth and converging pressures on the grid.

Regional Frameworks: The PJM Crisis as a National Inflection Point

The most consequential market signal of 2025 came from PJM Interconnection, the nation's largest grid operator serving 67 million people across 13 states and the District of Columbia. For the first time in PJM's history, the December 2025 capacity auction failed to procure sufficient capacity to meet its reliability standard, falling 5.2% short of requirements for the 2027/28 delivery year. PJM’s chief operating officer attributed the shortfall directly to rapid data center growth, noting that AI-driven demand continues to outpace new supply (PJM, 2025). The auction outcome shaped a debate about who should pay for the generation and transmission infrastructure required to serve large-scale AI loads, and how quickly that infrastructure could be built. They also exposed a structural weakness in PJM’s market design, whose planning tools were developed for incremental expansion rather than for evaluating numerous simultaneous 100-MW requests, leaving planners overextended. In January, 2026, the National Energy Dominance Council (NEDC[8]), together with 13 PJM-region governors, released a bipartisan Statement of Principles calling for emergency market reforms. The Statement called on PJM to implement a one-time Reliability Backstop Auction by September 2026, offering 15-year price certainty to attract new generation investment. In particular, the Statement directed that the costs of any new capacity procured through the auction would be allocated to data center customers that had neither self-procured capacity nor agreed to curtailment during periods of system stress. This established a structure intended to transfer demand risk from ratepayers to the companies driving load growth.

The Statement also called for extending the existing capacity price collar (low and high cap) through 2030, reforming load forecasting to exclude speculative demand, and reducing interconnection study timelines to 150 days. As a non-binding document, however, implementation of any of these measures would require PJM to file tariff changes with FERC for approval (NEDC 2026).

The PJM Board of Managers released a concurrent decisional letter on January 16, 2026, outlining five operational courses of action for the year, including establishing two interconnection pathways for new large loads: a voluntary "Bring Your Own Generation" option, under which customers self-supply with dedicated generation to avoid curtailment exposure, and a "Connect and Manage" framework, under which data centers may interconnect prior to the completion of transmission upgrades subject to priority curtailment during periods of system stress.

The PJM Board of Managers further directed the creation of an accelerated interconnection track for state-sponsored generation projects, the immediate initiation of a backstop generation procurement process to address near-term reliability needs through existing tariff mechanisms, and a holistic review of PJM energy, reserve, and capacity markets to assess their combined effectiveness in supporting new investment. In addition, PJM proposed significant reforms to load forecasting methodology, including an expanded role for states in demand projection processes (PJM 2026).

The gap between these two proposals illustrates the fundamental governance challenge. Retail utility customers, including data centers, are regulated at the state level, and there is currently no mechanism for direct participation of a retail customer in a wholesale capacity auction, meaning that implementing the White House framework could require changes in state law across PJM jurisdictions (as some states are fully deregulated, some partially and the mechanisms available to regulators, vary). The Statement of Principles carries political weight but no binding legal force, and any emergency auction requires PJM to file tariff changes and FERC to approve them. How PJM resolves these questions is likely to establish a reference point for every other region in the country.

Speculative Load and Flexible Interconnection

To address speculative loads in interconnection queues, system operators are beginning to adopt practical measures based on historical load materialization rates[9]. Queue reform proposals at MISO, PJM, and ISO-NE have identified readiness milestones[10] (Stevens et al. 2025), milestone

[8] Co-chaired by Interior Secretary Doug Burgum and Energy Secretary Chris Wright

[9] ERCOT roughly discounts large data center applications by half, and requires developers to disclose to a utility, if they are pursuing similar service requests with other electric utilities in Texas.

[10] Readiness milestone: required project development checkpoints that an interconnection customer must meet to remain in the queue (permits, detailed design information, study deadlines, financial commitments).

deposits[11], and site control requirements[12] as the most effective tools for reducing speculative applications, but implementation remains uneven. Standardizing flexible interconnection processes across all major system operators and managing speculative loads will bring interconnection queues closer to reflecting genuine demand, which is required for accurate load forecasting on which sound grid planning, resource procurement, and rate design all depend.

Emerging State Regulations

Several states and grid operators are moving to translate the technical potential of data center flexibility into formal regulatory and commercial frameworks. These approaches vary in their structure but share a common goal: connecting the speed-to-power problem with the grid management challenge by offering faster interconnection in exchange for verifiable flexibility commitments.

California

CA SB57directed the California Public Utilities Commission to assess cost-shifting from data center demand to residential ratepayers, establishing a public accountability framework.

Massachusetts

Massachusetts, through the Department of Public Utilities' grid modernization proceedings, has advanced flexible interconnection rules that include provisions for assumed load disconnection, allowing data centers to interconnect at full capacity while accepting contractual curtailment obligations during defined grid conditions.

Minnesota

Minnesota's HF 16, enacted in June 2025, established the first state-level water appropriation review process specifically applicable to data centers, requiring facilities consuming more than 100 million gallons annually to undergo evaluation of aquifer capacity, water conservation measures, and public health impacts prior to permit issuance HF16 MN (2025).

New Jersey

New Jersey's S 4143 NJ large load data center tariff, requires utilities to develop energy usage plans for qualifying data centers covering renewable energy sourcing, cooling efficiency, water optimization, and waste heat reuse before approving grid interconnection, and includes explicit clauses to ensure that costs of increased demand do not fall on other ratepayers.

New York

New York's ISO includes flexible large load projections in its reliability planning, modeling a potential reduction of approximately 1,200 MW during summer and winter peaks as a result of demand responsiveness from cryptocurrency mining, hydrogen production, and data center loads (NYISO 2024).

---

[11] Milestone deposits: escalating financial commitments projects must post at specific stages of the queue process, that are forfeitable if the project withdraws later.
[12] Site control requirements: developer to prove legal rights to the land where the proposed facility will be built (ownership, long-term lease, an option to purchase, or another legally binding agreement).

South Carolina

South Carolina's proposed H.4583, the Data Center Responsibility Act, is among the most stringent proposals yet introduced nationally, requiring complete onsite energy independence and closed-loop cooling systems as conditions of operation. While its passage is uncertain, it reflects a broader trend toward conditioning data center development approvals on environmental performance commitments that go beyond what existing utility interconnection processes require.

Texas

SB6 TX 2025 has drawn national attention as the first comprehensive state-level framework for large load interconnection. It establishes transmission cost allocation requirements for new large loads, mandates disclosure of onsite generation resources, and creates remote disconnect and curtailment protocols for emergency grid conditions, with an emergency disconnect provision. Texas lawmakers granted grid operators explicit authority to curtail AI data center load during system stress, a significant departure from prior practice and a model that several other states are now considering.

Virginia

Virginia, the nation's largest data center market, is amid a significant legislative response, with more than 27 bills introduced in 2025 and 2026 addressing siting, transmission cost allocation, affordability, environmental safeguards, and incentive conditions[13]. Tax incentives to data centers resulted in more than $1 billion in foregone revenue in 2024, which is nearly 80% of all Virginia's economic incentive spending that year. These events catalyzed legislative debate over whether continuation of state tax exemptions for data centers should require binding public benefit commitments (Sarte et al. 2024). Virginia HB 507 was amended to direct the Department of Environmental Quality to deny air permits for generators below EPA Tier IV emissions standards beginning in July 2026, while HB 897 would condition data center tax incentives on the use of emission-free backup power. Earlier versions of HB 507 that would have required battery storage as the primary backup source were substantially weakened in committee. Still, the combined effect of these measures represents the first state-level attempt to embed emissions quality standards directly into data center permitting conditions (Main 2026; Shonerd 2026).

Wisconsin

AB 840 WI 2026, among others, mandated closed-loop cooling systems for all data centers as a condition of operation and requiring annual water use reporting to the Department of Natural Resources. The bill passed the Assembly in January 2026, but failed to concur in March 2026.

Policy Trends

Bring Your Own Capacity Framework

The Bring Your Own Capacity (BYOC) framework reframes the question of who should pay for new capacity by directing large load customers to fund targeted demand-side management investments (ASE, 2026). Under the BYOC[14] model, the utility and the large load customers collaboratively design a geographically and technologically targeted portfolio of demand-side

---

[13] These bills reflect the intensity of Northern Virginia's data center concentration, which accounts for approximately 26% of statewide electricity demand compared with a national average of around 4%, and the affordability pressure this has placed on residential customers.

14 BYOC also known as BYODC – Bring Your Own Distributed Capacity

solutions, including energy efficiency, demand response, managed charging, and battery deployment. These are not system-wide programs, but locational investments directed at the specific transmission or generation constraints associated with expected load growth. The large load customers commit capital to fund incremental deployment of these programs, shifting early development costs away from ratepayers. The resulting savings are independently verified and formally accredited by the utility, the regulator, and, where applicable, the ISO/RTO. In return for this upfront capital commitment, the datacenter receives a defined benefit, most commonly an increased capacity reservation on a shorter interconnection timeline, providing the certainty that data center developers require. The tariff or contract structure governing these arrangements is publicly filed and available to all qualifying large loads, ensuring transparency, and targeting investment in locations and populations where it creates the greatest system value. As ASE (2026) notes, demand-side solutions can quickly create capacity at a fraction of the cost of traditional generation and infrastructure investment.

Development moratoriums

Development moratoriums are time-limited halts on new large data center approvals, and have been developed as mechanisms for creating regulatory breathing room in jurisdictions where the pace of development has outrun governance capacity. By early 2026, at least six states had introduced moratorium legislation, with proposals varying significantly in scope and duration. These include S 9144 NY imposing a minimum three-year-and-ninety-day halt on permits for facilities exceeding 20 MW pending completion of a generic environmental impact statement by the Department of Environmental Conservation and a ratepayer impact assessment by the Public Service Commission; SB 1488 OK would prohibit new facilities with electrical loads above 100 MW until November 2029 to allow study of impacts on water supply, utility rates, and property values, and SB 232 SD proposes a one-year pause on hyperscale development specifically (Shonerd 2026). At the local level, governments in at least fourteen states had enacted municipal or county-level moratoriums by the end of 2025, with Virginia's Loudoun County modifying its zoning rules to require legislative review and public hearings for all data center applications (Schneider et al. 2025).

Grid Edge Data Centers

While this paper focused on hyperscalers and colocation facilities, a buildout of smaller grid edge computing infrastructure is occurring largely outside the regulatory frameworks discussed above. Facilities below 5 MW typically fall below the thresholds that trigger large load interconnection studies, environmental review, or water permitting requirements (MN HF16). Individually, these facilities present manageable impacts. Cumulatively, in regions with dense deployment, they can impose aggregate demand, water, and land use impacts comparable to a mid-sized hyperscaler, while remaining invisible to the planning processes designed to manage those impacts. This regulatory gap deserves explicit attention as inference workloads proliferate and edge computing expands and represents a natural extension of the frameworks this paper describes.

Discussion

The regulatory landscape for large-scale data center integration is consolidating around a set of durable principles. After several years of ad hoc responses, driven largely by the urgency of individual interconnection crises, tax incentive competitions, and sporadic legislative efforts, a more comprehensive and deliberate policy architecture is beginning to take shape across the nation. This review identifies six emergent trends:

Restructuring data center tax incentives

States are increasingly restructuring data center tax incentives to condition public tax credits and incentives on measurable community and ratepayer benefit. Virginia, California, and New Jersey illustrate three distinct approaches, sunsetting or conditioning sales tax exemptions on binding public benefit commitments, establishing regulatory frameworks to assess cost-shifting from data center demand to residential ratepayers, and creating dedicated tariff structures that shield non-data center customers from infrastructure cost shifts (Sarte et al. 2024; SB 57 CA 2025; S 4143 NJ 2025).

Environmental performance standards regulation

A second trend involves the migration of environmental performance standards, particularly around backup power emissions and water consumption, from voluntary corporate commitments into binding regulatory conditions. This trend reflects a transition in which both the carbon intensity of backup power operations and the water intensity of cooling systems are increasingly treated as conditions of facility approval rather than as subjects of voluntary operator discretion.

Allocation of system costs.

Allocation of system costs involves the evolution of interconnection and capacity market policy toward frameworks that require large load customers to bear a defined share of the incremental system costs their demand imposes, rather than socializing those costs across the general rate base. Within the PJM region, this shift is most visible in the Critical Issue Fast Path (CIFP) process and its outcomes. The PJM Board's January 2026 decisional framework introduces a Non-Capacity-Backed Load (NCBL) construct under which large new loads, which are defined as facilities with interconnection requests at or above 50 MW, may connect to the grid but are designated as first-in-line for curtailment during reliability emergencies unless they elect the Bring Your Own Generation (BYOG) option and pair their demand with committed new supply (PJM 2026). At the state level, parallel legislative initiatives are embedding similar cost-assignment principles directly into large load tariff structures, requiring data centers to fund the grid infrastructure costs of serving their load rather than treating those costs as system-wide expenses eligible for socialization across the rate base (Shonerd et al. 2026). These policies reflect a structural shift toward front-loaded financial commitments that are designed to reduce speculative queue inflation and establish cost accountability as a precondition of grid access.

Data centers as grid citizens

Off-grid datacenters and behind-the-meter generation were once considered niche, but they are now gaining traction as structural responses to grid congestion. Proposals in California requiring at least 50% behind the meter energy supply as condition for tax incentive eligibility, South Carolina’s proposed requirement for full onsite energy independence, and PJM’s “Bring Your Own Generation” curtailment exemption pathway all signal growing policy support for development models that reduce pressure on shared transmission infrastructure. Total grid independence remains technically and economically challenging. However, regulatory policy is increasingly focused on the effectiveness of designs that minimize grid reliance and reward co-investment in local clean generation and storage.

Moratoriums halting data center development

By early 2026, at least six states had introduced moratorium legislation, with proposals ranging from one year pauses on hyperscale facilities to multiyear halts on permits exceeding specified load thresholds, each conditioned on completion of environmental impact studies (including water supply), utility rate analyses, or assessments of impacts on property values, and host communities. At the local level, governments in at least fourteen states had enacted municipal or county moratoriums by the end of 2025.These measures represent a shift in the regulatory posture, from project facilitation toward reconciliation of competing systemic interests: the speed to power, concentrated resource demand and massive capital deployment, that define large data center development, and the affordability, reliability and community and environmental welfare that the grid and human systems were designed to serve.

Conclusion

These trends reflect a shift from investment-focused policies to comprehensive standards that treat large-scale data center growth as part of public infrastructure governance. The pace of this transition varies by jurisdiction: Virginia and Texas are managing consequences that other states are still trying to prevent; Minnesota and New Jersey have advanced legislative frameworks; and Wisconsin and California are in the midst of contested proceedings that will shape regional precedent. Yet the overall direction is consistent across states and market structures. What remains unresolved is whether regulatory adaptation can keep pace with technological and commercial change. The tools now being deployed, including queue reform, cost allocation rules, water permitting, flexible interconnection frameworks, and community benefit conditions, are necessary but may prove insufficient if demand continues to accelerate.

References

Amazon, 2025. 2024 Amazon Sustainability Report, pp 57. https://sustainabilityexchange.amazon.com

AS 840 WI 2026. Relating to Certain Requirements Related to Data Centers. 2025–2026 Legislature. Introduced January 9, 2026; passed Assembly January 20, 2026 (53–44). Update: failed to concur 03/23/26. http://docs.legis.wisconsin.gov/2025/related/proposals/ab840

ASE, 2026. Bridging the Load Gap: A collaborative path for utilities, hyperscalers and customers. Adhoc Group, Alliance to Save Energy. White Paper. https://www.ase.org/sites/ase.org/files/large_load_customer_concept_report_-_jan_2026.pdf /ASE

Baumgarth, C. 2026. House Bill Could Benefit Communities Hosting Data Centers. Indiana Public Media, February 4. http://www.ipm.org/news/2026-02-04/house-bill-could-provide-benefit-to-communities-that-host-data-centers

Colangelo, P., Coskun, A., Megrue, J., Roberts, C., Sengupta, S., Sivaram, V., Tiao, E., Vijaykar, A., Williams, C., Wilson, D., Records, B., MacFarland, Z., Dreiling, D., Morey, N., Ratnayake, A. & Vairamohan, B. (2025). AI-centers as grid interactive assets. *Nature Energy*. http://doi.org/10.1038/s41560-025-01927-1

CRS 2025. Data Center Energy Demand and Grid Reliability: Federal Permitting Considerations. Congressional Research Service, Report R48762. December 2025. https://www.congress.gov/crs-product/R48762

Desislavov, R., F. Martinez-Plumed, and J. Hernandez-Orallo. 2023. Trends in AI Inference Energy Consumption. *Sustainable Computing: Informatics and Systems* 38. http://doi.org/10.1016/j.suscom.2023.100857

Garg, A., Booten, C., VanGeet, O., Pless, S., Stenger,K., Vidal, J., Jackson, R., Lavin, L., Prabakar, K. 2025. Considerations of Distributed Edge Data Centers and Use of Building Loads to Support Large Interconnections (NREL/TP-5500-96700). *National Laboratory of the Rockies.* https://research-hub.nrel.gov/en/publications/considerations-for-distributed-edge-data-centers-and-use-of-build/

Goldsmith, I., Byrum, Z. 2025. Powering the US Data Center Boom: Why Forecasting Can Be So Tricky. Washington, DC: *World Resources Institute*. http://wri.org/insights/us-data-centers-electricity-demand

Google 2018. Moving toward 24x7 Carbon-Free Energy at Google Data Centers: Progress and Insights. *Google Report* https://www.gstatic.com/gumdrop/sustainability/24x7-carbon-free-energy-data-centers.pdf

Google 2025. Google 2025 Environmental Report. Technical Report. Google Sustainability. https://www.gstatic.com/gumdrop/sustainability/google-2025-environmental-report.pdf

Hough, T. 2025. How ISOs and RTOs Are Addressing Large Load Growth in 2025. *Yes Energy*. http://yesenergy.com/blog/how-isos-and-rtos-are-addressing-large-load-growth-in-2025

Kunkel, C. 2025. PJM Capacity Auction Analysis. *Institute for Energy Economics and Financial Analysis (IEEFA)*. https://ieefa.org/resources/projected-data-center-growth-spurs-pjm-capacity-prices-factor-10

Kok, E., J. Rauer, P. Sachdeva, P. Pikul, D. Sutton, and R. Hasrouni. 2025. Scaling Bigger, Faster, Cheaper Data Centers with Smarter Designs. *McKinsey and Company*. https://www.mckinsey.com/industries/private-capital/our-insights/scaling-bigger-faster-cheaper-data-centers-with-smarter-designs?cid=soc-web

Luccioni, A., Strubell, E., and McCallum, A. 2025. From Efficiency Gains to Rebound Effects: The Problem of Jevons' Paradox in AI's Polarized Environmental Debate. *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. http://dl.acm.org/doi/10.1145/3715275.3732007

Green, A.Tai, H.,Noffsinger, J., Sachdeva, P., Bhan, A. Sharma, R. 2024. How Data Centers and the Energy Sector can State AI’s Hunger for Power. *McKinsey & Company* https://www.mckinsey.com/ industries/private-capital/our-insights/how-data-centers-and-the-energy-sector-can-sate-ais-hunger-for-power

Green, B. Nguyen, T. 2025. What happens when data centers come to town? *Science Technology and Public Policy Library, University of Michigan*. https://dx.doi.org/10.7302/26670

Microsoft 2024. 2024 Environmental Sustainability Report. Microsoft Corporation. Redmond. https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/msc/ documents/ presentations/ CSR/Microsoft-2024-Environmental-Sustainability-Report.pdf

HF 16 MN 2025. Data Center Regulatory Bill. 94th Legislature, 1st Special Session. Enacted June 15, 2025 (Chapter 12). http://www.revisor.mn.gov/bills/94/2025/1/HF/16/

Mytton, D. 2021. Data Centre Water Consumption. Nature npj Clean Water 4:11. https://www.nature.com /articles/s41545-021-00101-w

NCSL 2025. Electric Transmission Interconnection Queues. Denver: *National Conference of State Legislatures*. http://ncsl.org/energy/electric-transmission-interconnection-queues

NEDC 2026. Landmark Agreement to Restore Grid Reliability and Lower Costs. National Energy Dominance Council. https://www.doi.gov/pressreleases/NEDC-landmark-agreement-grid

Norris, T.H., Profeta, T., Patino-Echeverri, D., Cowie-Haskell, A. 2025. Rethinking Load Growth: Assessing the Potential for Integration of Large Flexible Loads in U.S. Power Systems. *NIEES, Duke University.* https://dukespace.lib.duke.edu/items/bb350296-d7a1-4d8f-acb0-2fba9b1f03de

NYISO 2024. Reliability Needs Assessment. A Report from the New York Independent System Operator. *NYISO*. https://www.nyiso.com/documents/20142/2248793/2024-RNA-Report.pdf

PJM 2026. PJM Board Outlines Plans to Integrate Large Loads Reliably. *PJM Inside Lines*. 1/16/26. http://insidelines.pjm.com/pjm-board-outlines-plans-to-integrate-large-loads-reliably/

Reali, M. 2025. Incentivizing the Digital Future: Inside America's Race to Attract Data Centers. Data Center Frontier. 8/5/25. datacenterfrontier.com/site-selection/article/55307797/incen tivizing-the-digital-future-inside-americas-race-to-attract-data-centers

Sarte, K., Gribbin, M. Miller, Berday-Sacks, E., Hopkins, K, Saunders, S. 2024. Data Centers in Virginia. Joint Legislative Audit and Review Commission,12/9/24.jlarc.virginia.gov/landing-2024-data-centers-in-virginia.asp

Sallam A. 2025. Targeted Interconnection Pathways for Near-Term Resource Adequacy in US ISO/RTO Markets. *Proceedings of the 2025 14th International Conference on Renewable Energy Research and Applications.*https://doi.org/10.1109/ICRERA66237.2025.11284084

Shonerd, D., Miller, K. 2026. State Data Center Legislation in 2026 Tackles Energy and Tax Issues. Multistate, February 20, 2026. https://multistate.us/insider/2026/2/20/state-data-center-legislation-in-2026-tackles-energy-and-tax-issues

S 4143 NJ 2025. Establishes Surcharge for Electric Public Utility Service to Certain Data Center Customers. 221st Legislature. http://www.njleg.state.nj.us/bill-search/2024/S4143

S 9144 NY 2026. NY Senate Bill S9144 Imposes a moratorium on data center permit issuance; and relates to data center rate impacts. Introduced. https://www.nysenate.gov/legislation/bills/2025/S9144

SB 57 CA, 2025. Ratepayer and Technological Innovation Protection Act. Enacted. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB57

SB 1488 OK 2026. An Act relating to the Corporation Commission; defining term; establishing moratorium; requiring certain study; requiring electronic submission of certain report; authorizing promulgation of rules; and providing for codification. Pending. https://www.oklegislature.gov/cf_pdf/2025-26%20INT/SB/SB1488%20INT.PDF

SB 232 SD 2026. An Act to impose a one-year moratorium on the construction or expansion of hyperscale data centers. Enacted. https://sdlegislature.gov/Session/Bill/27269/301228

SB 6 TX 2025. Relating to the planning for, interconnection and operation of, and costs related to providing service for certain electrical loads and to the generation of electric power by a water supply or sewer service corporation. Approved. https://legiscan.com/TX/bill/SB6/2025

Shehabi, A. Newkirk, A. Smith, S. Hubbard, A. Lei, N. Siddik, M. 2024 United States Data Center Energy Usage Report. *Lawrence Berkeley National Laboratory*. Report #: LBNL-2001637. http://dx.doi.org/10.71468/P1WC7Q

Schneider,D., Stolley,B. 2025. Loudoun County, Virginia, Eliminates By-Right Data Center Development. *Holland&Knight*. https://www.hklaw.com/en/insights/publications/2025/04/loudoun-county-virginia-eliminates-by-right-data-center-development

Stevens, J., Brooks, E. 2025. Adapting Utility Tariffs for Data Center Driven Load Growth. *Scott Madden*. December 19. http://www.scottmadden.com/insight/adapting-utility-tariffs-for-data-center-driven-load-growth/

Yañez-Barnuevo, M. 2024. Data Centers and Water Consumption. Environmental and Energy Study Institute. *Environmental and Energy Studies Institute*, Washington, DC. https://www.eesi.org/articles/view/data-centers-and-water-consumption

Wincele, R. 2026. Research Brief: Identifying Priority States for Data Center Regulation *ClimateXChange*. https://climate-xchange.org/2026/02/research-brief-identifying -priority-states-for-data-center-regulation/

Wolf, S. 2024. A Fundamental Shift: CenterPoint Sees 700% Increase in Data Center Interconnection Request Queue. *Renewable Energy World*. http://renewableenergyworld.com

Main, I. 2026. As 2026 Legislative Session Starts, Data Centers' Diesel Generators Are a Top Concern for Virginians. Virginia Mercury. January 14, 2026. https://tinyurl.com/mtd52u56