Advances in IT and Digital Security
ISSN 3143-4266
DAC Insight Publishers
https://journals.dacinsightpublishers.com/AIDY

ENERGIZATION RISK: A DECISION FRAMEWORK FOR POWERING DATA CENTRES, AND WHAT IT IMPLIES FOR INTERCONNECTION AND TARIFF POLICY

Daniel Raphael Ejike Ewim[1*], Paul Uche Didi [2]
[1]Centre for Engineering Education and Gender Studies, Ohio, USA
[2]University of New Haven, West Haven, Connecticut, USA
Corresponding Author*: daniel.ewim@ceegs.org · ORCID: https://orcid.org/0000-0002-7229-8980

ABSTRACT

Data-centre load is being connected to the United States power system faster than that system has historically added firm capacity. In most locations the binding constraint is not a shortage of energy resources but a coordination failure: developers cannot obtain credible dates for when power will arrive, utilities cannot distinguish real load from speculative load in their forecasts, and both sides sign decade-long commitments whose risks neither can quantify. This paper sets out a decision framework that treats energization risk, the joint uncertainty over how much firm capacity a site can obtain, by when, at what cost, and with what exposure if the plan slips, as the object of analysis, and draws out what it implies for the interconnection and tariff rules now being written. An illustrative application produces three results with direct policy consequences: the efficient set of supply plans changes depending on whether plans are judged at the median or the ninetieth percentile; modest correlation among queued requests widens the range of realised capacity by a factor of 2.4 without changing its average; and curtailment commitments buy speed in discrete steps rather than smoothly. The paper argues that FERC's June 2026 large-load orders can create, for the first time, a record capable of calibrating load forecasts, but only if the posting format preserves what was knowable when, and that minimum-take obligations, flexibility tariffs, and cost allocation should each be sized against distributions rather than single figures.

***Keywords**: Data-centre load growth; Large-load interconnection; Energization risk; Demand flexibility; Tariff design; Electricity regulation*

1. INTRODUCTION

Data-centre electricity consumption in the United States reached approximately 176 TWh in 2023, or 4.4% of national use, up from 58 TWh in 2014; Lawrence Berkeley National Laboratory projects a 2028 range of 325–580 TWh, or 6.7–12% of national consumption (Shehabi *et al.*, 2024). The width of that range is the salient feature. A factor-of-1.8 spread five years out is not an analytical failure but a faithful representation of uncertainty about chip efficiency, workload composition, utilisation, and the commercial fate of individual projects. The evidence base beneath it is also thin, with only about a third of published estimates peer-reviewed (Masanet *et al.*, 2020; Mytton & Ashtine, 2022).

The uncertainty compounds as it is disaggregated to the places where decisions are made. The North American Electric Reliability Corporation's 2025 Long-Term Reliability Assessment raised its ten-year summer peak growth forecast to 224 GW, a 24% increase over 2025 levels and roughly 69% above the same forecast a year earlier, and placed MISO, PJM, ERCOT, and parts of the Pacific Northwest at elevated reliability risk within five years (NERC, 2026). ERCOT's large-load interconnection queue exceeded 233 GW by late 2025, more than 70% of it data centres, after roughly 300% growth in a single year (Utility Dive, 2025). That most of that queue will not be built is not in dispute; which part, and by when, is.

One caveat belongs in the text rather than a footnote. The NERC figure is itself assembled from utility forecasts that aggregate queue requests, so it is not independent evidence that the requested load is real, and London Economics International (2025) argues that duplicate requests and cost-of-service capital incentives make those forecasts systematically upward-biased, while EPRI (2026) offers a competing institutional range of 9–17% of national electricity by 2030, a spread that is itself a measure of how little the aggregate is pinned down. The demand signal driving every downstream decision, from every interconnection study to every network upgrade and every tariff proceeding, is measured by instruments that have never been calibrated.

This paper makes three arguments. First, that the problem facing American data-centre power procurement is best understood as energization risk: the joint uncertainty over how much firm capacity a site can obtain, by when, at what cost, and with what financial exposure if the plan slips. Second, that a decision framework organised around that concept produces conclusions a conventional analysis does not, three of which are demonstrated here. Third, that those conclusions bear directly on rules being drafted now at FERC, at the regional transmission organisations, and at state commissions, and that several of those rules are being specified in ways that will not work.

Section 2 diagnoses the coordination failure. Section 3 states what existing tools and literature already supply and where the gap actually lies. Section 4 describes the framework. Section 5 reports three findings from an illustrative application. Section 6, the substantive core of the paper, sets out the policy implications. Sections 7 to 9 give the framework's scope limits, the study's limitations, and conclusions.

2. A COORDINATION FAILURE WITH THREE FACES

2.1 Demand: derating without calibration

Utilities already discount the load that developers request, and the adjustments are substantial. Berkeley Lab's Speed to Power review records PJM applying a 50% derate to large loads that miss milestone requirements within specified timeframes, California weighting loads by confidence level on a scale running from 10% for an initial inquiry to 100% for a signed agreement, and Santee Cooper making what it calls a risk-adjusted post-modelling adjustment (Kahrl & Frick, 2026). What is absent is not the practice but any basis for judging it. The same review finds that utilities "may use different methods, terminology, and definitions", so the adjustments are not comparable between one territory and the next; some appear in tariffs or regulatory filings while others remain internal methodology; and the question of whether any of them has been right before is one the published record does not address. A regulator comparing two utilities' forecasts cannot tell whether a difference between them reflects different circumstances or different guesswork.

This matters more than it may appear. The derating factor is the single number that converts a speculative queue into a capital plan. Set too generously, ratepayers fund network upgrades for load that never arrives. Set too severely, the utility under-builds and reliability suffers. There is at present no public basis for knowing which error is being made, or in which direction, in any given territory.

2.2 Supply: lead times that are long, correlated, and poorly characterised

Berkeley Lab recorded roughly 1,312 GW of generation and 749 GW of storage across some 8,200 active interconnection requests at the end of 2025, with queued gas capacity up 86% year over year and, against the prevailing narrative, total active queue volume down 10%. The median duration from request to commercial operation exceeded five years for projects reaching operation in 2025, and of requests submitted between 2000 and 2020, 13% reached operation while 75% were withdrawn (Rand *et al.*, 2026). Network upgrade costs have risen across all studied regions and are now the dominant cost driver (Seel *et al.*, 2023).

The load itself is built far faster. Hyperscale construction is commonly reported at eighteen to thirty months, and Idaho National Laboratory (2025) records the workshop finding that "data center construction is outpacing the multi-year cycles required for grid upgrades, permitting, and the procurement of critical equipment like large transformers". This asymmetry, a two-year load asset chasing a five-to-seven-year supply asset, is the structural fact around which every siting decision now turns. It is why bridge generation, behind-the-meter supply, co-location at existing plants, and curtailment-enabled interconnection have all moved from the margin to the mainstream.

Less widely appreciated is that these delays are correlated. Turbine delivery, engineering and construction labour, and transformer supply are constrained market-wide, not project by project. A developer holding four options does not hold four independent chances of being on time, and a utility screening forty requests is not observing forty independent bets. Treating them as independent is the most common way portfolio risk is understated, and Section 5.2 quantifies the size of that error.

2.3 Institutions: risk transferred against the wrong number

PJM's 2027/2028 Base Residual Auction cleared at the \$333.44/MW-day cap and procured 134,479 MW, 6,623 MW short of the reliability requirement (PJM, 2025), against projected load-weighted wholesale prices rising from roughly \$47 to \$51/MWh (EIA, 2025). Against that backdrop, on 18 June 2026 FERC issued show-cause orders under Section 206 of the Federal Power Act to six jurisdictional RTOs, defining large loads as those exceeding 50 MW peak above 69 kV and directing standardised study processes with 60-to-90-day timelines, cost-recovery agreements with credit support, hourly load forecasting and telemetry obligations, expedited study pathways for electrically proximate generation, and public posting of large-load additions and network upgrade costs (White & Case, 2026).

State tariff design has moved in parallel. Large-load tariffs now embed long contract terms, with filed examples running from five to twenty years; minimum billing demand commonly set at 80 to 85% of contract capacity; collateral or credit-rating requirements; and exit fees reaching, in one filed tariff, three years of minimum charges (Satchwell *et al.*, 2025). The combined effect is to transfer load-forecasting risk from the utility to the customer. That is defensible in principle, since the customer knows more about its own plans, but it is sized against nameplate capacity rather than against any distribution of what that customer will actually draw. Whether the transfer is efficient, and who bears the residual, is contested (Martin & Peskoe, 2025).

These instruments are being written quickly, in dozens of proceedings, on a factual record that Section 2.1 has just described as uncalibrated. That combination has a political cost as well as an economic one. Where a commission approves network investment on a load forecast it cannot verify, and retail bills subsequently rise for reasons that are difficult to disentangle, the resulting loss of public consent falls on the whole programme of grid investment, including the transmission build that reliability requires irrespective of data centres. Getting the forecasting basis right is therefore not a technical nicety at the edge of the problem; it is a precondition for sustaining the investment consensus the sector depends on.

A siting decision is therefore simultaneously an engineering decision, a procurement decision, a regulatory decision, and a contingent financial commitment of several hundred million dollars, taken at the moment of least information.

3. WHAT ALREADY EXISTS, AND WHERE THE GAP ACTUALLY IS

Three bodies of work bear on this problem, and it is worth being precise about what each supplies. The research literature. Bottom-up demand estimation is established (Masanet *et al.*, 2020; Shehabi *et al.*, 2024), as is the finding that facility efficiency turns on cooling architecture, hardware, and design choices that no analysis of the grid alone can capture, and that technological measures alone are insufficient without accompanying policy (Ewim *et al.*, 2023). Two-stage siting analysis combining spatial exclusion with weighted criteria is standard practice (Malczewski, 2006) and has been applied to data centres directly (Hussain *et al.*, 2026; Ayyildiz *et al.*, 2025). Johnston *et al.* (2023) model queue entry, waiting, and withdrawal econometrically. Norris *et al.* (2025) show that the twenty-two largest US balancing authorities could host roughly 76–126 GW of additional load if that load accepted curtailment of 0.25–1% of its annual energy, though without modelling transmission constraints. The physical achievability of such flexibility is now demonstrated rather than assumed: Colangelo *et al.* (2026) sustain a 25% power reduction for three hours during peak demand on a 256-GPU cluster at a hyperscale facility in Phoenix, Arizona, while maintaining quality-of-service guarantees and without hardware modification or energy storage, and Williams *et al.* (2026) demonstrate rapid load reduction, sustained curtailment, and carbon-aware operation in a real-world deployment on a 130 kW GPU cluster. Most directly, Kim *et al.* (2026) show that operational flexibility expands the siting frontier by 9–17% for a 1 GW load and 19–21% for a 2 GW load relative to firm operation, and argue that a planner-initiated, pre-certified process could deliver first energization in twelve to eighteen months against a conventional five-to-eight-year timeline. Farmer *et al.* (2025) translate curtailment-enabled headroom into a concrete interconnection-process design for rulemaking.

Commercial tools. It is sometimes claimed that siting software treats power as an afterthought. In 2026 that is no longer true, and a gap statement resting on it is not credible. On their own public descriptions, retrieved in August 2026 and reported here as vendor claims rather than verified capability, Enverus markets parcel-level buildability scored against more than 23,000 interconnection points with available capacity; PVcase markets consultant-grade interconnection studies with grid-upgrade cost estimation; Nira Energy estimates point-of-delivery available capacity including the effect of competitor load and queued generation; and Build.inc models P50/P75/P90 energization scenarios from FERC and state dockets, benchmarked against historical delivery timelines. Pearl Street Technologies' interconnection study automation has been adopted by MISO.

Set against this, Chen and Zheng (2026) caution that the benefits are context-dependent: increasing flexibility does not necessarily translate into less generation capacity required, cost savings across their cases range widely, and extending the deferral window yields progressively smaller returns.

What is therefore missing. Not capability, but accountability. No tool in this space, commercial or academic, publishes a record of how its own past forecasts performed. None states which percentile its recommendations are computed at. None discloses whether the correlations underlying its risk estimates were measured or assumed. And no utility publishes the derating rule that converts its queue into its capital plan, still less any evidence about whether that rule has been right before.

That gap is narrower and more tractable than a naive reading of the market would suggest. It is also, for reasons Section 6 develops, substantially a regulatory question rather than a technical one.

4. THE FRAMEWORK

The framework organises the problem around a single idea. For any candidate site there exists a power-availability frontier: the set of combinations of firm capacity, delivery date, cost, and reliability actually achievable there, given the resources, interconnection pathways, and contract structures reachable from that location. Conventional siting analysis asks whether a site can be powered. This framework asks what the shape of its frontier is, how that compares to alternatives, and which point on it best matches a particular decision-maker's needs. Speed, cost, and reliability trade against one another, and the terms of that trade are specific to each site.

Every output is expressed in three currencies at once: megawatts of firm deliverable capacity over time, months to each capacity milestone with the uncertainty stated, and dollars at risk, the exposure created by commitments that must be made before the outcome is known. Reporting one alone misleads in a predictable direction. A schedule with no exposure attached invites over-commitment; a cost estimate with no schedule uncertainty invites plans that are cheap and late.

Four components produce these outputs (Figure 1).

Load materialisation and siting estimates, for a requested project, the probability that it reaches energization by a given date, replacing the uncalibrated derating of Section 2.1 with a scored instrument fitted to realised outcomes rather than to expert opinion. It draws on the maturity indicators utilities already use informally: land control, permitting stage, executed agreements, equipment orders, tenant commitment, capital raised, developer track record. Two design choices matter for policy. Projects are grouped by developer, by territory, and by supply chain, so the model can represent the correlated failure of Section 2.2 rather than assuming projects fail independently. And the model is explicitly predictive, not causal: maturity indicators are as much consequences of progress as causes of it, so a utility must not read them as levers to be pulled. The framework also separates the population-level question a utility asks about third parties from the ramp a data-centre operator forecasts for its own facility. These are different questions requiring different information, and conflating them produces exactly the mis-sized obligations Section 6.2 identifies.

Siting itself follows established practice: hard constraints eliminate parcels outright, then graded criteria score the survivors on electrical proximity and headroom, transmission capacity, land, water, permitting duration, fiscal regime, labour, community conditions, and hazard exposure. This is an application of existing method rather than a contribution, and the framework's only insistence is that individual criterion scores be shown alongside any composite, with weights visible and set by the user rather than buried in a default.

Dispatchable-power matching assembles supply plans that meet a load programme over time. Each resource (network service, a purchase agreement, on-site generation, co-location, storage, a bridge installation, or a flexibility commitment) carries not a single delivery date but a range of possible dates, decomposed by the process that drives it: study completion, agreement execution, permitting, equipment delivery, construction, commissioning. Crucially those ranges are linked rather than independent, so that a supply-chain squeeze delays several resources at once. Because load arrives faster than firm supply, the answer is usually a sequence: bridge generation succeeded by network service, or curtailment-enabled service succeeded by firm service. The framework therefore evaluates sequences rather than choosing single resources.

Reliability and cost scoring converts plans into comparable figures, and does two things current practice does not. It reports two distinct reliability quantities and refuses to combine them: exposure arising from the contracted supply portfolio, and local delivery reliability at the point of delivery, which is driven by distribution topology and comes from published utility statistics rather than from any model of the site. Blending them produces a number insensitive to what actually differentiates one candidate site from another. On cost, it reports a delivered cost of energization that partitions cash flows by who is paid and under what instrument (capital, operations, fuel, network upgrade allocation, interconnection facilities, and tariff charges net of directly assigned recovery), so a user can see whether a site is expensive because of energy price, upgrade allocation, tariff structure, or the premium paid for speed. Those are four different problems with four different responses, and a single blended figure conceals which one is operating.

Decision support presents these results to five user classes: data-centre developers, utilities and load-serving entities, power developers, equipment suppliers, and, for reasons Section 6.4 sets out, regulators and consumer advocates. Uncertainty is displayed by default, and where the analysis cannot reliably distinguish two options the interface shows them as a tier rather than a ranking.

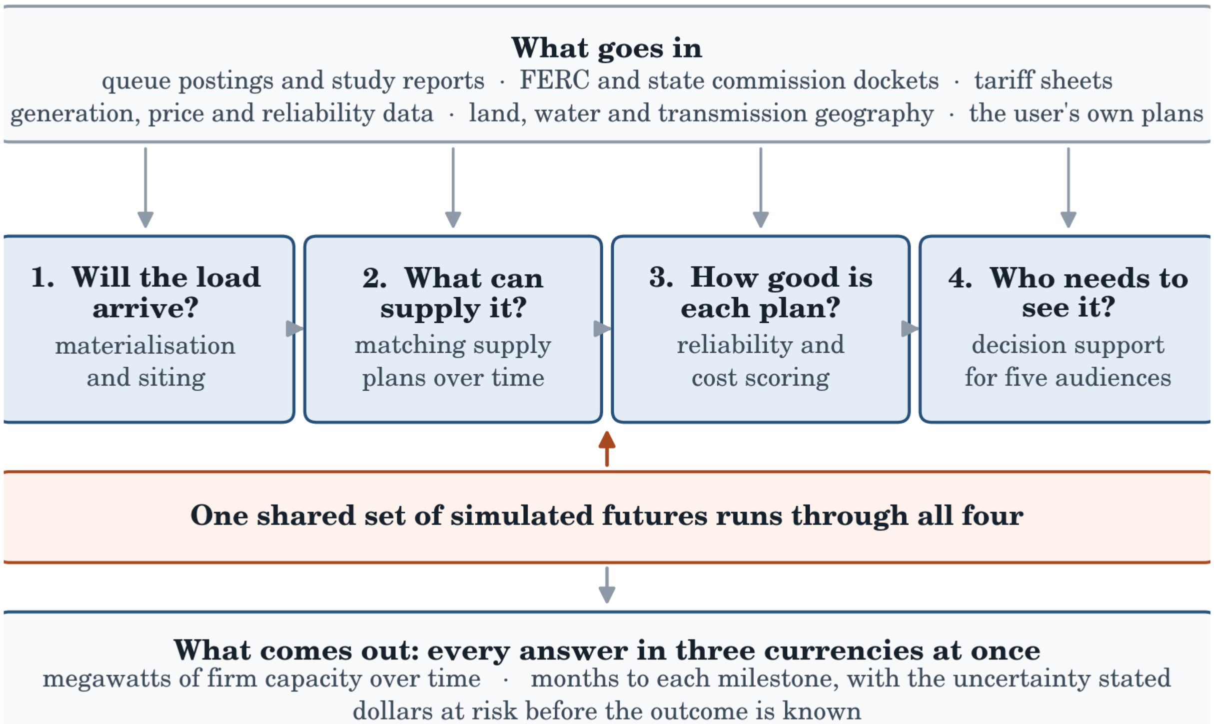

Figure 1: The four components of the framework and the data they consume. A single set of simulated futures is shared across all components, so that a supply-chain squeeze or a regulatory change moves several estimates together rather than being averaged away inside each one separately.

5. THREE FINDINGS FROM AN ILLUSTRATIVE APPLICATION

The framework was applied to a stipulated case: a 300 MW facility-load programme reaching 60 MW at month 18, 150 MW at month 30, and 300 MW at month 48, evaluated over eleven years at two candidate sites: one a brownfield beside an existing 345 kV substation with 60 MW of existing headroom, the other a greenfield beside gas transportation. Nine supply plans were compared across 120,000 simulated futures.

This is a demonstration, not evidence. Every parameter is stipulated for plausibility; none is estimated from data; no claim is made about any real site, resource, or market. What the exercise establishes is that certain properties follow from the structure of the problem rather than from the particular numbers chosen, and those properties have policy consequences.

5.1 The efficient choice depends on which percentile you plan against

Judged at the median, three of the nine plans were efficient, in the sense of not being beaten by any other plan on both time and cost. Judged at the ninetieth percentile, the efficient set changed. One plan, a bridge-plus-curtailment sequence terminating in an expedited but contested firm-service pathway, was efficient at the median and beaten in the tail. Another, pairing a bridge with network service and a co-located combined-cycle plant, was the reverse.

The mechanism is straightforward. The first plan had the earliest median date, 53 months, but the widest spread, its range from the fiftieth to the ninetieth percentile running to 59 months, because an expedited pathway that may be contested carries a long tail. The second had a later median but a much tighter distribution.

This is not an artefact of numbers chosen to be dramatic. It is the generic consequence of comparing options whose risk profiles differ in shape, and it is invisible to any procedure reporting a single expected date. A decision-maker shown only the median view would shortlist the first plan; one shown only the tail view would reject it (Figure 2).

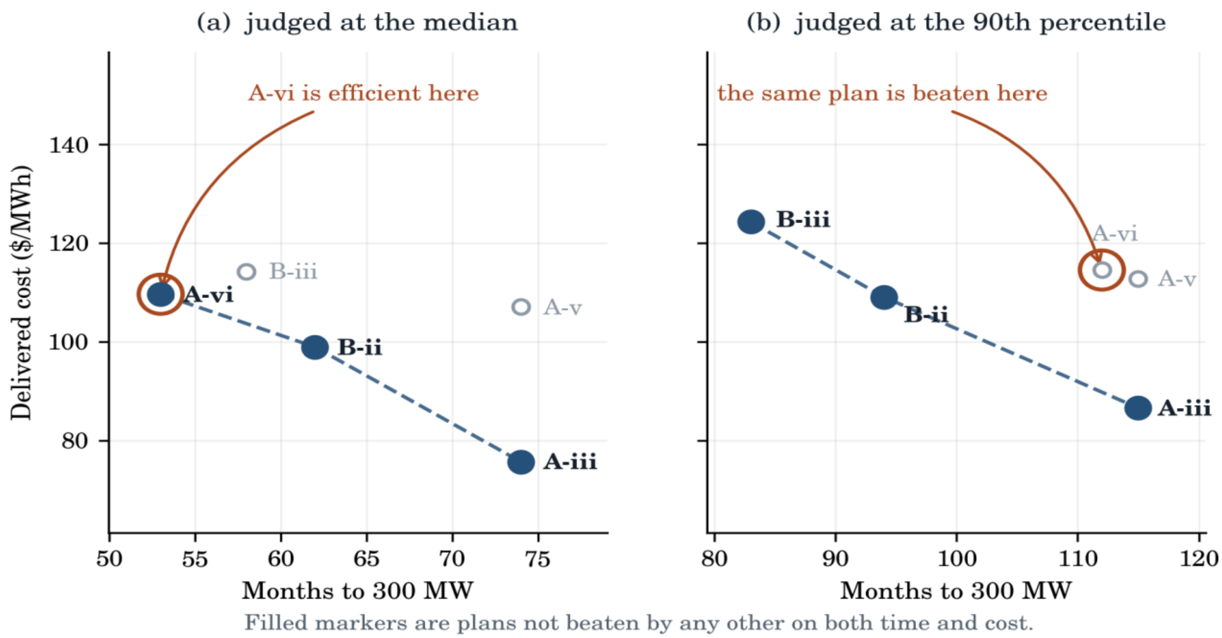

Figure 2: The efficient set of supply plans changes with the percentile it is judged at. Filled markers are plans not beaten on both time and cost. Plan A-vi is efficient at the median (a) and beaten at the ninetieth percentile (b); plan B-iii is the reverse. Both panels hold reliability at the same level. Stipulated parameters; illustrative only.

5.2 Correlated failure, not project count, dominates portfolio risk

Consider a utility screening forty queued large-load requests of 250 MW each, where each has a 35% chance of materialising within five years. Treated as independent, realised capacity falls between roughly 2,250 MW and 4,750 MW ninety per cent of the time, a range of 2,500 MW around an average of about 3,500 MW.

Introduce modest correlation between them, the kind that arises when a handful of hyperscalers, a shared set of turbine and transformer suppliers, and a common regulatory regime govern most of the megawatts, and the average does not move at all. The range does. At a pairwise correlation of 0.20 it widens to 6,000 MW, a factor of 2.4. At 0.50 it widens to 8,750 MW and the fifth percentile falls to zero: a materially probable outcome in which none of the forty requests materialises, to which an independence assumption assigns essentially no probability (Figure 3).

Utilities size procurement, network upgrades, and reserve margins against precisely this range. Any screening process that derates requests one at a time, which is to say current practice, produces a planning interval too narrow by roughly that factor, and narrow in the direction that has consequences.

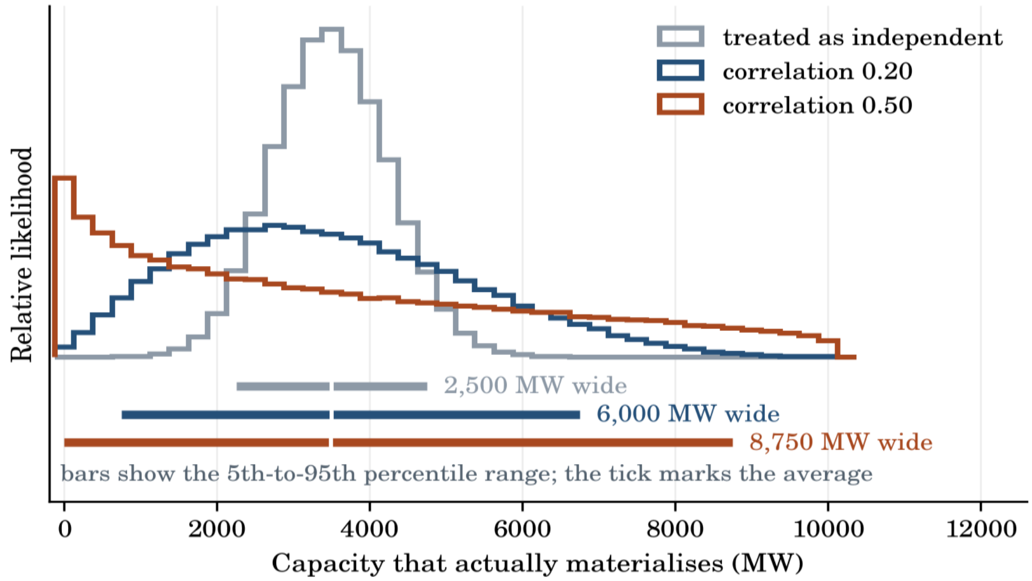

Figure 3: Realised capacity from forty queued requests. Each request has the same 35% chance of materialising in all three cases; only the correlation between them changes. The average is unmoved, the range is transformed. Stipulated parameters; illustrative only.

5.3 Flexibility buys speed in steps, not smoothly

At the brownfield site, accepting curtailment of 0.25% of annual energy, about 88 hours a year at an average depth of 25% rather than full shutdown, advanced the 150 MW milestone from month 74 to month 26. That is a 48-month acceleration, at an increase in delivered cost of $26.6/MWh and a forgone compute value of roughly $211M.

Increasing the obligation beyond that bought nothing further for a long stretch. At 1% of annual energy the same milestone still arrived at month 26, while the forgone compute value rose to $946M, an additional $735M transferred for no additional speed. Only at 2% did a second step appear, when the unlocked capacity became sufficient to reach the full 300 MW without waiting for the firm upgrade at all.

The shape is a staircase, not a slope: value accrues where accumulated capacity crosses a milestone requirement and is flat between crossings (Figure 4). The relationship between curtailment and unlocked capacity is stipulated here rather than measured, being precisely the quantity that must come from utility hosting-capacity data, but the staircase structure follows from the fact that load arrives in tranches, and that is not an assumption. The direction is consistent with Chen and Zheng (2026), who find that flexibility does not monotonically substitute for capacity and that extending the deferral window yields diminishing returns.

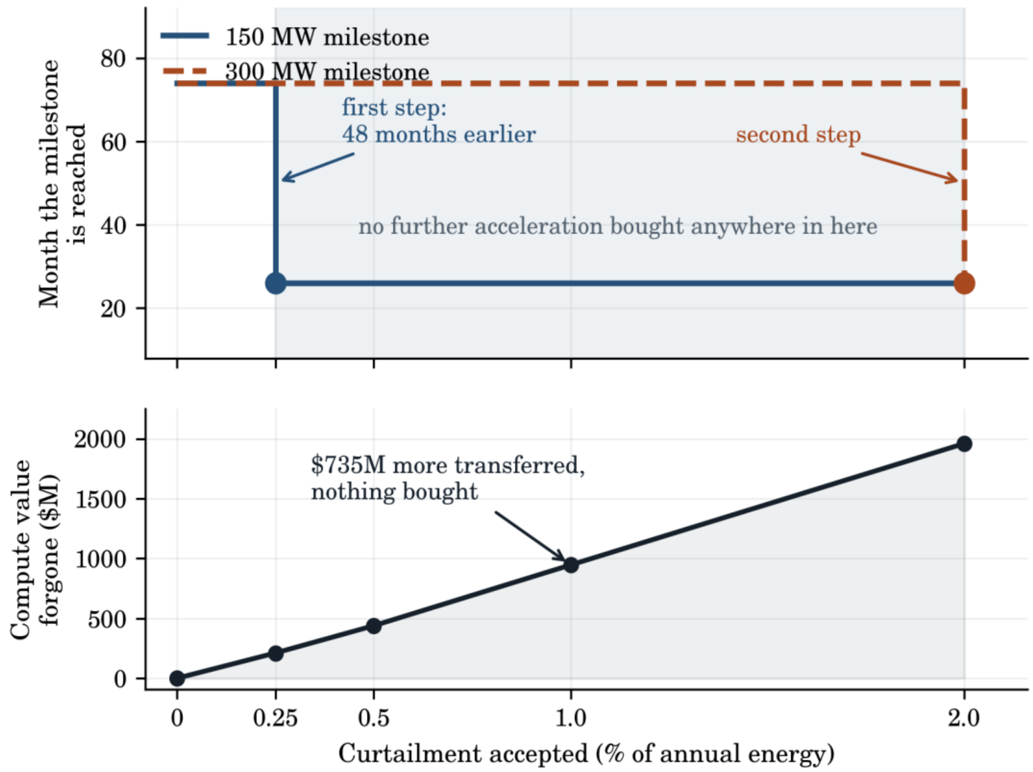

Figure 4: Curtailment accepted against firm capacity unlocked and compute value forgone. The first increment unlocks most of the available capacity; later increments transfer substantial value for little additional speed. The capacity-unlock relationship is stipulated, not measured (Section 5.3).

5.4 A fourth result, about analysis itself

One further comparison deserves reporting. If plans are evaluated by asking, in each simulated future, which option happened to arrive first, a natural-seeming procedure that a competent analyst might adopt without noticing the error, the resulting estimates understate expected cost by 19.4% and expected schedule by 12.9 months, against the same plans evaluated as commitments that must be made before the future is known.

That is the same direction, and roughly the same order of magnitude, as the schedule optimism the sector's own record exhibits. It suggests that a meaningful share of the industry's chronic over-optimism is not carelessness but a reproducible artefact of how the analysis is set up.

6. POLICY IMPLICATIONS

The findings above are not primarily commercial. Each bears on a rule being written now. Table 1 summarises the recommendations that follow, with the body responsible for each and the window in which it is actionable.

Table 1. Summary of recommendations.

#	Recommendation	Responsible body	Window
1	Require large-load postings to be immutable and retrieval-time stamped, preserved as dated snapshots rather than overwritten	FERC; RTO compliance filings	Immediate; settled in current compliance proceedings
2	Require derating parameters to be filed alongside the forecasts they produce, with realised outcomes reported against them	State commissions; FERC	1–2 years
3	Step minimum-take obligations with the contracted ramp rather than applying them against nameplate from energization	State commissions; utilities	Current tariff proceedings
4	Denominate curtailment obligations in annual energy, not hours, and state the assumed average depth	State commissions; RTOs	Current tariff proceedings
5	Structure flexibility credits around capacity steps rather than a smooth marginal price	State commissions; utilities	1–2 years
6	Publish hosting-capacity and substation-loading data at a granularity supporting site-level headroom estimates	Utilities; state commissions	1–3 years
7	Require analysis in large-load proceedings to report allocated and unallocated upgrade cost separately	State commissions	Immediate
8	Fund independent analytical capacity for consumer advocates in large-load proceedings	State legislatures; commissions	1–2 years
9	Require any decision-support analysis introduced into a proceeding to state its percentile, disclose whether correlations were measured or assumed, and publish its calibration record	State commissions; FERC	2–3 years

6.1 FERC's large-load orders can create a calibratable record, if the posting format is specified for it

The most consequential provision of the June 2026 show-cause orders, from a measurement standpoint, is among the least discussed: the requirement that RTOs publicly post large-load additions and the associated network upgrade costs (White & Case, 2026). Section 2.1's diagnosis, that derating rules are uncalibrated because there is no outcome record to calibrate them against, has held because the necessary record has never existed in usable form. These orders could create it within a few years.

Whether they do depends on a detail that will be settled in compliance filings rather than in the orders themselves. Queue data as presently published is a current-state snapshot: statuses are mutated in place, withdrawal dates must be inferred from a project's disappearance between one posting and the next, and there is no way to reconstruct what was knowable at a past date. A record with those properties cannot be used to test whether a forecasting rule would have worked, because the analyst cannot avoid contaminating the test with information that arrived later.

The remedy is inexpensive and specific. Postings should be immutable and retrieval-time stamped: each posting preserved as a dated snapshot rather than overwritten, with status changes recorded as new entries carrying the date on which the change became known. This is routine data-architecture practice and adds essentially nothing to compliance cost. Without it, publication produces a record that looks usable and is not.

Commissions specifying these formats should also require that the derating parameters a utility applies to requested load be filed alongside the forecasts they produce, and that realised outcomes be reported against them as they resolve. A derating factor is a forecast, and forecasts that are never scored do not improve. This complements a recommendation the Idaho National Laboratory (2025) workshop reached from the other direction: data centres already hold detailed load forecasts but rarely share them, and standardised protocols for sharing them with utilities would reduce the conservatism utilities apply in their absence.

6.2 Minimum-take obligations are sized against the wrong quantity

Minimum billing demand set at 80 to 85% of contract capacity, now common (Satchwell *et al.*, 2025), transfers load-forecasting risk to the party with better information about its own plans. That is the right direction of transfer. The problem is the quantity it is sized against.

A contract-capacity figure is a single number describing a facility that will in fact ramp over several years with genuine uncertainty at each stage. Sizing a take-or-pay obligation against nameplate rather than against a ramp distribution guarantees the obligation is wrong: too loose if the facility fills quickly, punitive if it fills slowly for reasons outside the customer's control. Neither error serves the utility, since a customer facing punitive charges on capacity it cannot use has every incentive to litigate, renegotiate, or walk away.

Two changes would improve this. Obligations should step with the contracted ramp rather than applying at full capacity from energization, and the schedule of steps should be a negotiated term informed by the customer's own forecast rather than a tariff default. And where a customer's stated ramp is materially more aggressive than comparable projects have achieved, that divergence should be visible to the commission reviewing the arrangement, which requires exactly the outcome record Section 6.1 describes.

6.3 Flexibility tariffs should price energy, not hours, and should be written in steps

Two errors are available in designing curtailment-linked tariffs, and both are being made.The first is a units error. The curtailment percentages in the flexibility literature are fractions of annual energy, not fractions of hours. A 0.5% obligation corresponds to roughly 177 hours a year at an average depth of about 25%, not 177 hours of full shutdown. A tariff specifying curtailment in hours where the underlying analysis was in energy misprices the obligation by approximately a factor of four, and the error runs in the direction of extracting far more from the customer than the analysis contemplated.

The second is a shape error. Section 5.3 showed that the value of flexibility accrues in steps. A tariff pricing curtailment as a smooth marginal quantity, so much credit per additional hour, will be wrong almost everywhere: too generous in the flat stretches, too stingy at the steps. The analytically important quantity is the location of the steps, which depends on how a load's capacity milestones line up with what curtailment unlocks locally.

That points to a third recommendation. Site-level headroom cannot be inferred from the balancing-authority aggregates published in the flexibility literature, which are computed without transmission constraints, because headroom is local and transmission is what binds locally. Expedited interconnection conditioned on flexibility therefore requires utilities to publish hosting-capacity and substation-loading data at a granularity that supports site-level estimates. Without it, a flexibility tariff is priced against a number nobody can compute.

6.4 Cost allocation is a two-sided problem and should be analysed as one

Network upgrade costs appear in any developer-facing analysis as a cost to the developer. Their allocation between the large load and the general ratepayer is the contested question in current tariff proceedings (Martin & Peskoe, 2025). A tool that optimises the developer's allocated share without surfacing the residual borne elsewhere is optimising one side of a two-sided problem, and a proceeding informed only by such tools is structurally imbalanced.

Two practical consequences follow. Analysis submitted in large-load proceedings should report allocated and unallocated upgrade cost separately, so the residual is visible rather than implied. And regulators and consumer advocates need access to the same load forecasts and cost decompositions the applicant relies on, which is why they are treated here as first-class users of the framework rather than as recipients of its outputs. An asymmetry of analytical capability inside a rate case is itself a distributional outcome.

The asymmetry is not primarily one of legal representation, which consumer advocates in most states have, but of analytical capacity. A hyperscaler's application arrives supported by proprietary siting analysis, load forecasts, and cost modelling that an advocate's office cannot replicate on its budget and cannot easily interrogate without the underlying assumptions. The remedy is unglamorous and cheap relative to the sums at stake: standing analytical support for advocates in large-load proceedings, funded through the same mechanisms that already fund intervenor participation in several states. Recommendation 8 in Table 1 reflects this. Without it, the transparency provisions of the FERC orders will improve what is published without improving what is contested.

A related point deserves stating because it is easy to encode without noticing. Siting analyses routinely score short local permitting durations as a virtue. They may reflect efficient administration; they may equally reflect weak local process and limited community participation.

Treating permitting velocity as an unambiguous positive takes a normative position on that question, and it should be taken explicitly if at all.

6.5 The analytics themselves should be held to a published standard

A framework of this kind would be used to inform commitments of several hundred million dollars, and in some cases to inform filings before commissions. That imposes an obligation its authors, including these, should accept rather than assert.

Three commitments are proposed as a standard for any analysis of this kind offered to regulated entities or introduced into proceedings. State the percentile. A recommendation computed at the median and one computed at the ninetieth percentile can name different plans, as Section 5.1 shows; a recommendation that does not say which was used is not interpretable. Disclose whether correlations were measured or assumed. The dependence structure driving portfolio risk cannot presently be estimated from available data. The series are short, aggregated, and highly autocorrelated, so any framework reporting a single tail figure is reporting an assumption, and should say so and give a range across plausible assumptions instead. Publish the calibration record. Confidence claims should rest on evidence of how past forecasts performed, published and updated as outcomes resolve; until such a record exists, any confidence rating should be labelled as resting on input quality rather than demonstrated accuracy.

This last commitment is what separates a public-interest instrument from a private information advantage in a bilateral negotiation. Better-calibrated load forecasts have a genuine claim on both sides of the error: over-building imposes costs on ratepayers, under-building imposes costs on reliability. That claim holds only if the calibration record is actually published. If it is not, this framework, and every commercial tool alongside it, should be evaluated as what it would then be.

6.6 Sequencing: what is actionable now, and what has to wait

The nine recommendations differ sharply in how quickly they can bind, and treating them as a single package would waste the one that is genuinely time-critical.

Now, and only now. Recommendation 1 has a closing window. The posting formats that will govern large-load transparency for the next decade are being settled in compliance filings responding to the June 2026 orders. Formats are difficult to change once systems are built against them, and a snapshot-based posting regime, once implemented, will not be retrofitted into an archival one without cost that nobody will volunteer to bear. This is the rare case where a technical specification decided in the next few months determines whether a policy question can be answered at all in five years. It deserves attention disproportionate to its apparent triviality.

Within current proceedings. Recommendations 3, 4, and 7 concern tariff terms and evidentiary requirements already before commissions. None requires new data, new authority, or new analytical capability. Denominating curtailment in energy rather than hours, in particular, is a drafting correction rather than a policy change, and one that a commission can make on the existing record once the distinction is pointed out.

Over one to three years. Recommendations 2, 5, 6, and 8 require something to be built: an outcome-reporting practice, a granular hosting-capacity dataset, an analytical function inside advocates' offices. Each is worth beginning now precisely because the payoff is delayed, and because the alternative is to arrive in 2029 with the same evidentiary vacuum that exists today.

Last, and dependent on the rest. Recommendation 9, holding decision-support analysis to a published standard, cannot bind before the record that would make calibration possible exists. Requiring a calibration record in 2026 would require the impossible and be ignored accordingly. Requiring it in 2029, of tools that have by then had three years of postings to score themselves against, is reasonable and enforceable. The sequencing matters: the standard should be announced early and applied late, so that vendors and utilities build the logging that will later be required of them.

7. SCOPE: WHAT THIS FRAMEWORK DOES NOT DO

Because the framework is intended for use near regulatory processes, its boundaries need stating plainly rather than in a disclaimer.

It is analytical and advisory. It informs siting, procurement, and timing decisions, and its outputs are comparative assessments, ranges, cost decompositions, and scenario analyses.

It does not perform grid operations. It issues no dispatch instruction, setpoint, switching order, or protection setting, and has no control interface to any generator, substation, or energy management system. It works on planning horizons of months to decades, not operating horizons.

It does not replace interconnection studies, engineering analyses, or regulatory reviews. Feasibility, system impact, and facilities studies remain with transmission providers and grid operators. Power-flow, short-circuit, stability, protection-coordination, and thermal-rating analyses remain with licensed professional engineers. Prudence review, cost allocation, and tariff approval remain with regulators. Where the framework estimates what a study will conclude, that estimate is labelled as a screening figure the authoritative process supersedes; when an authoritative result arrives it replaces the estimate rather than being averaged with it, because blending a screening estimate with a licensed determination would blur exactly the boundary this section draws.

Responsibility divides along disciplinary lines. Framework design, the evaluation approach, and this paper are the authors'. The engineering inputs the framework consumes (equipment ratings, thermal and stability limits, protection assumptions, constructability and cost estimates, cooling engineering, and interconnection precedent) are externally supplied parameters requiring validation by licensed engineers and power-sector specialists. Each is held with a named validating professional, a validity date, and a stated range rather than a single value, so that it is always possible to say which professional stands behind which number and so that uncertainty in the inputs survives into the outputs. Expired or unvalidated parameters cause the affected output to be withheld rather than silently defaulted.

The framework is also designed against foreseeable misuse. No output may be presented as an interconnection study result or an engineering certification; exports carry scope language; and a query approaching a boundary returns the applicable parameters and directs the question to the appropriate authority rather than computing an answer. Confidential customer material is partitioned and excluded from cross-client use except in aggregated, consented form, a restriction sitting in genuine tension with the transparency Section 6.5 demands, and one the architecture does not resolve.

8. LIMITATIONS

Five limitations concern this study rather than the framework, and are stated separately for that reason.

The framework is unimplemented and has not been evaluated. Sections 4 and 6 describe a design and a set of policy inferences drawn from it; neither has been tested beyond the illustrative application. No claim of accuracy, or of superiority over any named alternative, is supported here.

No parameter is estimated. The application uses stipulated values throughout. In particular the relationship between accepted curtailment and unlocked capacity, which drives the most quotable result in Section 5.3, is assumed, and it is exactly the quantity Section 6.3 argues cannot be inferred from published aggregates. The staircase shape follows from tranched load arrival; the specific step locations do not.

The evidence base is substantially grey literature. Queue statistics, tariff design, regulatory developments, and market outcomes are drawn from national-laboratory reports, regulatory filings, and trade press, because that is where they are published. Several claims, notably the details of the June 2026 FERC orders, rest on secondary legal-alert coverage and should be verified against the primary dockets by any reader relying on them for policy purposes.

The design reflects the authors' reading of documentary evidence. It has not been subjected to structured expert elicitation, and the user classes of Section 4 were derived analytically rather than from interviews with the people described.

Scope is United States-specific. The framework encodes ISO/RTO market design, FERC jurisdiction, and state tariff practice. Given how much of it is institutional rather than physical, transfer to other jurisdictions should not be assumed.

9. CONCLUSIONS

The constraint on powering data centres in the United States is not, in most locations, a shortage of energy resources. It is a coordination failure among parties who lack a shared, calibrated, auditable way to answer the question governing every one of their decisions: how much capacity, by when, at what cost, and with what exposure if the plan slips.

A framework organised around that question produces three conclusions conventional analysis does not. The efficient supply plan depends on whether one plans against the median or the tail, so a recommendation that does not state its percentile is not interpretable. Correlated project failure, not the number of projects, dominates the uncertainty in a utility's queue, so derating requests one at a time understates the planning range by a factor that matters. And flexibility buys speed in discrete steps, so obligations priced on a smooth curve are set at the wrong level almost everywhere.

Each has a counterpart in rules being drafted now. FERC's posting requirements can create the outcome record that would let load forecasts be calibrated for the first time, but only if the format preserves what was knowable when. Minimum-take obligations should step with a contracted ramp rather than applying against nameplate. Flexibility tariffs should be denominated in energy rather than hours, and structured around the steps rather than a slope. And cost allocation should be analysed from both sides, with regulators and consumer advocates equipped to interrogate the same figures the applicant relies on.

The framework's most consequential commitment is not analytical but epistemic: to bind the confidence it displays to a calibration record it publishes, and, until that record exists, to say so plainly. That commitment is specified here rather than discharged. Whether it is honoured, by this framework and by the commercial tools now shaping billions of dollars of siting and tariff decisions, is the test that ultimately matters.

References

1. Ayyildiz, E., Yildirim, B., & Aydin, N. (2025). Location selection methodology for data center with renewable energy integration. Renewable Energy, 250, 123270. https://doi.org/10.1016/j.renene.2025.123270
2. Build.inc. (2026). Data center power delivery schedule tracking [Product description]. Retrieved August 2026 from https://build.inc/insights/data-center-power-delivery-schedule-tracking
3. Chen, Y., & Zheng, X. (2026). To defer or to shift? The role of AI data center flexibility on grid interconnection (arXiv:2604.05376). https://arxiv.org/abs/2604.05376
4. Colangelo, P., Coskun, A. K., Megrue, J., Roberts, C., Sengupta, S., Sivaram, V., Tiao, E., Vijaykar, A., Williams, C., Wilson, D. C., Records, B., MacFarland, Z., Dreiling, D., Morey, N., Ratnayake, A., & Vairamohan, B. (2026). AI data centres as grid-interactive assets. Nature Energy, 11, 254–261. https://doi.org/10.1038/s41560-025-01927-1
5. Enverus. (2026). Data center siting solutions [Product description]. Retrieved August 2026 from https://www.enverus.com/products/data-center-siting-solutions/
6. Ewim, D. R. E., Ninduwezuor-Ehiobu, N., Orikpete, O. F., Egbokhaebho, B. A., Fawole, A. A., & Onunka, C. (2023). Impact of data centers on climate change: A review of energy efficient strategies. The Journal of Engineering and Exact Sciences, 9(6), 16397-01e. https://doi.org/10.18540/jcecvl9iss6pp16397-01e
7. Hussain, S. N., Al-Mandhari, M., Ali, S. M. F., Zaib, A., & Ghosh, A. (2026). GIS-driven regional assessment for sustainable data center siting in the United Kingdom. Land, 15(3), 516. https://doi.org/10.3390/land15030516
8. Idaho National Laboratory. (2025). Bridging the gap for powering data centers: Technical workshop report (INL/RPT-26-89901). https://inl.gov/content/uploads/2026/01/Bridging-the-Gap-for-Powering-Data-Centers.pdf
9. Johnston, S., Liu, Y., & Yang, C. (2023). An empirical analysis of the interconnection queue (NBER Working Paper 31946). National Bureau of Economic Research. https://www.nber.org/papers/w31946
10. Kahrl, F., & Frick, N. M. (2026). Speed to power: Solutions for accelerating large load connections. Lawrence Berkeley National Laboratory. https://emp.lbl.gov/publications/speed-power-solutions-accelerating
11. Kim, D., Dong, L., & Xie, L. (2026). Flexibility-aware framework for efficient planner-initiated siting of data center. Nature Communications, 17, 6512. https://doi.org/10.1038/s41467-026-72324-9
12. London Economics International. (2025). Uncertainty and upward bias are inherent in data center electricity demand projections. Prepared for the Southern Environmental Law Center. https://www.selc.org/wp-content/uploads/2025/07/LEI-Data-Center-Final-Report-07072025-2.pdf
13. Malczewski, J. (2006). GIS-based multicriteria decision analysis: A survey of the literature. International Journal of Geographical Information Science, 20(7), 703–726.

14. Martin, E., & Peskoe, A. (2025). Extracting profits from the public: How utility ratepayers are paying for Big Tech's power. Environmental and Energy Law Program, Harvard Law School. https://eelp.law.harvard.edu/extracting-profits-from-the-public-how-utility-ratepayers-are-paying-for-big-techs-power/
15. Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. Science, 367(6481), 984–986. https://doi.org/10.1126/science.aba3758
16. Mytton, D., & Ashtine, M. (2022). Sources of data center energy estimates: A comprehensive review. Joule, 6(9), 2032–2056. https://doi.org/10.1016/j.joule.2022.07.011
17. Nira Energy. (2026). Data centers [Product description]. Retrieved August 2026 from https://www.niraenergy.com/data-centers
18. 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 US power systems. Nicholas Institute for Energy, Environment & Sustainability, Duke University. https://nicholasinstitute.duke.edu/publications/rethinking-load-growth
19. North American Electric Reliability Corporation [NERC]. (2026). 2025 long-term reliability assessment. https://www.nerc.com/globalassets/our-work/assessments/nerc_ltra_2025.pdf
20. Pearl Street Technologies. (2026). SUGAR interconnection study automation [Product description]. Retrieved August 2026 from https://pearlstreettechnologies.com/
21. PJM Interconnection. (2025). PJM auction procures 134,479 MW of generation resources [Press release]. https://www.pjm.com/-/media/DotCom/about-pjm/newsroom/2025-releases/20251217-pjm-auction-procures-134479-mw-of-generation-resources.pdf
22. PVcase. (2026). Prospect: Data center siting [Product description]. Retrieved August 2026 from https://pvcase.com/data-center-siting
23. Rand, J., Cheyette, A., Talley, C., Zhang, S., Gorman, W., Wiser, R. H., Seel, J., Jeong, S., & Kahrl, F. (2026). Queued up: 2026 edition. Characteristics of power plants seeking transmission interconnection as of the end of 2025. Lawrence Berkeley National Laboratory. https://emp.lbl.gov/publications/queued-2026-edition-characteristics
24. Satchwell, A., Frick, N. M., Cappers, P., Sergici, S., Hledik, R., Kavlak, G., & Oskar, G. (2025). Electricity rate designs for large loads: Evolving practices and opportunities [Technical brief]. Lawrence Berkeley National Laboratory. https://eta-publications.lbl.gov/sites/default/files/2025-01/electricity_rate_designs_for_large_loads_evolving_practices_and_opportunities_final.pdf
25. Seel, J., Mulvaney Kemp, J., Rand, J., Gorman, W., Millstein, D., Kahrl, F., & Wiser, R. H. (2023). Generator interconnection costs to the transmission system. Lawrence Berkeley National Laboratory. https://emp.lbl.gov/publications/generator-interconnection-costs
26. Shehabi, A., Smith, S. J., Hubbard, A., Newkirk, A., Lei, N., Siddik, M. A., Holecek, B., Koomey, J. G., Masanet, E. R., & Sartor, D. A. (2024). 2024 United States data center energy usage report. Lawrence Berkeley National Laboratory. https://doi.org/10.71468/P1WC7Q
27. U.S. Energy Information Administration [EIA]. (2025, November 13). Short-term energy outlook. https://www.eia.gov/outlooks/steo/

28. Utility Dive. (2025, November 12). ERCOT's large load queue jumped almost 300% last year, official says. https://www.utilitydive.com/news/ercots-large-load-queue-jumped-almost-300-last-year-official/808820/
29. White & Case LLP. (2026, June). FERC orders grid operators to promptly revise or justify interconnection rules for data centers and large loads. https://www.whitecase.com/insight-alert/ferc-orders-grid-operators-promptly-revise-or-justify-interconnection-rules-data
30. Williams, C., Colangelo, P., Coskun, A., Levine, E., Neale, A., Roberts, C., Sengupta, S., Shirolkar, N., Sivaram, V., Soares, S., Tiao, E., Underwood, S., Wilson, D., Sharp, F., Wainwright, L., Petty, H., Wallace, S., & Records, B. (2026). Power-flexible AI data centers: A new paradigm for grid-responsive compute (arXiv:2606.25098). https://arxiv