THE IMPACT OF AI GROWTH ON THE POWER SYSTEM AND THE ROLE OF ENERGY REGULATORS

How Balance Can Be Achieved: A Regulatory Framework Across Five Dimensions

Working paper by ***Luca Lo Schiavo***

Regulatory Specialist, ***ERRA*** *– Energy Regulators Regional Association*

30 March 2026

This working paper was prepared as an original submission for the ICER Distinguished Scholar Award ("Next Practices" category), in connection with the XI World Forum of Energy Regulation (WFER), Tbilisi, 2026. It has not been previously published and does not replicate any prior work of the author, although it builds on and expands arguments developed in previous working papers and articles duly cited.

The author wishes to thank Pres. Jorge Vasconcelos, former Chair of ERSE (Portugal) and member of ERRA's Strategic Advisory Board, for a conversation that gave this paper its central thesis: that the energy transition has placed sustainability alongside efficiency as a co-equal regulatory objective, with structural consequences for how independence itself must be understood.

Energy Regulators Regional Association (ERRA)
secretariat@erranet.org • https://erranet.org/
Tel.: +36 1 477 0456 • Mob.: +36 70 392 5986
Logodi utca 44/B, 1012 Budapest, Hungary

Summary

Executive Summary

Artificial intelligence is transforming the electricity system from both sides of the meter simultaneously: as an optimization tool for system operations, including renewable integration and grid resilience, and as one of the most intensive categories of electricity demand in the history of the sector. This article examines the regulatory challenges that the rapid growth of AI-driven data center demand is generating for power system governance, drawing on recent regulatory decisions, technical documentation, and policy proposals from North America, Europe, and Australia.

The analysis is organized around four substantive dimensions — long-term predictability, operational reliability, economic affordability, and environmental sustainability — that trace a spectrum from the primarily technical to the irreducibly political. A fifth dimension, regulatory independence, cuts across all four: not a separate domain of analysis, but the institutional condition that determines whether (and which) governance of the other four is possible at all.

On long-term predictability, aggregate US utility load forecasts submitted to FERC have increased sixfold in three years, from 24 GW to 166 GW of projected additional peak demand by 2030. The gap between utility projections and independent analyst estimates —around 40 percent overestimation relative to market-based benchmarks — is not primarily a data problem. It reflects a structural misalignment of incentives: under cost-of-service regulation, utilities earn returns proportional to their rate base, making overestimation of demand growth a rational response to how they earn returns. As a priority, regulators shall engage in mitigating capex-bias for utilities that provide demand forecasts; moreover, methodological standardization of large load forecasts, probabilistic rather than deterministic planning assumptions, and cross-utility coordination on hyperscaler location strategies are further remedies.

On operational reliability, large AI data centers introduce specific patterns and features — instantaneous power-electronic ramping, voltage sensitivity, absence of rotational inertia, correlated simultaneous disconnection — that existing grid standards were not designed to manage. In US and Canada, NERC has documented incidents in the Eastern Interconnection and ERCOT involving data center-driven load trips of up to 1,500 MW. In Europe, ENTSO-E warns that the postponement of revision of EU-wide Network Code on Demand Connection is unlikely to influence the 2030 energy mix, given implementation timelines. In Australia, AEMO's decade of experience managing inverter-based generation provides a transferable technical template. The lesson across all three systems: regulatory frameworks take several years to build from first recognition of the risk to binding requirements. Given current growth trajectories, that lead time must be initiated before the first large-scale incident occurs.

On economic affordability, the cross-subsidy risk is structural: when grid infrastructure triggered by large load connections is socialized across all ratepayers, households and small businesses bear the cost of hyperscale investment. The Australian Energy Regulator's proactive scrutiny of utility demand projections and cost-reflective connection standards, Ireland's non-firm flexible connection framework, and Spain's competitive demand access tender — which scores applicants on greenhouse gas emissions avoided — represent the most coherent regulatory responses. The evidence from three continents is consistent:

where independent regulators have been given the mandate and the will to enforce the causer-pays principle, they have done so effectively.

On environmental sustainability, no energy regulator can govern alone the full set of questions that data center sustainability raises. Additionality requirements, temporal matching standards, and dispatchable generation conditions fall within the energy regulator's toolkit. Land use planning, water allocation, and district heating integration do not (or not necessarily). The Irish CRU's explicit acknowledgment that medium-term data center governance requires "*a State-led approach encompassing spatial planning, targets-based and plan-led infrastructure development*" is the correct institutional response: identify the boundary of the mandate precisely, exercise what authority exists within it rigorously, and pass the baton explicitly to the actors who hold the remaining instruments.

On regulatory independence — the institutional condition that runs beneath the other four dimensions — the comparative evidence reveals a spectrum of institutional configurations: from FERC, where formal independence coexists with a sustainability agenda systematically excluded by political framing, to Italy's ARERA, where a regulator with both energy and environmental mandate has been rendered structurally irrelevant by government choices in the strategy for attractivity of DC investments. Between them, the Irish case demonstrates honest institutional self-awareness, and the Spanish one demonstrates what productive government-regulatory collaboration can produce. The lesson is that independence is necessary but not sufficient: when the challenge crosses institutional boundaries, the appropriate response is deliberate interdependence — structured coordination with climate, planning, natural resources (land and water) regulators and industrial policy actors, with transparent handoffs at the boundaries of each mandate.

The actionable conclusion for energy regulators is that the toolkit to govern AI-driven load growth already exists. The gap between what regulation could achieve and what it is currently achieving is primarily institutional and political, not technical. Across the four substantive dimensions, independent regulators with a clear mandate and the will to exercise it have demonstrated the capacity to manage the challenge effectively. Whether that capacity is deployed depends on the fifth: the institutional condition of independence itself, and its deliberate extension — where the problem crosses jurisdictional boundaries — into interdependence.

Three conditions determine whether balance is achievable: mandate clarity, proactive engagement, and institutional coordination. Where these conditions are in place, the evidence shows that effective governance of AI infrastructure is within reach. Where they are not, the evidence is equally clear about the cost.

Introduction

Artificial intelligence is simultaneously the most powerful optimization tool the energy sector has ever encountered and one of its most demanding new consumers. It can sharpen the forecasting of variable renewable generation, improve the dispatch efficiency of congested grids, and accelerate the integration of electric vehicles and heat pumps into system operations. It can also, through the data center infrastructure it requires, add tens of gigawatts of new demand to systems that took decades to build, strain the transmission networks that serve existing consumers, and — if the energy sourcing is not carefully managed — entrench fossil fuel generation for years beyond the dates at which it was expected to retire.

In short, where not properly regulated, AI development can jeopardize energy transition.[1] This article examines the regulatory challenges that the rapid growth of AI-driven data center demand is generating for power system governance, and assesses the tools and frameworks that regulators in North America, Europe, and Australia are developing in response.

The analysis is organized around four substantive dimensions — *long-term predictability*, *operational reliability*, *economic affordability*, and *environmental sustainability* — that trace a spectrum from the primarily technical to the irreducibly political. A fifth dimension, *regulatory independence*, cuts across all four: it is not a separate domain of analysis but the institutional condition that determines whether governance of the other four is possible at all. The chapters that follow develop this argument in turn, before the Conclusions examine how independence and interdependence must be combined to achieve balance.

The four substantive dimensions are not independent of each other: a connection policy that addresses affordability by requiring data centers to bring their own generation capacity has direct implications for sustainability if that generation burns gas; a forecasting framework that overstates demand may produce stranded infrastructure that worsens affordability for decades. The analytical value of treating them separately lies not in pretending they are separable, but in giving each the focused attention it requires before examining how they interact.

A thread runs through all five: data centers bring the challenge, and regulation must provide the solution. The operators of large AI infrastructure are making rational decisions within the frameworks they face; the outcomes those decisions produce — for grid security, for ratepayers, for the climate — are a function of regulatory design, not of the intent of any individual actor. The appropriate instrument on the first three dimensions is independent expertise exercised within a clear mandate; the appropriate instrument where sustainability and democratic legitimation enter is *interdependence* — coordinated action across institutional boundaries, with each actor exercising its own mandate and transparent handoffs at the points where that mandate runs out.

[1] Ernesto Bonafé, "Is the Energy Transition a *'Con Job'* or rather the *'Trend of our Times'*?", *Global Energy Law and Sustainability*, vol. 6, n. 2, 19 February 2026, https://doi.org/10.3366/gels.2025.0140. Bonafé frames his question around the energy transition broadly; this article focuses on one of its most consequential stress tests — the effects on power system, including sustainability, of AI-driven demand growth.

1. Long-Term Predictability: When Forecasts Become the Problem

In May 2024, construction began on a data center in southwest Memphis, Tennessee. By Labor Day weekend — less than three months later — it was operational. Colossus, the training facility for xAI's Grok model, occupies a building larger than a dozen football fields on the site of a former oven factory, beside a steel mill and a rail yard. To connect it to power fast enough, xAI did not wait for the local grid: it installed up to 35 natural-gas turbines on-site and argued that no permit was required because the units would run for less than a year. When fully operational, Colossus and two nearby xAI facilities are projected to draw nearly two gigawatts of power: roughly twice the annual electricity consumption of Seattle.[2]

The speed is not incidental. It is the business model. The capital expenditures of Amazon, Microsoft, Meta, and Google have exceeded $600 billion since the launch of ChatGPT in November 2022 — more, in inflation-adjusted terms, than the federal government spent to build the entire interstate highway system — and much of that spending has gone toward data centers that each company is racing to bring online before its competitors. OpenAI has announced facilities requiring more than 30 gigawatts of total power: more than the highest recorded demand peak for all of New England. *"These are the largest single points of consumption of electricity in history,"* Jesse Jenkins, a climate modeler at Princeton University, has observed. Dominion Energy, the dominant utility in Virginia — home to Loudoun County, which alone hosts 13 percent of global data-center capacity — describes what it is witnessing as *"the largest growth in power demand since the years following World War II"*.[3]

For two decades, electricity demand in the United States was essentially flat. Grid operators, system planners, and regulators built their institutions, standards, and reliability frameworks around that stability. The question that Chapter 1 examines — whether the demand forecasts now being filed by utilities are accurate — matters enormously for the system cost.[4] This chapter examines why forecasting this transformation has proved so difficult — and why the difficulty is not merely technical.

Section 1.2 identifies the methodological failures that occur when statistical tools designed for incremental demand are applied to step-change loads. Section 1.3 argues that behind those methodological failures lies a structural misalignment of incentives: in cost-of-service regulatory regimes, overestimating demand growth is not an error but a rational response to how utilities earn returns. Sections 1.4 and 1.5 compare how three governance systems — NERC in North America, AEMO in Australia, and ENTSO-E in Europe — have responded to the challenge, and identify the institutional conditions under which accurate, consistent forecasting of transformative new loads becomes possible.

---

[2] Matteo Wong, "Inside the Dirty, Dystopian World of AI Data Centers," *The Atlantic*, March 13, 2026.

[3] *Ibidem*. Jenkins directs the ZERO Lab at Princeton University, whose research on flexible grid connections is also discussed in Chapter 3 (see footnote 35).

[4] L. Lo Schiavo, "*AI Data Centers and Energy Regulation: When Traditional Frameworks Meet Unprecedented Demand Growth*", Global Energy Law and Sustainability, Vol 6 issue 2 (hereinafter: Lo Schiavo, *AI Data Centers and Energy Regulation,* 2026*),* https://www.euppublishing.com/doi/10.3366/gels.2025.0141, in particular ch. 4 on the affordability crisis due to increase of reliability auctions costs in PJM related to forecasts of DC demand. See also in this paper sect. 3.3 and Conclusions.

1.1 A Structural Discontinuity

For more than two decades, electricity demand in most advanced economies barely moved. In the United States, annual consumption grew at under 1% per year throughout the 2000s and 2010s. Grid planners built models around incremental residential electrification, slow commercial growth, and occasional large industrial arrivals — all highly predictable. Then, in the span of three years, the picture changed completely.

By 2025, aggregate utility load forecasts submitted to the US Federal Energy Regulatory Commission (FERC) — the authority that collects five-year demand projections from all major utilities — were projecting **166 GW of additional peak load between 2025 and 2030**.[5] In 2022, the same process had yielded a figure of just 24 GW. A six-fold increase in forecast demand growth, in three years, for the same five-year horizon. No new industrial revolution. No sudden population surge. The culprit, at least nominally, is one: data centers built to run artificial intelligence workloads.

Data centers accounted for approximately 4.4% of US electricity consumption in 2023.[6] Lawrence Berkeley National Laboratory projects that figure could reach **12% by 2028** — representing roughly 580 TWh annually. In Europe, ENTSO-E's 2025 TYNDP scenarios project that energy demand from data centers and the broader ICT sector will **triple by 2030** compared to 2019 levels.[7] In Australia, AEMO's 2025 Electricity Statement of Opportunities (ESOO) shows data centers reaching **10% of nation demand by 2035** under its Step Change scenario, with independent forecasts commissioned from Oxford Economics pointing to 32 TWh by 2050.[8] Ireland offers perhaps the starkest regional example: data centers grew from 5% of total electricity consumption in 2015 to **21% in 2023**, accounting for 85% of all demand growth over that period.[9]

These are not marginal adjustments to existing planning frameworks. They are structural discontinuities — the kind that expose whether regulatory systems are built for the world they were designed to govern, or only for the world as it was.

1.2 Why Traditional Forecasting Fails

The core problem is methodological. Utilities have built their long-term demand forecasts on regression-based and autoregressive time series models: sophisticated statistical tools

[5] Grid Strategies LLC, *"National Load Growth Report 2025. Power Demand Forecasts Revised Up for Third Year Running, Led by Data Centers",* by J. D. Wilson, S. Meyer, Z. Zimmerman, and R. Gramlich, November 2025, https://gridstrategiesllc.com/wp-content/uploads/Grid-Strategies-National-Load-Growth-Report-2025.pdf hereinafter: Grid Strategies, *National Load Growth*, 2025)

[6] Lawrence Berkeley National Laboratory, *"2024 United States Data Center Energy Usage Report"* by A. Shehabi et al., December 2024, https://escholarship.org/content/qt32d6m0d1/qt32d6m0d1.pdf

[7] ENTSO-e, *"Position on the need for national connection requirements to ensure EU power system stability,"* approved 11 December 2025, https://www.entsoe.eu/2025/12/22/entso-e-position-on-the-need-for-national-connection-requirements-to-ensure-eu-power-system-stability/ (hereinafter: ENTSO-e Position on Connection Requirements, 2025). Forecast data from ENTSO-e TYNDP 2026 scenarios (cut-off date 24 December 2024), https://2026.entsos-tyndp-scenarios.eu/download/

[8] AEMO, *"2025 Transition Plan for System Security,"* version 1.1, February 2026, Section 10.3 (Large Inverter-Based Loads), https://www.aemo.com.au/energy-systems/electricity/national-electricity-market-nem/nem-forecasting-and-planning/transition-planning/

[9] Commission for Regulation of Utilities (CRU), *"Large Energy Users Connection Policy – Proposed Decision,"* CRU202504, February 2025 https://consult.cru.ie/en/consultation/review-large-energy-users-connection-policy (hereinafter: CRU *Large energy users proposal*, 2025); Ireland data: EirGrid/SONI, *"Generation Capacity Statement 2023–2032,"* January 2024.

that work excellently when past behavior is a reliable guide to future behavior. For decades, that assumption held. Residential and commercial demand moved in step with GDP, weather patterns, population, and appliance saturation — all slowly varying and historically documented variables.

Large AI data centers break every one of these assumptions. They arrive not incrementally but as step changes. A single campus can exceed 1 GW — roughly equivalent to the peak demand of a mid-sized European city. Their power consumption depends on computational workloads that follow no seasonal or demographic logic. Their location is determined by land availability, fiber connectivity, tax incentives, and proximity to power — not by where population lives or industries operate. Their construction timeline (typically 12–18 months for a building, but years for full campus buildout) creates a sharp and sudden energization date. And many of the largest operators are simultaneously developing sites in multiple regions, creating a geographic arbitrage dynamic that individual utilities — who see only their own service territory — cannot observe.

The Energy Systems Integration Group (ESIG), in its December 2025 task force report on large load forecasting,[10] identified nine findings after reviewing more than two dozen utility load forecasts across the United States. The first and most fundamental: **large load forecast methods lack transparency and consistency**. Different utilities apply different assumptions to the same type of load, and most do not describe their methodology in a way that allows cross-comparison. The report formalized five metrics that any rigorous large load forecast should quantify:

Metric	Definition
Project realization rate	Share of announced projects that actually enter service
Energization date	When commercial operation begins, accounting for realistic delays
Load realization	Actual peak demand as a percentage of the contracted request
Load ramping	Monthly demand profile during the ramp-up period
Load factor / load shape	Average energy use relative to peak capacity; hourly demand profile

Table 1 – Metrics for large load forecast (*source*: ESIG report)

The gap in current practice is systematic: most utility forecasts apply the same realization assumptions to all large loads regardless of their type, maturity, or financial commitments — treating a signed interconnection agreement from a hyperscaler with a fixed commissioning date the same as an early-stage inquiry from a speculative developer. As ESIG noted, very few forecasts differentiate among subcategories of data centers: traditional cloud, AI training (which can exhibit violent intra-hour load spikes), cryptocurrency mining (which is highly price-sensitive and can curtail completely in seconds), and hybrid configurations.

NERC's July 2025 white paper on emerging large loads echoes this: *"The electrical demand is different for training an AI model versus using an AI model to create inferences"*.[11] But at

[10] Energy Systems Integration Group (ESIG), *"Forecasting for Large Loads: Current Practices and Recommendations,"* Large Loads Task Force, December 2025, https://www.esig.energy/large-loads-task-force/forecasting/ hereinafter: ESIG *Forecasting for Large Loads,* 2025

[11] NERC Large Loads Task Force, *"Characteristics and Risks of Emerging Large Loads,"* White Paper, July 2025, Figure 2.2 (AI Training Data Center Demand Curve, EdgeTunePower)

the same time, AI load is not categorically tracked in any publicly available utility forecast. There is also a systematic back-end risk blind spot. ESIG's Recommendation 3 points out that most forecasts address front-end uncertainty (will the project connect? when?) but almost entirely ignore back-end uncertainty: what happens if a large load customer reduces demand, cancels, relocates, or moves workloads behind the meter? The asymmetry is telling — forecasters worry more about under-serving demand than about stranding infrastructure — which, as we shall see, has a structural explanation.

1.3 The Incentive Problem

Behind the methodological failures lies something more troubling: a misalignment of incentives. The ITIF (Information Technology & Innovation Foundation) published an analysis in November 2025 that identifies a structural bias in the US utility regulatory model toward capital overinvestment.[12] Under traditional cost-of-service regulation, utilities earn returns proportional to their rate base — the value of physical assets in service. More infrastructure built means higher revenues. Within this framework, overestimating demand growth is not an error: it is the rational outcome of incentive structures that were designed for a different era.[13]

As the ITIF report notes, since 2006 FERC has approved more than $53 billion in transmission investments through 123 incentive-based development programs. The Grid Strategies November 2025 national load growth report documents the result: aggregate utility forecasts project roughly 90 GW of data center demand by 2030, while market analysts — using methods based on actual chip shipments (TD Cowen), tracked project databases (Cleanview), and financial modeling (Bloomberg NEF, Wood Mackenzie, S&P Global, McKinsey) — converge around 65 GW. A gap of approximately 25 GW, equivalent to 40% overestimation of the market-based benchmark. To put it in tangible terms: 25 GW is roughly twice the peak demand of New York City, or about half of Italy's national peak demand.[14]

Grid Strategies is careful not to claim that utilities are consciously manipulating their forecasts. The data submitted to FERC are derived from customer-supplied information — connection requests, development plans, demand estimates — that utilities aggregate and weight using professional judgment. The bias enters through the weighting: under-applying cancellation rates, over-applying load factors, using optimistic energization timelines. Each individual assumption may seem defensible in isolation; in aggregate, the direction of error is consistently upward.

The problem is compounded by an invisible coordination failure. A utility planning in Virginia sees only the data centers in its service territory. It does not know — because data

---

https://www.nerc.com/globalassets/who-we-are/standing-committees/rstc/3_doc_white-paper-characteristics-and-risks-of-emerging-large-loads.pdf (hereinafter: NERC *Emerging Large Loads*, 2025)

[12] ITIF (Information Technology & Innovation Foundation), *"The United States Needs Data Centers, and Data Centers Need Energy, but That Is Not Necessarily a Problem,"* November 2025, https://itif.org/publications/2025/11/24/united-states-needs-data-centers-data-centers-need-energy-but-that-is-not-necessarily-a-problem/

[13] Amon European regulators, large attention is dedicated to mitigate "capex bias" from economic regulation of utilities. See CEER (Council of European Energy Regulators), Incentives in Regulatory Frameworks with a Focus on OPEX/CAPEX Neutrality, Report 5 May 2025, https://www.ceer.eu/wp-content/uploads/2025/05/CEER-Paper-on-incentives-in-regulatory-frameworks-with-a-focus-on-OPEX-CAPEX-neutrality.pdf

[14] Grid Strategies *National Load Growth,* 2025

center developers do not tell it — that the same hyperscaler has filed parallel interconnection requests in Georgia, Ohio, and Texas, and will ultimately select only one location. ESIG Finding 6 documents this explicitly: for the most part, individual utility forecasts do not consider whether prospective customers may be evaluating alternative sites, because utilities lack awareness of those alternatives.

1.4 Three Governance Models, Three Responses

How regulatory systems have responded to this forecasting challenge reveals a great deal about their institutional capacity to absorb structural discontinuities.

In North America, NERC's approach has been diagnostic. Its Large Loads Task Force produced a July 2025 white paper characterizing the phenomenon,[15] followed in March 2026 by a gap analysis identifying specific regulatory deficiencies.[16] The gap analysis is explicit: large loads are not NERC-registered entities, they face no standardized data-sharing or reporting obligations, and there is no national framework for transmission planning methodologies applicable to this class of customer. NERC identifies the forecasting problem but places it outside its own mandate — correctly, since demand forecasting is a state-regulated activity in the US, governed by individual Public Utility Commissions and aggregated by FERC. The gap analysis proposes modifying NERC's Rules of Procedure to reactivate a Load Serving Entity registration category (removed approximately a decade ago), and commits to beginning Reliability Standards development in 2027. It is a roadmap, not yet a solution.

The structural fragmentation of the US system — 66 entities filing separate five-year demand forecasts to FERC, each with different methodologies, different definitions of "large load," and different relationships to interconnection queues — means that even with NERC's reforms, achieving consistent national forecasting practice will require either federal legislative action or a FERC rulemaking of considerable breadth.

In Australia, the National Electricity Market operates under a fundamentally different architecture. AEMO serves simultaneously as system operator, market operator, and integrated planner — publishing the Integrated System Plan (ISP) every two years and the Electricity Statement of Opportunities (ESOO) annually. Both are produced by AEMO directly, using its own modeling, rather than aggregating data from independent utilities. The forecasting question for data centers in Australia is therefore not "how do we reconcile 66 different utility submissions?" but "how do we incorporate a new demand category into our own planning models?" The answer, documented in AEMO's December 2025 Transition Plan for System Security, is straightforward: AEMO has added data centers as a **separately forecasted demand category** in both ISP and ESOO methodology, commissioned an independent demand analysis from Oxford Economics (published 2025), and announced a dedicated LIBL (Large Inverter-Based Load) study in the 2026 Grid and Power System Review Report. New South Wales alone had received over 10 GW of data center connection enquiries by the end of the 2024–25 financial year, with individual facilities exceeding 1 GW. AEMO's response was methodologically integrated: the new demand category feeds

---

[15] NERC *Emerging Large Loads*, 2025

[16] NERC Large Loads Working Group, *"Assessment of Gaps in Existing Practices, Requirements, and Reliability Standards for Emerging Large Loads,"* White Paper, March 2026 https://www.nerc.com/globalassets/our-work/guidelines/reliability/white-paper---assessment-of-gaps.pdf (hereinafter: NERC *Assessment of Gaps*, 2026)

directly into transmission augmentation planning and system security assessments. The forecasting and the planning are institutionally co-located.

In Europe, the challenge takes a different form. ENTSO-E's December 2025 position paper on national connection requirements documents a structural gap not in forecasting per se, but in the regulatory framework surrounding demand forecasts: the review Network Code for demand connection (NC DC 2.0) — which would extend technical requirements to data centers and other new load types — would have been deprioritized by the European Commission, with no adoption timeline communicated.[17] ENTSO-E warns explicitly that the updated codes are *"unlikely to influence the 2030 energy mix"* given the time required for national implementation after EU adoption. Some Member States (Germany, Ireland, the Netherlands) have moved ahead with national measures, but the result is a patchwork that undermines the single market logic of harmonized grid connection standards. The EU forecasting architecture itself — centered on ENTSO-E's TYNDP (Ten-Year Network Development Plan), produced every two years — incorporates data center projections at the EU level through the National Energy and Climate Plans submitted by member states. However, these bottom-up national submissions face the same methodological inconsistencies that characterize US utility forecasts: different assumptions about realization rates, load factors, and locational distributions, aggregated without a common methodology.

1.5 Paths Forward

The emerging consensus across all three systems points toward a set of common remedies.

Methodological standardization is the first priority. ESIG's Recommendation 8 calls for a national database of large load forecast data, categorized by load type and tracking all five key metrics reported in Table 1 — allowing historical realization rates to inform probabilistic rather than deterministic forecasting. Grid Strategies' report documents that at least one utility (Georgia Power) has already developed a systematic model incorporating cancellation probabilities and load realization adjustments; the challenge is to generalize this practice.

Tariff and contract reform is the second lever. Grid Strategies documents a growing set of large-load-specific tariff mechanisms being adopted or proposed across US jurisdictions: minimum-bill obligations (customers pay for a minimum percentage of contracted capacity regardless of actual usage), long-term contracts (5–15 years), and creditworthiness requirements. These create financial incentives for data center developers to submit accurate demand information and reduce speculative connection requests. Texas SB 6 (enacted June 2025), Oregon's POWER Act, and proposals in Virginia and Indiana represent a wave of state-level responses to the forecasting distortion problem.

Flexible connection frameworks — explored in the Nicholas Institute/Duke University "Rethinking Load Growth" report,[18] cited by NERC itself — offer a third path: allowing large

[17] *"According to the latest information shared by the EC at the Grid Connection European Stakeholders Committee (GC ESC) chaired by ACER (on 18 Sept.2025), the adoption of CNC 2.0 has been deprioritised, and no specific timeline has been communicated so far"* (ENTSO-e *Position on Connection Requirements*, 2025).

[18] Nicholas Institute for Energy, Environment & Sustainability, Duke University, *"Rethinking Load Growth: Assessing the Potential for Integration of Large Flexible Loads in US Power Systems"*, by T. H. Norris, T. Profeta, D. Patino-Echeverri, A. Cowie-Haskell, 2025,

loads to connect faster and at lower infrastructure cost in exchange for accepting curtailment during grid stress events. A large load that accepts flexible service terms removes the need for the utility to plan firm infrastructure to its full contracted peak — reducing both the investment requirement and the incentive to overstate demand. The Irish energy (and water) regulator, CRU, has developed analogous “non-firm” connection mechanisms that provide approximately 50% savings on connection costs.

Finally, there is the question of **institutional architecture**. The US case illustrates the cost of forecasting fragmentation: 66 independent submissions, no standardized methodology, no visibility across service territories. Australia’s case illustrates the benefit of centralized integration: a single planner seeing the full system, incorporating data centers as a named category, and linking the demand forecast directly to transmission planning. For regulators designing or reforming their planning frameworks, two are the most fundamental lessons: first, **accurate long-term forecasting of transformative new loads requires not just better methods, but an institutional home with the authority and information to apply them consistently**. And second, regulators should engage in **minimizing capex-bias that provide utilities with perverse incentives to overestimate demand for connection**.

2. Operational Reliability: When Large Loads Become a System Security Challenge

2.1 A Shift in the Nature of Load

The traditional power system security framework rested on three assumptions about load: that it was passive (it consumed power but did not interact with grid conditions), predictable (its aggregate behaviour could be forecast with reasonable accuracy over operational timescales), and dispersed (no single consumer or cluster of consumers could materially alter system-wide frequency or voltage). Large data centres challenge all three assumptions simultaneously.

They are active in the sense that their internal control systems — uninterruptible power supplies (UPS), variable-frequency drives, power distribution units, and high-density cooling infrastructure — respond to grid conditions in ways that can amplify rather than absorb disturbances. They are unpredictable in that AI training runs, cryptocurrency mining cycles, and demand response activations can shift hundreds of megawatts within seconds, creating ramp events that exceed those associated with conventional generation outages. And they are concentrated: a single hyperscale campus may present a load equivalent to a mid-sized city, localised at a single point (or very few points) of interconnection, with consequences for system strength and short-circuit current levels that extend far beyond that point (or those points).

The regulatory response emerging across all three jurisdictions examined in this article — the European Union, the United States and Canada, and Australia — reflects a recognition that large electronic loads require treatment analogous to generation resources for

https://nicholasinstitute.duke.edu/publications/rethinking-load-growth hereinafter Nicholas Institute, Duke University *Rethinking Load Growth* 2025 reliability purposes. The sections that follow examine how that recognition has translated into regulatory action, or the absence of it, in each case.

2.2 The United States and Canada: incidents, analysis, and the limits of voluntary frameworks

2.2.1 The NERC Large Load Task Force

The North American Electric Reliability Corporation (NERC) formally recognised the reliability implications of large load growth in mid-2024, when its Reliability and Security Technical Committee (RSTC) established the Large Load Task Force (LLTF). The LLTF's mandate was to characterise the risks posed by emerging large loads and to assess whether existing Reliability Standards, planning practices, and operational procedures were adequate to address them.

The Task Force published its first white paper in July 2025.[19] That document remains the most comprehensive systematic analysis of large load reliability risks produced by any regulatory body to date. It identifies seven broad risk categories: observability and data quality; long-term resource adequacy planning; operations and balancing; stability (encompassing ride-through, voltage, angular, resonance, and forced oscillations); power quality; cyber and physical security; and load shedding and restoration. Of these, the LLTF assigns high priority to four: ride-through capability deficiencies, voltage stability under large load concentrations, angular stability during rapid load changes, and reserve adequacy given the increased variability of large electronic loads. A second white paper, published later in 2025, catalogued the gaps between the identified risks and existing NERC Reliability Standards.[20]

2.2.2 Documented incidents

The white paper analyses several operational incidents that illustrate the systemic character of the risk. The fault ride-through capability is particulary relevant: "a 230 kV transmission line fault led to customer-initiated simultaneous loss of approximately 1,500 MW of voltage-sensitive load that was not anticipated by the BES operator" on 10 July 2024, due to a lightning arrestor failure on a 230 kV transmission line in the Eastern Interconnection that triggered a sequence of faults over 82 seconds.[21] The incident demonstrated that large loads can have effects of the same order of magnitude (and opposite versus in change), from the system operator's perspective, as unscheduled generation trips. Simultaneous large load losses have two effects on the electric system: First, frequency rises on the system as a result of the imbalance between load and generation; second, voltage rises rapidly because less power is flowing through the system.

The ERCOT interconnection has documented 26 events at cryptocurrency mining facilities between 2023 and 2025 in which load reductions of 17 to 95 percent occurred during transmission fault events. The root causes are technically distinct from those affecting AI data centres. Cryptocurrency mining facilities typically lack UPS systems; their constant-power switching supplies draw increasing current as voltage depresses, exacerbating

[19] NERC *Emerging Large Loads,* 2025

[20] NERC *Assessment of Gaps,* 2026

[21] NERC, *Incident Review: Considering Simultaneous Voltage-Sensitive Load Reductions*, January 2025 https://www.nerc.com/globalassets/our-work/reports/event-reports/incident_review_large_load_loss.pdf ; NERC *Emerging Large Loads*, 2025, Chapter 3, pp. 23–24; p. 20 (xAI Colossus ramp rates); pp. 28–31 (resonance and converter-driven stability; forced oscillation case, Dominion system, 2023).

rather than buffering the disturbance; VFD cooling systems are programmed to trip at undervoltage thresholds as high as 0.96 per unit; and neutral overcurrent protection can disconnect entire facilities during asymmetric faults. Recovery to full load requires minutes to hours, depending on the restart sequence required.[22]

AI training facilities present a different risk profile. The xAI Colossus facility in Memphis — at the time of the LLTF white paper, the world's largest AI training cluster — has been documented shifting between 35 and 70 MW within one minute as training runs start and stop. On a larger scale, a single large AI campus can ramp from approximately 450 MW to 7 MW in 36 seconds, a ramp rate that materially affects the Area Control Error (ACE) of the balancing authority. A separate incident involved a 1 Hz forced oscillation traced to UPS units at a data centre in the Dominion Virginia service territory in 2023, demonstrating that large electronic loads can introduce power quality disturbances with system-wide propagation characteristics.

2.2.3 The structural gap

The most significant finding of the Large Load Task Force (LLTF) analysis is institutional rather than technical. Large loads are not NERC-registered entities. They are not subject to Reliability Standards. They are not required to provide modelling data to transmission planners, real-time telemetry to system operators, or post-event reports to reliability coordinators. The entire analytical and operational visibility infrastructure that governs generation resources — registration, modelling obligations, telemetry, standards compliance, enforcement — does not extend to load, regardless of size.

The invisibility problem has a further dimension that existing frameworks are not equipped to address. The largest hyperscalers already practice geographic workload shifting — routing computational tasks dynamically to whichever facility in their global network is served by the cleanest or cheapest energy in a given hour.[23] From a sustainability perspective, this is an optimization tool; from a system operator's perspective, it is an unobservable demand event. A balancing authority managing frequency and reserves cannot distinguish a data center reducing load because of a grid disturbance from one reducing load because an algorithm has redirected a training run to a facility in another continent. NERC has flagged this practice explicitly as a material operational risk: the triggers for large load movements are invisible to the operators responsible for system security.[24] The sustainability opportunity and the reliability risk are thus two sides of the same regulatory gap.

The LLTF has identified the registration threshold question as one requiring careful design. The figure of 75 MW appears frequently in discussions, but the Task Force has cautioned against reliance on a single-number definition: ramp rate, voltage sensitivity, power electronics content, and behind-the-meter configuration all affect whether a facility poses systemic reliability risks independent of its nameplate demand. A 100 MW load with high ramp variability and no ride-through capability may be more consequential for system

[22] NERC *Assessment of Gaps,* 2026, pp. 15–16, p. 38 (ERCOT cryptocurrency mining incidents).

[2323] Google has declared that they already practice *geographic workload shifting* — routing computational tasks to whichever facility in a global network is served by the cleanest energy in a given hour: see https://blog.google/innovation-and-ai/infrastructure-and-cloud/global-network/data-centers-work-harder-sun-shines-wind-blows/

[24] NERC, *Emerging Large Loads,* 2025, because system operators lack visibility into load movements whose triggers they cannot observe.

security than a 300 MW load with stable demand and well-designed power conversion equipment.

2.3 The European Union: from generation-centric codes to large-load requirements

2.3.1 The ENTSO-E position

In December 2025, the European Network of Transmission System Operators for Electricity (ENTSO-e) published a position paper on the need for harmonised national technical connection requirements for large energy users (LEUs).[25] The document marked the first explicit recognition at the pan-European level that the Network Codes applicable to generation and conventional demand are structurally inadequate for large power electronic loads.

ENTSO-e's analysis focuses on three specific gaps. First, existing connection requirements do not specify ride-through capability obligations for large loads — that is, the ability to remain connected and continue drawing power through voltage and frequency disturbances within defined envelopes. Second, there is no standard specifying the reactive power behaviour of large electronic loads during disturbances; uncontrolled reactive absorption can destabilise voltage in regions with concentrated data centre demand. Third, there is no mandatory telemetry obligation for large loads above any threshold, leaving system operators without the real-time visibility needed to manage large-load events.

The position paper calls on member states to develop national connection requirements that address these gaps and on the European Commission to consider whether a revision of the relevant Network Codes — in particular the Requirements for Generators (RfG) regulation — is warranted to extend analogous obligations to large consuming units. The paper explicitly cites the RfG as the regulatory template: the principle that entities capable of materially affecting system stability should be subject to technical capability requirements and registration obligations, regardless of whether they are generators or consumers, is the central argument.

2.3.2 Ireland as the frontier case

Among European systems, Ireland presents the most advanced and most acute instance of the policy challenge.[26] The country hosts one of the highest concentrations of hyperscale data centre capacity relative to system size in the world. Data centres accounted for approximately 21 percent of total electricity consumption in Ireland in 2023, and projections by EirGrid suggest that this share could reach 30 percent by the end of the decade.

CRU's proposed decision of February 2025 on the review of the Large Energy Users connection policy acknowledges explicitly that the collective behaviour of data centres — simultaneous disconnection during disturbances followed by simultaneous reconnection as backup systems exhaust their energy reserves — is already exacerbating system disturbances on the Irish transmission system. EirGrid is developing mandatory Grid Code standards for data centres, including fault ride-through requirements that would oblige large loads to remain connected through defined contingency events rather than

[25] ENTSO-e *Position on Connection Requirements*, 2025; European Commission, Regulation (EU) 2016/631 (Requirements for Generators Network Code), cited by ENTSO-e as the generation-side precedent for future large-load requirements.
[26] CRU *Large energy users proposal*, 2025), pp. 40–43

transferring to backup generation. The CRU proposed decision treats this as an urgent reliability measure rather than a medium-term planning aspiration.

Ireland's situation is analytically significant beyond its own borders because it represents, in compressed form, the trajectory that other European systems will follow as data centre penetration increases. The technical problems now visible in Ireland — voltage support deficits, frequency response degradation, reserve adequacy under concentrated load variability — are predictive of conditions that will emerge in other member states within a five-to-ten year horizon.

2.3.3 Recent developments in Spain

Spain's approach illustrates the institutional division that effective governance requires. MITECO has initiated a draft Royal Decree establishing mandatory technical connection requirements for demand installations — including fault ride-through obligations, power quality standards, and oscillation damping requirements — explicitly designed to anticipate the EU Network Code revision that ENTSO-E has identified as unlikely to arrive before 2030.[27] In parallel, CNMC has opened public consultation on the flexible access permit framework that determines how those technical requirements translate into connection rights. The two instruments are complementary and institutionally coherent: the ministry sets the technical floor; the independent regulator governs the access conditions. Neither substitutes for the other. The contrast with Italy — where ARERA holds the mandate but has been *de facto* excluded from the data center policy process — is striking.

2.4 Australia: transferable experience from inverter-based generation

The Australian Energy Market Operator (AEMO) enters the data centre reliability discussion with an institutional advantage that neither ENTSO-e nor NERC possesses: a decade of operational experience managing the stability consequences of large-scale inverter-based resource (IBR) integration. The rapid expansion of utility-scale solar and wind generation in the National Electricity Market (NEM) has required AEMO to develop analytical tools and operational procedures for managing system strength shortfalls, inertia deficits, and fast frequency response gaps that are directly applicable to the reliability challenges posed by large electronic loads.

AEMO's 2025 Transition Plan for System Security identifies three system security services as simultaneously at risk: frequency control, system strength, and inertia.[28] The plan notes that the erosion of these services reflects the simultaneous retirement of synchronous generation and the addition of large power-electronic resources — both on the generation and, increasingly, on the consumption side.

Australia does not yet have data center-specific Grid Code standards of the kind that Ireland is developing. However, the institutional machinery already in place — the Rules framework, connection registration and modelling requirements, system strength assessment methodology, and the market for inertia and frequency services — provides a

---

[27] Ministerio para la Transición Ecológica, *Proyecto de Real Decreto por el que se establecen requisitos mínimos de diseño, equipamiento, funcionamiento y seguridad de las instalaciones que se conecten a las redes de transporte y distribución de energía eléctrica, audiencia pública 24 febrero – 16 marzo 2026* (not yet adopted at time of writing). CNMC*, Propuesta de resolución sobre permisos de acceso flexible de la demanda, consulta pública marzo 2026* (RDC/DE/003/25).
[28] AEMO, *2025 Transition Plan for System Security*

ready infrastructure for extending reliability obligations to large loads. The transition from general IBR rules to data-centre-specific requirements is primarily a technical calibration task rather than a foundational institutional design challenge.

2.5 Comparative assessment

The three jurisdictions share a common diagnostic: existing frameworks designed for passive, predictable, dispersed load are structurally inadequate for large electronic loads that are active, variable, and geographically concentrated. The convergent regulatory response — mandatory ride-through requirements, real-time telemetry obligations, system security assessments as a condition of connection — reflects the same underlying logic across all three cases. The differences lie in institutional velocity and the starting point from which each jurisdiction is moving.

Dimension	EU / ENTSO-e	US–Canada / NERC	Australia / AEMO
Documented incidents	Ireland: DC collective disconnect/reconnect exacerbates faults (EirGrid, 2024–25)	EI: 1,500 MW loss Jul 2024; ERCOT: 26 crypto events 2023–25	No DC-specific incidents documented; IBR experience provides precedent
Ride-through requirements	ENTSO-E calls for harmonised national rules; Ireland Grid Code FRT under development	ERCOT LEL rules in progress; NERC Reliability Guideline in development	No DC-specific FRT rule yet; IBR rules in NEM provide template
Registration / telemetry	ENTSO-E calls for mandatory telemetry; not yet binding EU-level law	Large loads not NERC-registered; gap identified as top priority	Connection registration framework exists; system strength assessments required at POI
System strength / inertia	No EU rules yet; RfG Network Code cited as model for future large-load requirements	Resonance and converter stability identified as medium-priority risk in LLTF white paper	Most advanced: inertia, system strength, FFR formally co-optimised in NEM design
Regulatory stage	Policy framework emerging; Ireland leads implementation	Risk taxonomy complete; standards in development	Rules framework mature; DC-specific rules pending

Table 2 – Summary of reliability issues related to data centers in three regions (EU, US-CAN, AUS).

AEMO operates the most technically complete framework for managing power electronic loads in general — inertia, system strength, and fast frequency response are formally co-optimised within the NEM design — but this maturity was built to address the consequences of renewable generation, not of large consumption loads. Adapting it to data centres is a near-term institutional task rather than a foundational one.

NERC has produced the most analytically rigorous taxonomy of risks — the LLTF white paper and its gap analysis represent a level of systematic documentation that neither ENTSO-e nor AEMO has matched — but NERC's institutional role is advisory rather than prescriptive, and the translation of its risk analysis into binding Reliability Standards remains work in progress.

ENTSO-e occupies an intermediate position: its December 2025 position paper articulates the right requirements at the level of principle, but the absence of a binding EU-level instrument means that implementation will be fragmented across member states. Ireland, as the system under greatest stress, is moving fastest; other systems with lower current data centre penetration have less urgency and may wait for the next adverse event before acting.

Of the three, AEMO operates the most technically complete framework for managing power-electronic resources — but that maturity was built for renewable generation, not consumption loads; adapting it to data centers is a near-term task rather than a foundational redesign. NERC has produced the most analytically rigorous risk taxonomy, but its advisory rather than prescriptive role means translation into binding standards remains work in progress. ENTSO-e articulates the right principles, but the absence of a binding EU instrument risks fragmented national implementation, with Ireland — under greatest stress — moving fastest and others waiting for the next adverse event.

The pattern across all three is consistent: incidents precede policy, and the scale of the incident determines the speed of response. **The regulatory framework takes three to five years to build from first recognition of the risk to binding requirements. Given current growth trajectories, that lead time must be initiated now, not after the first incident occurs.**

3. Economic Affordability: Who Pays for the System Expansion?

There is a structural asymmetry at the heart of the large load challenge that no amount of technical ingenuity can dissolve. Data centers generate the problem; regulation must provide the solution. This is not a polemical observation — it is a precise description of how the economics of system expansion work. When a hyperscaler requests hundreds of megawatts of firm grid capacity at a single interconnection point, the transmission infrastructure required to serve that request at peak demand under all system conditions must be built, financed, and recovered in tariffs. If the regulatory framework does not assign that cost to the entity that caused it, the cost does not disappear — it is redistributed to all other ratepayers, most of whom are households and small businesses with no share in the economic benefits of AI development. This is the cross-subsidy risk, and addressing it is the central affordability challenge of our time for energy regulators.

For generation, things are even more complex. When PJM Interconnection held its capacity auction in July 2024 for the 2025/2026 delivery year, clearing prices reached $269.92 per megawatt-day — roughly ten times the level of the two previous auctions.[29] The proximate cause was a sharp upward revision in demand forecasts driven by data center growth, compounded by the retirement of thermal generation. The additional cost to consumers in the PJM region was estimated at approximately $14.7 billion annually.[30] That figure represents a transfer — from the ratepayer base to the owners of capacity assets — that

[29] PJM Interconnection, 2025/2026 RPM Base Residual Auction Results, July 2024. The clearing price of $269.92/MW-day in the RTO zone compared with $28.92/MW-day in the 2024/2025 delivery year.
[30] The cost impact figure is derived from the auction clearing prices applied to PJM's total capacity procured. Various industry analyses place the annual incremental cost to consumers in the range of $13–$15 billion. See Grid Strategies, *National Load Growth*, 2025, for discussion of the forecast drivers behind the auction outcome.

was triggered in significant part by demand growth that may itself have been overstated, as discussed in Chapter 1. The affordability dimension of large loads is, in this sense, inseparable from the predictability dimension: overforecasting demand not only risks stranding physical infrastructure, but can immediately and directly raise energy costs for millions of consumers through wholesale market price formation.

3.1 The Cost Structure of Grid Expansion

To understand why cost allocation is so contested, it is necessary to understand how grid expansion costs are structured. Under traditional regulatory practice in most jurisdictions — both in North America and in Europe — transmission network costs are socialised: they are recovered through tariffs paid by all network users on the basis of consumption or connected capacity, broadly independent of which specific load caused a given investment.

This socialisation principle rests on the public-good character of transmission infrastructure and on the difficulty of attributing shared network benefits to individual users. It works tolerably well when network expansions serve diffuse, incremental demand growth spread across many customers.

It breaks down when a single load, or a small cluster of loads, triggers a disproportionate share of new investment. In that case, socialisation of costs is not cost-sharing — it is cost transfer. The distinction matters because it alters incentive structures: if a large load knows that the cost of grid reinforcement will be spread across all ratepayers, it has no incentive to locate in areas of existing surplus capacity, to accept non-firm connection arrangements, or to invest in on-site generation that would reduce its net draw on the grid. The socialisation principle thus creates a form of moral hazard precisely among the customers whose connection decisions have the largest systemic impact.

NERC has documented this dynamic in its July 2025 characterisation of large loads, noting that most large loads interconnecting to the grid request firm utility service — meaning the utility must provide transmission infrastructure capable of serving the load's peak demand request at all times — regardless of how often that peak is actually reached or whether the load could technically tolerate occasional curtailment.[31] The implication is direct: the grid is sized for declared peaks, and declared peaks are not always real peaks, as discussed in Chapter 1. The combination of overforecasting and firm connection requests produces a systematic bias toward infrastructure overbuilding that is paid for by all ratepayers, not only by those who caused it.

3.2 The "Causer-Pays" Principle and the American Laboratory

The appropriate regulatory response to this asymmetry is the *causer-pays* principle: the entity whose connection decision triggers a network investment should bear the incremental cost of that investment. This principle is not new — it underpins the shallow/deep cost allocation debate in generation interconnection that has occupied FERC for two decades — but its application to demand-side large loads has until recently been largely unexplored, because demand of this scale and concentration was simply unprecedented.

[31] NERC *Emerging Large Loads*, 2025, Chapter 2 ("Firm Load vs. Flexible Load"), p. 14.

The United States has in the past three years become an unintentional laboratory for causer-pays mechanisms on the demand side, as individual states and utilities have responded to the data center buildout with a variety of tariff innovations. The results are instructive for European regulators, both in what has worked and in the fragmentation that has resulted from the absence of a harmonised framework.

The common architecture of what has come to be called *large load tariffs* includes several elements. A size threshold — typically between 20 and 100 MW — triggers applicability, above which a distinct regulatory regime applies. The load is required to contract for a defined minimum billing obligation, typically expressed as a percentage of contracted capacity (commonly 80 to 100 percent) that is payable even in periods of reduced consumption; this ensures that the infrastructure built to serve the load generates a minimum revenue stream regardless of utilization. The load assumes direct responsibility for the cost of network upgrades required to accommodate its connection, documented in an interconnection agreement. Long-term contract commitments — ranging from five to fifteen years — with exit fees or stranded cost provisions are standard, and some jurisdictions require creditworthiness demonstrations or collateral posting.[32]

The proliferation of these tariffs within a single nation illustrates both the legitimacy of the underlying concern and the cost of regulatory fragmentation. Each jurisdiction has designed its own threshold, its own minimum bill percentage, its own contract term, and its own cost responsibility methodology. The result is a patchwork of rules that large loads must navigate when selecting interconnection locations — and that creates opportunities for regulatory arbitrage, as operators choose jurisdictions with more lenient cost-recovery requirements over those with stricter ones. This is not a hypothetical risk: there is already evidence that the attractiveness of specific US markets for data center investment correlates with the relative leniency of connection and cost-allocation rules, as well as with energy prices.[33]

The deeper tension in the American approach is between two legitimate goals that do not always point in the same direction. Causer-pays principles correctly protect existing ratepayers from bearing the cost of infrastructure built for private large-scale commercial interests. But tariff designs that are too stringent, or that create excessive contractual rigidity, may deter investment entirely — or push it to jurisdictions with fewer safeguards, producing both economic and regulatory competition that serves no one's long-term interest. The right calibration is a regulatory judgment, not an economic formula.

Spain's RDL 7/2026 operationalises a comparable logic at an earlier stage of the process: permit holders are required to pay a monthly capacity reservation fee from the date of grant — before any infrastructure is built or any power consumed — calculated as a multiple of

[32] The common elements described are drawn from Grid Strategies, *National Load Growth*, 2025, pp. 19–21, which surveys six jurisdictions that have adopted or proposed large load tariff mechanisms. Indiana Michigan Power adopted a large load tariff applicable to loads of 70 MW or more, requiring an 80 percent minimum bill and a twelve-year contract commitment. AEP Ohio operates a dedicated data center tariff for loads exceeding 25 MW with minimum billing up to 85 percent of contracted capacity. Texas Senate Bill 6 (2023) directed the Public Utility Commission of Texas to ensure that large loads above 75 MW make an appropriate contribution to grid costs, currently being implemented through PUCT rulemaking. Oregon's House Bill 3546, the POWER Act, establishes a ten-year minimum contract requirement for loads above 20 MW, with implementation delegated to the Oregon Public Utility Commission.

[33] See Nicholas Institute, Duke University *Rethinking Load Growth* 2025, for a comparative analysis of the factors influencing large load siting decisions.

existing network tariffs and increasing over time. The instrument directly addresses the speculative queue occupation problem identified in Chapter 1: approximately 90 percent of granted distribution access capacity in Spain was reported as unused at the time of the decree's adoption.[34]

3.3 Non-Firm Connection and the Logic of Demand-Side Flexibility

A parallel and complementary strand of regulatory innovation addresses the affordability problem from a different angle: not by assigning the cost of firm infrastructure to the large load, but by questioning whether firm infrastructure is in fact necessary to serve the load's actual operational needs. The key concept is the *non-firm* or *flexible connection*. Under a firm connection, the grid operator guarantees delivery of the contracted power at all times and under all system conditions — hence the obligation to build infrastructure capable of serving the declared peak demand in the worst-case scenario. Under a non-firm or flexible connection, the load accepts a defined probability or frequency of curtailment during periods of high system stress, in exchange for which the grid operator dimensions the required infrastructure to a lower level. The physical infrastructure needed is determined by the load's realistic operational profile — including the hours per year it is willing to operate on curtailed supply — rather than by its declared maximum capacity.[35]

This is not a novel concept in electricity regulation: interruptible supply tariffs have existed in many European jurisdictions for decades, originally designed for industrial loads capable of temporary shutdown. What is new is its systematic application to very large loads at transmission voltage level, and its quantification as a network planning tool rather than merely as an emergency demand response mechanism. NERC has noted that concepts around mandatory load curtailments by ISOs or utilities during stressful grid conditions have been explored, which might allow transmission planners to assume a lower peak demand for the load and potentially reduce the transmission buildout exclusively needed to serve it.[36] The practical challenge is that most large loads — particularly data centers providing cloud services and AI inference — have contractual uptime commitments of 99.999 percent or higher to their customers, making any form of guaranteed curtailment operationally and commercially problematic. The industry consultation conducted by Ireland's CRU confirmed this reluctance: industry respondents emphasised that data centers require certainty of power availability in order to meet service level agreements, and that non-firm connections would not be viable for many types of large energy users.[37]

The resolution of this apparent impasse lies in recognising that there are different layers of non-firmness. A data center need not accept curtailment of its *computational operations* in order to accept non-firm *grid import* during specific system stress events, provided it has on-site resources capable of bridging the gap for the limited hours during which curtailment would be called. This is precisely the architecture that underpins the most sophisticated current proposals: the large load accepts conditional grid curtailment, but maintains continuous operations through on-site generation or storage that activates

[34] Real Decreto-ley 7/2026, de 20 de marzo, *por el que se aprueba el Plan Integral de Respuesta a la Crisis en Oriente Medio,* BOE-A-2026-6544, 21 marzo 2026, Arts. 11–13 and Disposiciones Transitorias 3ª–5ª. Convalidated by the Spanish Congress on 26 March 2026. Art.11; the fee is structured as an advance payment on future network tariffs, recoverable once the installation enters operation.

[35] The concept is explained in NERC *Emerging Large Loads*, 2025, p. 14, citing Nicholas Institute, Duke University *Rethinking Load Growth* 2025.

[36] NERC *Emerging Large Loads*, 2025, p. 14.

[37] CRU *Large energy users proposal*, 2025, Section 8.5 ("Non-firm demand connections"), p. 96.

during the curtailment window. From the system operator's perspective, the load is non-firm; from the data center operator's perspective, operations are uninterrupted.

3.4 Bring Your Own Generation, Bring Your Own Capacity

Two specific instruments have emerged to operationalise this architecture, and it is important to distinguish them clearly because they operate at different levels of the system and have different regulatory implications.

Bring Your Own Generation (BYOG) requires the large load to install dispatchable generation capacity on-site or in proximity, sufficient — after applicable derating — to match its maximum import capacity from the grid. This is the approach adopted by Ireland's Commission for Regulation of Utilities (CRU) in its February 2025 proposed decision on large energy user connections, the most detailed and advanced large load connection policy in Europe to date.[38] Under the CRU's proposed framework, every new data center connecting to the Irish electricity network must procure dispatchable generation equal to its Maximum Import Capacity (MIC), with that generation required to participate in the wholesale Single Electricity Market (SEM) rather than being reserved exclusively for the data center's own backup use. This is a crucial design choice: the generation is not a private backup asset — it is a market participant that contributes to system adequacy, receiving capacity market remuneration that partially offsets the data center's capital outlay, while granting the data center an exemption from Mandatory Demand Curtailment provisions that apply to other large energy users.

The logic is elegant. Instead of requiring the grid to build the generation capacity needed to serve a new large load, the CRU requires the large load itself to bring that capacity into the system. The net effect on system adequacy is equivalent to the traditional approach; the effect on cost allocation is radically different. The cost of the generation asset is internalised by the data center rather than socialised across ratepayers. And because the generation asset participates in the wholesale market, it generates revenue that makes the arrangement commercially viable over the asset's lifetime, rather than representing a pure capital burden.

Bring Your Own Capacity (BYOC) is a market-oriented variant, developed primarily in the United States in the context of capacity markets such as PJM, in which the large load directly procures capacity credits from the wholesale capacity market — or contracts bilaterally with generation owners — rather than installing physical generation on-site. The operational effect on the capacity market is similar to BYOG: the large load's demand does not place a net additional burden on the capacity procurement process, because it arrives self-supplied.

The distinction from BYOG is that BYOC does not require the large load to own or operate physical generation; it requires only that it secure the financial capacity commitments that the market would otherwise need to procure on its behalf. This is more flexible than BYOG and better suited to jurisdictions with liquid capacity markets, but it requires careful regulatory design to ensure that the self-supply genuinely corresponds to

[38] CRU *Large energy users proposal*, 2025. The proposed decision requires dispatchable onsite or proximate generation and/or storage matching the data center's Maximum Import Capacity (subject to derating factors), with mandatory participation in the Single Electricity Market (SEM).

incremental market capacity rather than simply displacing existing commitments.[39] The quantitative case for BYOC combined with flexible connection arrangements has been made with unusual rigour by a December 2025 study commissioned by Google from Camus Energy, encoord, and Princeton University's ZERO Lab. Analysing six candidate sites in the PJM territory for a hypothetical 500 MW data center, the study found that a flexible connection with 20 percent conditional firm capacity — meaning the data center accepts curtailment for up to 20 percent of its contracted capacity during system stress events — combined with BYOC, allows the facility to reach full operations within approximately two years, three to five years faster than a traditional firm interconnection would permit.[40] During that accelerated operational period, on-site resources would need to be deployed for only 40 to 70 hours per year — less than 0.8 percent of annual hours — to bridge curtailment windows. Crucially, the cost coverage analysis shows that for every GW of new demand served through this model, the data center directly internalises approximately 96 percent of the incremental system costs it generates, compared to the near-zero internalisation of a traditional firm connection with socialised upgrade costs.[41] The residual 4 percent represents costs that remain with the broader system — a share that, while not zero, is vastly smaller than the current default.

The Camus/Princeton study was commissioned by Google and should be read with that context in mind. However, the methodology is transparent and the findings are consistent with the directional logic of the non-firm/BYOC framework: reducing the firm capacity obligation of a large load reduces the infrastructure required to serve it, and the on-site resources needed to substitute for that firm capacity under real-world operating conditions are far smaller than the peak capacity they nominally replace, because the overlap between grid stress events and data center peak demand is limited.

3.5 Non-Discrimination as a Design Constraint

Designing large load tariffs and BYOG/BYOC requirements raises an immediate question for European regulators: do these instruments comply with the principle of non-discriminatory access to the network, which is a foundational element of the EU's internal energy market framework? [42]

The concern is real but ultimately manageable, provided the regulatory architecture is carefully constructed. The risk of non-compliance arises if cost-allocation or connection requirements are designed in ways that single out specific types of users — for instance, if

[39] The BYOC concept as applied in the PJM context is discussed in Camus Energy, encoord, and Princeton University ZERO Lab, *Rethinking Grid Connections: How Flexible Data Centers Can Accelerate Grid Access and Control Costs*, December 2025 (commissioned by Google LLC), pp. 4–6 https://zero.lab.princeton.edu/wp-content/uploads/2025/12/Flexible-data-centers-report-dec-2025.pdf (hereinafter: Camus Energy et al., *Rethinking Grid Connections*, 2025).

[40] Camus Energy et al., *Rethinking Grid Connections*, 2025, p. 5. The two-year timeline assumes parallel procurement of BYOC and deployment of on-site resources, compared with a five-to-seven-year typical timeline for traditional firm interconnection in constrained PJM zones.

[41] Camus Energy et al., *Rethinking Grid Connections*, 2025, p. 6. The 96 percent cost coverage figure is the sum of BYOC capacity costs ($326 million/GW), energy payments ($329 million/GW), and the avoided new-build savings from the flexible connection design ($78 million/GW), expressed as a fraction of the $764 million/GW incremental system cost of a traditionally served GW of new demand.

[42] For a wider comparative discussion of regulatory approaches in EU and US, see Lo Schiavo, L., *Regulatory and Policy Frameworks for AI Data Centers in EU and US*, ERRA Working Paper, August 2025, https://erranet.org/regulatory-and-policy-frameworks-for-ai-data-centers-in-eu-and-us/ (hereinafter: ERRA working paper, *Regulatory and Policy Frameworks for AI Data Centers in EU and US* 2025).

BYOG is required only of data centers but not of other large loads with comparable system impact, or if connection conditions vary across users with identical technical profiles for reasons unrelated to their actual grid impact.

The resolution lies in ensuring that special requirements are grounded in objective technical criteria applicable to all users meeting those criteria, regardless of their sector or identity. The CRU's Irish framework illustrates this approach: the proposed BYOG requirement applies to all data centers above a threshold Maximum Import Capacity, defined by objective size and connection characteristics rather than by sectoral identity. The CRU's justification for applying the policy only to data centers — and not, for instance, to manufacturing facilities — rests on documented characteristics specific to this class of user: the speed and scale of growth, the high load factor (near-constant consumption at or near maximum capacity), and the geographic and temporal concentration of new connection requests.[43] These are technical differentiators that a regulator can defend before a court or a competition authority; they are not preferences or industrial policy dressed up as regulation.

Spain's *concurso de demanda*, introduced in July 2025 implementing Royal Decree-law 8/2023, offers a complementary illustration from a different angle.[44] Rather than imposing differentiated technical requirements on specific user categories, Spain replaced the traditional first-come-first-served capacity allocation with a competitive scoring mechanism that ranks connection applications by their contribution to decarbonisation objectives. The mechanism is formally non-discriminatory — it is open to all categories of demand — but it produces differentiated outcomes based on objective, publicly disclosed criteria aligned with EU energy policy goals. A data center powered entirely by renewable energy scores higher than one relying on grid-average power; a flexible load scores higher than a firm-demand user of equivalent size. This is not special treatment for large loads: it is the application of consistent principles that happen to reward characteristics that data centers can acquire through investment and commitment.

Both approaches — differentiated technical requirements justified by objective criteria, and competitive allocation based on transparent policy-relevant scoring — are compatible with the non-discrimination principle as understood in EU law, provided the criteria are predetermined, publicly available, and applied consistently. What would not be compatible is the ad-hoc granting of preferential connection terms to specific investors, the exclusion of specific user categories from connection queues, or the exemption of favoured projects from cost-allocation obligations that apply to others. The non-discrimination principle does

[43] CRU *Large energy users proposal*, 2025, Section 3.1 ("Scope of Application"), pp. 47–54.

[44] *Resolución de 11 de julio de 2025, de la Secretaría de Estado de Energía por la que se convocan los concursos de capacidad de acceso de demanda en determinados nudos de la red de transporte,* Boletín Oficial del Estado, 17 July 2025 (BOE n.171-2025-14863) https://www.boe.es/boe/dias/2025/07/17/pdfs/BOE-A-2025-14863.pdf (hereinafter: Resolución de 17 de julio de 2025, *BOE* núm. 171. N.14863) The mechanism applies to demand facilities at transmission network nodes and ranks applications by greenhouse gas avoidance criteria, replacing chronological priority. Further details are provided in ERRA working paper, *Regulatory and Policy Frameworks for AI Data Centers in EU and US* 2025, with comparison to other policy frameworks like German law on energy efficiency (Energieeffizienzgesetz – EnEfG, https://www.gesetze-im-internet.de/enefg/BJNR1350B0023.html) that sets requirements for data centers only in terms of PUE (Power Usage Effectiveness).

not prevent regulators from designing sophisticated, differentiated frameworks; it prevents them from designing arbitrary or opaque ones.

3.6 The Australian Model: Economic Regulation as a Proactive Shield

The Australian National Electricity Market (NEM) offers a particularly instructive case for the affordability dimension, not least because it demonstrates how a well-structured institutional architecture — with a clear separation between system operation and economic regulation — can enable proactive responses to cost allocation challenges before they become political crises.

Australia's energy governance rests on a deliberate tripartite design. The Australian Energy Market Commission (AEMC) sets the rules; the Australian Energy Market Operator (AEMO) manages the system and markets in real time and leads long-term planning; and the Australian Energy Regulator (AER) enforces those rules, sets maximum revenues for network businesses, and oversees retail affordability. Each is an independent decision-maker with distinct statutory functions and accountabilities. The separation matters: it means that AEMO's system security planning (discussed in Chapter 2) and the AER's cost allocation decisions can proceed simultaneously, on their respective technical merits, without either function contaminating the other.

The AER has engaged the data center question directly, and in a way that illuminates the broader cost allocation problem articulated earlier in this chapter. Network costs represent close to half of Australian household electricity bills, and those costs are rising. As Energy Consumers Australia has noted, the rapid growth of data centers has driven increases in peak electricity demand, requiring network upgrades — and the central policy question is who should bear those costs.[45] The AER's answer has been unambiguous: the cost-causation principle applies. Connection costs triggered by a specific load must be borne by that load, not socialized across the existing customer base.

This principle has been applied with notable rigour. In its 2025-2030 revenue determination for Jemena, the Melbourne distribution network, the AER rejected 38% of forecasted capex, finding that the projected doubling of demand driven by data centre connections had not been adequately justified.[46] The decision is significant not only for its quantum but for its logic: it applies the same scrutiny to utility demand forecasts — and their attendant capital expenditure claims, that Chapter 1 identifies as structurally absent in the US context.

The Ausgrid case in New South Wales illustrates the operational implementation. When four data centres sought to connect to Ausgrid's Macquarie Park substation — which lacked spare capacity — the network required the developers to fund a new substation entirely. The data centers are charged based on the maximum capacity they have reserved, not merely the capacity they consume during ramp-up, and must meet predefined "capacity commitment milestones" — such as purchasing land or procuring transformers — to

---

[45] Energy Consumers Australia, *How Data Centres are Reshaping Australia's Energy Landscape*, 15 October 2025 https://energyconsumersaustralia.com.au/news/how-data-centres-are-reshaping-australias-energy-landscape.

[46] Australian Energy Regulator, Draft Decision — Jemena Electricity Networks Distribution Determination 2026–31, 30 September 2025. Jemena's proposed forecast capex was AUD 1,366.3 million ($2025–26); the AER's alternative estimate was AUD 843.3 million, a reduction of 38%. The total draft revenue allowed was AUD 1,600.1 million ($2025–26, smoothed), to be finalised by 30 April 2026. Available at: https://www.aer.gov.au/industry/registers/determinations/jemena-determination-2026-31 .

maintain their position in the connection queue.[47] This structure simultaneously solves two problems identified in Chapter 1: it aligns financial exposure with actual demand commitments, reducing the incentive for speculative queue occupation; and it ensures that the infrastructure investment risk is borne by the party that creates the demand rather than by existing ratepayers.

There is a further dimension worth noting. The AER's approach draws an explicit distinction between data centres and other forms of growing electricity demand, such as electric vehicle charging. EV charging can improve network utilisation by spreading demand across the day, potentially reducing the need for infrastructure upgrades; data centres, running at constant high load, are more likely to drive peak demand and require dedicated capacity.[48] This distinction — between load types that are complementary to the existing system and load types that are structurally additive — is precisely the kind of technically grounded differentiation that economic regulation, insulated from political pressure to treat all investment equally, is well positioned to make and enforce.

The Australian case thus completes a picture that earlier sections of this chapter assembled from the US and European experiences. Where the US has struggled with cost allocation partly because regulatory incentive structures reward capital expenditure and partly because political pressure now increasingly overrides tariff design, and where European frameworks vary considerably in their approach to large load cost recovery, the AER demonstrates that a well-resourced, independent economic regulator with a clear consumer protection mandate can address the affordability challenge proactively — through scrutiny of utility revenue claims, cost-reflective connection charging, and firm enforcement of the cost-causation principle — before data centre growth transforms from an infrastructure planning question into a household bill crisis.

3.7 Regulatory Implications: Three Models and a Common Lesson

The three regulatory experiences examined in this chapter do not converge on a single template. They were designed in different institutional contexts, for different market structures, and under different legal constraints. But they share a common diagnostic and point toward a common set of principles.

The American experience demonstrates what happens in the absence of a coherent framework: US jurisdictions have improvised individually, producing a patchwork that creates regulatory arbitrage and imposes navigation costs on investors and regulators alike. The underlying logic of causer-pays is sound; its fragmented implementation is not. The PJM capacity auction outcome — $14.7 billion in additional annual consumer costs, triggered in part by overstated demand forecasts — is the most quantified illustration of what poorly structured cost allocation produces at scale. And when the political cost became visible, it took a presidential intervention rather than a regulatory decision to resolve it. [49]

[47] AER, Ausgrid — Increased Customer Demand Requirements in the Macquarie Park Area — Contingent Project Application, February 2025, which documents the AED 162 million substation project and the cost-recovery mechanism https://www.aer.gov.au/industry/networks/contingent-projects/ausgrid-increased-customer-demand-requirements-macquarie-park-area-contingent-project .

[48] Energy Consumers Australia, op. cit.

[49] This has been the case of last evolution of PJM price crisis, see Matthew Zeithlin, "*How Trump Made an Electricity Price Deal With Democrats"*, Heatmap News, 16 January 2026

The European experiments — Ireland's BYOG requirement, Spain's *concurso de demanda* — **demonstrate that frameworks can be simultaneously ambitious on cost allocation, compatible with the non-discrimination principle, and grounded in objective technical criteria**. These are exportable models, not local solutions. The Australian case adds the institutional lesson the other two leave implicit: what distinguishes the AER's approach is not a superior legal framework but an independent regulator with a clear consumer protection mandate and the institutional will to apply the cost-causation principle before pressure builds rather than after. That is a governance lesson, not a tariff design lesson.

For European regulators, the risk of competitive fragmentation is real, but the remedy is not harmonisation for its own sake. It is harmonisation of the foundational principles all three models share: large loads must not be cross-subsidised by existing ratepayers; connection requirements must rest on objective technical criteria; and non-discrimination permits — indeed requires — differentiated treatment where differentiated technical circumstances genuinely justify it.

The affordability dimension is ultimately a question of distribution: who bears the cost of serving the digital economy, and on what basis? Regulation cannot determine who wins or loses in the AI market. But it can and must ensure that the costs of powering that market are assigned to those who generate them. Where independent regulators have been given the mandate and the will to enforce this principle, they have done so effectively. Where they have not, the cost has been borne by the ratepayer.

4. Environmental Sustainability: When Regulation Crosses Institutional Boundaries

On 13 March 2026, residents of Stockton, Alabama — a community of fewer than a thousand people — learned that a utility-scale solar farm was to be built on land adjacent to their homes to power a Meta data center in the state capital. The local county had no authority to regulate land use in unincorporated areas. The reaction escalated to the state legislature, where a senator introduced a blanket statewide moratorium on all new utility-scale solar projects. A data center operator doing exactly what decarbonisation policy encourages — procuring renewable energy — triggered a legislative instrument that would have blocked the infrastructure that decarbonisation requires.[50]

The story is not exceptional. It is the sustainability problem of large AI loads in miniature: a challenge that simultaneously implicates energy markets, climate governance, land use

https://heatmap.news/energy/pjm-auction-trump-shapiro and Lo Schiavo, *AI Data Centers and Energy Regulation,* 2026, https://www.euppublishing.com/doi/10.3366/gels.2025.0141.

[50] The Alabama solar moratorium bill (Senate Bill 354) was introduced by State Senator Greg Albritton following community opposition to a Meta-linked solar project in Baldwin County, south of the state capital of Montgomery. The episode is documented J. Holzman, "*Alabama Is Threatening to Ban Utility-Scale Solar*", *Heatmap* March 13, 2026. It illustrates a structural pattern visible across multiple jurisdictions: in the absence of a preemptive planning framework at the relevant institutional level, local opposition escalates to blunt legislative instruments — with consequences that extend well beyond the original dispute.

planning, and water management — each governed by different institutions at different levels, none of which alone has the tools to address it.

4.1 The Four-Dimensional Challenge

The sustainability problem of energy demand growth due to data centers has four analytically distinct dimensions. The first is *additionality*: whether data center renewable procurement drives new clean generation or simply re-routes existing certificates, leaving the marginal carbon intensity of the grid unchanged. The second is *temporal matching*: whether renewable energy is consumed hour by hour in correspondence with actual data center load, or whether annual accounting on paper conceals periods of high-carbon consumption. The third is *social acceptance*: whether the physical infrastructure for renewable energy can actually be built where it is needed, at the pace demand requires. The fourth involves natural resources usage, like land use and water consumption.

Underlying all four dimensions is a more fundamental question that regulatory frameworks have not yet systematically addressed: whether the absolute growth of AI electricity demand is itself a governance variable, or whether it is simply a parameter to which regulation must adapt. The energetic burden of AI systems has already structurally shifted from one-time model training to recurring, potentially unbounded inference — with inference accounting for over 90% of total lifecycle energy consumption in high-traffic deployments, and energy consumption per query increasing by a factor of approximately 80x for advanced reasoning models relative to standard large language models.[51]

Each dimension requires a regulatory response, and the responses interact. Annual certificate matching is cheap and produces no additionality. Hourly matching is technically demanding but increasingly feasible and is the standard toward which the most advanced operators — Google's 24/7 Carbon-Free Energy programme, the EED reporting framework under Delegated Regulation 2024/1364 — are moving.[52] Additionality without temporal matching is incomplete; temporal matching without locational planning produces the Alabama outcome. Whether the fifth — absolute demand growth itself — can or should become a governance variable is a question this article notes but does not resolve, as its

---

[51] Belova et al., *An Alternative Trajectory for Generative AI*, Princeton University, arXiv:2603.14147 (2026), §10. The analysis identifies a Jevons paradox dynamic: efficiency gains from smaller, more optimised models are likely to be absorbed by the expansion of new applications, such that efficiency alone does not guarantee reduced total consumption and must be paired with deliberate constraints or carbon pricing. The competition and industrial policy dimensions of the same problem have attracted growing regulatory attention in parallel. In December 2025, the French *Autorité de la concurrence* published a study identifying the emergence of "AI frugality" — the deliberate reduction of model size and energy intensity — as a competitive parameter capable of reshaping market structure, and noting that without reliable transparency on environmental footprint, a handful of large operators risk securing a decisive informational advantage over competitors. *See* Autorité de la concurrence, *Les questions concurrentielles relatives à l'impact énergétique et environnemental de l'Intelligence Artificiell*, Study of 17 December 2025, https://www.autoritedelaconcurrence.fr/en/press-release/autorite-publishes-its-study-competition-issues-surrounding-energy-and-environmental

[52] The distinction between annual and hourly accounting is increasingly operationalized in regulatory frameworks. EU Delegated Regulation 2024/1364, implementing Article 12 of the revised Energy Efficiency Directive (2023/1791), requires data center operators above 500 kW to report renewable energy use and PUE from September 2024 — creating the data infrastructure for more granular matching requirements in future.

focus lies in regulatory framework design rather than the technological trajectory that generates the demand.

4.2 Additionality: The Irish Benchmark

The most rigorous regulatory attempt to embed additionality directly into connection policy is the CRU's December 2025 large energy user decision. All new data centers above the 1 MVA threshold must meet at least 80 percent of annual demand with additional indigenous renewable electricity — new or fully repowered projects, contracted through PPAs not covered by state support schemes, generating within the Irish system.[53] A six-year glide path from energisation acknowledges realistic renewable development timelines.

The emphasis on indigenous generation is deliberate: certificates from Norwegian or Spanish renewable projects do not increase the amount of clean energy in the Irish grid. The Spanish *concurso de demanda* takes a complementary approach, embedding greenhouse gas emissions avoided as an explicit scoring criterion in the competitive access allocation process — making sustainability a selection variable rather than a reporting obligation.[54]

Both models share a common logic: sustainability requirements must be embedded upstream, at the point of grid access, rather than left to voluntary corporate commitments whose verification is structurally difficult and whose additionality is structurally uncertain.

4.3 The Gas Paradox and Its Resolution

The CRU's requirement that data centers install on-site dispatchable generation as a condition of connection creates an apparent sustainability paradox. Mandating gas-fired capacity for the sector that should lead the clean energy transition seems counterproductive.

The CRU's response is analytically sound: the system-level effect depends on how the asset operates, not on its fuel type. In a system targeting 80 percent renewables by 2030, dispatchable gas capacity runs as a balancing resource — for a limited number of hours per

[53] The CRU final decision is: Large Energy Users Connection Policy, Decision paper CRU2025236, December 2025 https://cruie-live-96ca64acab2247eca8a850a7e54b-5b34f62.divio-media.com/documents/CRU2025236_Large_Energy_User_connection_policy_decision_paper.pdf (hereinafter. The February 2025 proposed decision had attempted to decarbonise the on-site generation asset directly, requiring gas generators to be *"sufficiently futureproofed in terms of facilitating low/zero emissions going forward, e.g. if using gas generation, the equipment should be compatible with being run on hydrogen or biofuels."* The final decision substantially softened this, encouraging futureproofing *"where possible and appropriate"* and broadening the fuel reference to *"lower or zero-carbon fuels"* more generally. The shift is analytically sound: the proposed approach was commercially difficult and dependent on hydrogen timelines that remain speculative at scale. The final decision concentrates the decarbonisation obligation on the energy supply side, where the 80% renewable requirement is both enforceable and technology-neutral.

[54] Resolución de 17 de julio de 2025, *BOE* núm. 171. N.14863. The three scoring criteria are: (a) date of consumption commencement, (b) volume of investment, (c) greenhouse gas emissions avoided. The third criterion is the most significant regulatory innovation: it converts sustainability from a post-connection reporting obligation into an ex ante competitive selection variable, creating a direct financial incentive to bring demonstrable decarbonisation commitments to the connection application itself.

year, during periods of low wind or solar output.[55] The paradox dissolves once the distinction between *energy supply* — addressed by the renewables requirement — and *system services* — provided by the dispatchable asset — is maintained.

4.4 Land Use: The Governance Difference

The Alabama outcome and the Lombardy response to the same underlying pressure illustrate how differently the land use challenge can be governed. In Lombardy — which hosts 60%of Italian data center capacity currently connected (317 MW, out of total 513 MW at nation-level), with connection requests growing from 1 GW in 2021 to 23 GW by May 2025 — local opposition to a proposed greenfield data center in the Milano metropolitan area (site of Zibido San Giacomo) produced not a moratorium but a regional legislative proposal for structured siting criteria: priority for brownfield and already-urbanised areas, land consumption limits, a permanent regional observatory for monitoring and community participation.[56]

The difference is not cultural but regulatory. Lombardy has a regional planning framework that can be engaged and refined. Alabama does not have a state solar siting policy against which to evaluate competing interests. Preemptive planning frameworks — even imperfect ones — are vastly preferable to reactive responses. The data center boom is arriving fast enough that jurisdictions without them will face the Alabama dilemma soon.

4.5 Water and Waste Heat: The Nexus

Water is the sustainability footprint that has largely escaped regulatory attention while generating its own version of the Alabama problem in water-stressed regions. A 100 MW facility using conventional evaporative cooling can consume in excess of one billion litres per year. EU Delegated Regulation 2024/1364 mandates WUE (Water Usage Effectiveness) disclosure for facilities above 500 kW from September 2024 — but this is transparency without performance standards, with no regional differentiation for water stress contexts where the same WUE figure has entirely different implications in northern Finland and in Aragon, Spain.[57]

---

[55] At 10 percent annual utilisation — consistent with a high-renewables system where dispatchable generation operates as a last-resort balancing resource — 3.7 TWh of biomethane could support over 2 GW of such capacity. The CRU's calculation (CRU2025236, technical annex) illustrates how the system-level carbon impact of BYOG depends not on the existence of gas capacity but on the dispatch regime within which it operates.

[56] The Zibido San Giacomo case was documented in local and regional Italian media in February 2026. The mayor's statement explicitly distinguished between the rejected greenfield proposal and an approved Amazon facility on already-industrialised land — a distinction that reflects the presence of a functioning land use instrument (the municipality had reduced land consumption by 50 percent in its revised *piano regolatore*) rather than opposition to data center investment as such. The Lombardy regional legislative proposal, introduced by councillor Matteo Piloni (www.milanotoday.it/politica/governo-data-center-legge-regionale.html), represents the institutional escalation from case-by-case resistance to structured governance — the appropriate regulatory response to a challenge of this scale and pace.

[57] The European water governance gap can be shortly illustrated by case-by-case national responses that substitute for the missing EU framework. In the Netherlands, Microsoft's Middenmeer data center consumed 84 million litres of drinking water in 2021 — four to seven times the volume committed in the permitting process — prompting Vitens, the local water utility, to place 45 industrial applicants on waiting lists. In southern Europe, Spain's river basin authority flagged Meta's proposed Talavera de la Reina complex as requiring 665 million litres annually from an already drought-stressed basin; Catalonia imposed emergency bans on permits above 1 MW during drought conditions. These are reactive, case-by-

The energy-water nexus cannot be governed by optimising either metric in isolation: switching from evaporative to air cooling reduces WUE to near zero but typically increases PUE by 15 to 25 percent, raising electricity consumption and emissions. Nordic jurisdictions demonstrate what becomes possible when the two are governed together — Google's Hamina facility achieves PUE of 1.09 with zero potable water consumption through seawater cooling; Denmark's abolition of the surplus heat tax transformed the economics of waste heat recovery, enabling Meta's Odense campus to supply heat to 12,000 homes and eliminating cooling tower water consumption simultaneously.[58]

Waste heat recovery — currently treated in most European jurisdictions as an externality to be managed rather than a resource to be captured — is the clearest example of how regulatory design determines whether a physical phenomenon is a cost or a benefit. Italy's first significant implementation in another site of Milano metropolitan area (Rozzano), where TIM's data center was connected to Getec's district heating network supplying 5,000 social housing units with an estimated saving of 3,500 tonnes of CO_2 annually, demonstrates that the framework can be built and the benefits are concrete.[59] The question for regulators is whether to wait for bilateral arrangements as happened in Rozzano, or to create systematic incentives — as Denmark has done through its waste heat legislation — that make recovery the default rather than the exception.

4.6 Regulatory Implication: The Boundary of the Energy Regulator's Mandate

The sustainability dimension of the large load challenge is where the limits of energy regulation as conventionally defined become most visible. Additionality requirements, temporal matching standards, and dispatchable generation conditions fall within the energy regulator's toolkit. Land use planning, water allocation, and district heating integration do not — or not alone. The CRU's explicit acknowledgment that the medium-term governance of data centers requires "a State-led approach encompassing spatial planning, targets-based and plan-led infrastructure development, which drives synergies between energy, environmental and enterprise policy" is precisely the right diagnosis.[60] It is also the bridge to the analysis of regulatory independence and interdependence that Chapter 5 develops.

---

case responses to a structural imbalance that requires a structural framework — which the Commission's forthcoming Data Centre Energy Efficiency Package, expected in April 2026, should provide.

[58] Google Hamina (Finland): closed-loop seawater cooling through repurposed granite tunnels; PUE 1.09, zero potable water consumption. Meta Odense (Denmark): 215,000 MWh/year of waste heat supplied to approximately 12,000 homes (Ramboll project documentation). The Danish enabling condition was the abolition of the surplus heat tax: first reduced in January 2022, eliminated entirely in July 2025 — a fiscal instrument whose removal transformed the commercial viability of waste heat recovery across the sector.

[59] The Rozzano project (Milano metropolitan area) was completed in early 2025. TIM's data center was connected to Getec Italia's district heating network through the subsidiary Atmos, with heat exchangers and heat pumps distributing recovered thermal energy to 5,000 Aler social housing units in the Rozzano district. The estimated environmental benefit is approximately 3,500 tonnes of CO_2 avoided annually (https://www.today.it/speciale/game-changer/a-rozzano-il-calore-dei-server-riscalda-cinquemila-case.html).

[60] CRU2025236, section on future evolution of policy.

5. Regulatory Independence and the Challenge of AI Impact: Towards Interdependence

Regulatory independence is not a binary condition, nor is it equivalent to institutional isolation. An independent energy regulator operates within a web of accountability relationships — to the legislature that defines its mandate, to the consumers whose interests it is charged with protecting, to the courts that review its decisions, and to the broader public whose trust sustains its legitimacy. What independence requires is that decisions within the regulator's technical domain are made on the basis of evidence, expertise, and defined statutory objectives rather than executive preference or political convenience.

The rapid growth of AI data centers has subjected that condition to pressures that reveal something more fundamental than the familiar tension between regulatory autonomy and political interference. The challenge is not only that governments are pressing regulators to move faster, allocate costs differently, or favour particular industrial actors. It is that the AI infrastructure challenge — when its environmental sustainability dimension is added to its grid reliability and affordability dimensions — crosses the boundaries of what any energy regulator, however independent and however capable, can govern alone. The cases examined in this chapter illustrate a progression: from independence under pressure, through independence acknowledging its own limits, to independence rendered irrelevant by design. Together, they point toward a common conclusion — that the appropriate response to the AI challenge requires not the weakening of regulatory independence but its extension through deliberate interdependence with other institutional actors.

5.1 FERC: Independence Preserved, Sustainability Excluded

The Federal Energy Regulatory Commission's formal independence from the executive branch is constitutionally robust. Its commissioners serve fixed five-year terms and cannot be removed for policy disagreements; its decisions on rates, transmission access, and market design are insulated from direct executive override by the deliberate architecture of the Department of Energy Organization Act, which authorises the Secretary of Energy to "propose" rules to FERC but requires the Commission to "consider" rather than adopt them — a calibrated distinction between direction and deliberation.[61]

That architecture was tested visibly in the autumn of 2025. Secretary of Energy Chris Wright invoked Section 403(a) to transmit to FERC a detailed Advance Notice of Proposed Rulemaking on large load interconnection, setting April 30, 2026 as the deadline for final action — compressing a process that typically takes two to three years into approximately six months.[62] The proposal's fourteen principles address interconnection scope, size thresholds, joint studies, financial commitments, hybrid facility arrangements, cost allocation, and NERC compliance. None of the fourteen principles addresses the environmental sustainability of the loads to be connected — their carbon footprint, their energy sourcing commitments, their water consumption, or their compatibility with US climate commitments. The thirteenth and fourteenth principles, addressing transition

[61] 42 U.S.C. § 7173(a)-(b).

[62] Chris Wright, Secretary of Energy, Letter to FERC re Large Load Interconnections, 23 October 2025, with attached ANOPR. Available at www.energy.gov.

planning and NERC compliance respectively, are procedural and technical. The word “environment” does not appear in the document.

This silence is not incidental. It reflects a deliberate policy frame established at the highest level of the administration, of which the withdrawal from IPCC and IRENA processes forms part of the same coherent orientation. The FERC that Chair Swett leads — having declared that her priority is to “connect and power data centers as quickly and as durably as possible” — is formally independent and substantively aligned with an agenda in which sustainability is not a regulatory variable.[63] The most significant form of pressure on an independent regulatory commission is not illegal instruction but the reshaping, through personnel strategy and agenda-setting, of what the institution understands its job to be.

This pattern extends beyond large load interconnection. On 19 February 2026, FERC issued two orders expanding Categorical Exclusions from environmental review under NEPA for hydropower license terminations and ancillary construction activities — actions that had previously required Environmental Assessments or Environmental Impact Statements (Docket Nos. RM26-7-000 and CX26-1-000). Chair Swett described the move as “cutting unnecessary red tape” to accelerate decision-making “while ensuring we meet our statutory obligations under NEPA”.[64] The phrase is instructive: what counts as “unnecessary” is itself a regulatory choice, and one that consistently resolves in the direction of reduced environmental scrutiny across FERC’s portfolio.

The result is a regulatory framework in which large load interconnection is governed with technical rigour on dimensions of reliability, cost allocation, and non-discrimination — and with structural silence on sustainability. Independence, in this configuration, is real. But it operates within a frame that has been defined by political choice rather than technical necessity.

5.2 CRU: Independence and the Explicit Call for Interdependence

The Irish Commission for Regulation of Utilities presents the most instructive case in this comparative analysis — not because it succeeded where FERC struggled, but because it identified with unusual precision exactly where the boundaries of its own competence lie, and drew the institutional consequences.

Throughout the process culminating in its December 2025 final decision on large energy user connections, the CRU repeatedly acknowledged that the Climate Action and Low Carbon Development Act 2015 does not provide it with “sufficient legal basis to mandate explicit emissions reduction and offsetting measures” through electricity connection policy.[65] The regulator could require an 80 percent renewables obligation and mandate onsite dispatchable generation with future-proof fuel compatibility — and did. But it could

[63] Laura Swett, remarks at FERC open meeting, 20 November 2025, as reported in Diana DiGangi, "New FERC commissioners say connecting data centers is key priority", *Utility Dive*, 21 November 2025.

[64] FERC, Order H-1 (Docket No. RM26-7-000) and Order H-2 (Docket No. CX26-1-000), 19 February 2026; Chair Swett quoted in FERC press release, "FERC Streamlines Environmental Reviews", 19 February 2026, as reported in Cara M. MacDonald, "FERC Moves to Streamline Hydropower Environmental Reviews", Pillsbury Law, 13 March 2026.

[65] CRU2025236; the quotation is at page 5, paragraph “A State-led approach”. The CRU final decision was adopted and published few days after the publication of Government’s report with 30 recommendations for accelerating energy infrastructure development “*Accelerating Infrastructure Report and Action Plan*”, 3 December 2025 https://www.gov.ie/en/department-of-public-expenditure-infrastructure-public-service-reform-and-digitalisation/publications/accelerating-infrastructure-report-and-action-plan/

not impose emissions caps or carbon offsetting conditions, because its statutory remit, as it correctly interpreted it, did not extend that far.

This interpretive restraint had real consequences. As independent analysts noted, the CRU's decision was likely to exacerbate Ireland's overshooting of its sectoral emissions ceilings — renewable procurement is not the same as emissions reduction in a system where fossil generation still fills the gaps left by intermittent renewables.[66] The CRU's own assessment of this tension in the final decision document is remarkable in its directness: the medium- to long-term development of the data center sector requires, the CRU stated, *"a State-led approach encompassing spatial planning, targets-based and plan-led infrastructure development, which drives synergies between energy, environmental and enterprise policy*".[67] A regulator acknowledging in a formal decision document that the regulatory framework it operates within is inadequate for the problem at hand — and calling for legislative and governmental action to fill the gap — is a rare form of institutional self-disclosure. It is also the correct response: rather than papering over the gap with expansive interpretations of existing powers that would be legally fragile, the CRU identified the problem precisely and passed it, explicitly, to the institutions with the constitutional competence to address it.

This is interdependence in its most honest form. The CRU did not abandon its independence in calling for government action — it exercised it, by defining the boundary of its mandate with analytical rigour and refusing to pretend that boundary did not exist.

5.3 Spain: Sustainability Embedded in the Access Mechanism

The Spanish *concurso de demanda*, established by resolution published in the *Boletín Oficial del Estado* on 17 July 2025, illustrates a different model of institutional response — one in which the sustainability challenge is addressed not by the energy regulator alone but through a government-designed competitive mechanism that incorporates sustainability criteria into the access allocation process itself.

The *concurso* replaces the traditional first-come, first-served logic of grid access with a competitive scoring system. Applications for connection capacity at congested transmission nodes are ranked on three criteria: date of consumption commencement, investment volume, and — explicitly — greenhouse gas emissions avoided. The third criterion operationalises sustainability as a selection variable rather than a reporting obligation: projects that demonstrate a material contribution to decarbonisation obtain a competitive advantage over projects that do not.

Whether this design was developed in coordination with the CNMC, Spain's integrated energy and competition regulator, or primarily within the *Secretaría de Estado de Energía* cannot be confirmed from the available documentation. What the instrument demonstrates is that the integration of sustainability into grid access policy is technically and legally feasible — and that the appropriate institutional locus for that integration may be a government-regulatory collaborative rather than the regulator acting alone. The Spanish model is, in this sense, a functional illustration of the interdependence that the CRU called for in principle.

[66] Hannah Daly, "A new era of fossil fuels to facilitate data centres?", personal blog, 12 December 2025, https://hannahdaly.ie/2025-12-12-CRU-data-centre-policy/
[67] CRU2025236, section on future evolution of policy.

Spain has since taken this further. The Real Decreto-ley (RDL) 7/2026, adopted on 20 March 2026, introduces a structural reform of demand access regulation that goes beyond the *concurso de demanda* mechanism.[68] Its most significant innovations for the purposes of this analysis are three: a mandatory capacity reservation fee payable from the date of permit grant, operationalising the causer-pays principle at the queue stage rather than only at the connection stage (see above paragraph 3.2); a formal flexible access category that unlocks previously blocked grid capacity; and a shift from "first-come-first-served" to a readiness-based ordering criterion aligned with the European Grids Package guidance[69]. These are coherent extensions of the regulatory logic already present in the *concurso de demanda*.

The RDL 7/2026 also introduces a formal prioritisation hierarchy for demand access — with housing, essential services, and strategically designated industrial projects ranking above standard applications — and a readiness-based ordering criterion within each tier ("*first ready, first served*"). The criteria for 'strategic' designation are still to be defined in subsequent implementing measures, leaving some residual uncertainty about whether the process will be grounded in transparent, objective criteria of the kind the *concurso* had exemplified.

5.4 Italy: Sustainability Competence, but Unused

ARERA — Italy's Regulatory Authority for Energy, Networks and Environment — possesses, by European standards, a sophisticated regulatory toolkit and a mandate that explicitly encompasses both energy and environmental dimensions. Its experience with incentive regulation, capacity markets, demand flexibility schemes, and renewable integration is internationally recognised.[70] Yet ARERA's involvement in Italian data center policy in 2025-2026 has been marginal on the most consequential questions. The government's data center strategy published by MIMIT in July 2025 made no mention of ARERA's role in connection policy, cost allocation, or sustainability requirements for large energy users.[71] The recent Decree-law delegated to ARERA the technical implementation of connection rules but as for data center permitting procedures the regulator is not at all involved.[72] The subsequent discussion of classifying AI data centers as energy-intensive users *(energivori)*

---

[68] Real Decreto-ley 7/2026, de 20 de marzo, *por el que se aprueba el Plan Integral de Respuesta a la Crisis en Oriente Medio*, BOE-A-2026-6544, 21 marzo 2026, Arts. 11–13 and *Disposiciones Transitorias* 3ª–5ª. Convalidated by the Spanish Congress on 26 March 2026.; for flexible access Art. 2.k RD 1183/2020 as amended; for the strategic prioritisation framework Art. 13.

[69] European Commission, Guidance on efficient and timely grid connections, Commission notice, C/2025/6703, 12 December 2025, https://eur-lex.europa.eu/eli/C/2025/6703/oj/eng/pdf

[70] For instance, in terms of mitigating capex bias, ARERA launched the so-called ROSS (*Regolazione per Obiettivi di Spesa e di Servizio*) program (https://blueprint.raponline.org/casestudies/italy-step-by-step-to-modern-revenue-regulation/). Another important regulatory mechanism to avoid overforecasting is the "last regret criteria" used when assessing, according different future scenarios, target capacities at interconnections or at bidding zone boundaries (see E. M. Carlini *et alii*, *"Cost-Effective Target Capacity Assessment in the Energy Transition: The Italian Methodology"*, Energies, Volume 17, Issue 12, https://www.mdpi.com/1996-1073/17/12/2824)

[71] Ministero per l'industria e il *made in Italy* (MIMIT), "Strategia per l'attrazione degli investimenti esteri nei data center", July 2025 https://www.mimit.gov.it/it/strategie/datacenter

[72] *Decreto-legge* 20 febbraio 2026 n. 21 ("Decreto Bollette 2026"); in the attached technical report, there are impressive figures on data centers: 339 requests pending for total demand of 55,9 GW (equivalent to whole Italy peak demand!), half of which concentrated in Lombardy (https://www.greenreport.it/news/nuove-energie/60274-relazione-tecnica-del-decreto-bollette-ecco-come-ha-ragionato-il-governo-ma-i-conti-restano-incerti)

— a cost allocation decision with distributional consequences directly equivalent to the cross-subsidy risks analysed in Chapter 3 — is being developed as a matter of EU State aid positioning.[73]

The Italian case is not a failure of regulatory independence in the conventional sense — ARERA has not been pressured to reach a decision it would not otherwise have reached. It is a failure of regulatory relevance: a government choosing to develop data center policy through industrial policy channels in a space where the energy regulator has both the formal mandate and the technical competence to contribute, but has not been asked to. The paradox is acute: Italy's regulator covers energy *and* water — a combined mandate that is precisely the institutional configuration required for governing the full sustainability footprint of large data centers. That mandate is available and unused.

5.5 A Spectrum, Not a Typology

The four cases do not constitute a typology of regulatory failure. They constitute a spectrum of institutional responses to a challenge that exceeds, in each jurisdiction and for different reasons, what an energy regulator operating within its conventional mandate can address alone. At one end, FERC demonstrates that formal independence and the substantive exclusion of sustainability from the regulatory agenda are not contradictory. At the other, Italy shows that a capable and formally independent regulator with the right mandate can be rendered structurally irrelevant by the choices of the government it serves. Between them, Ireland shows what honest institutional self-awareness looks like, and Spain shows what the productive collaboration it calls for can produce.

The common lesson is that the AI infrastructure challenge cannot be governed by energy regulators acting alone, however independent and however competent. The appropriate institutional response is interdependence: clearly defined mandates, explicit coordination mechanisms between energy regulators and the governmental actors responsible for climate, planning, and industrial policy, and the institutional will to acknowledge the boundaries of what any single regulator can legitimately decide. The cases where that interdependence has begun to emerge — however imperfectly — are also the cases where regulatory outcomes most closely reflect the full range of public interests at stake.

Conclusions: When Balance Is Achieved — Independence, Interdependence, and the Governance of AI-Driven Load Growth

Jorge Vasconcelos, first Chair of Portugal's regulatory authority in the 1990s, has observed that the energy transition introduced a structural change in what regulation is asked to do: alongside efficiency — historically the core objective — it placed sustainability as an equally binding constraint.[74] The two do not always point in the same direction. An efficient connection is not always a sustainable one; a cost-reflective tariff does not automatically

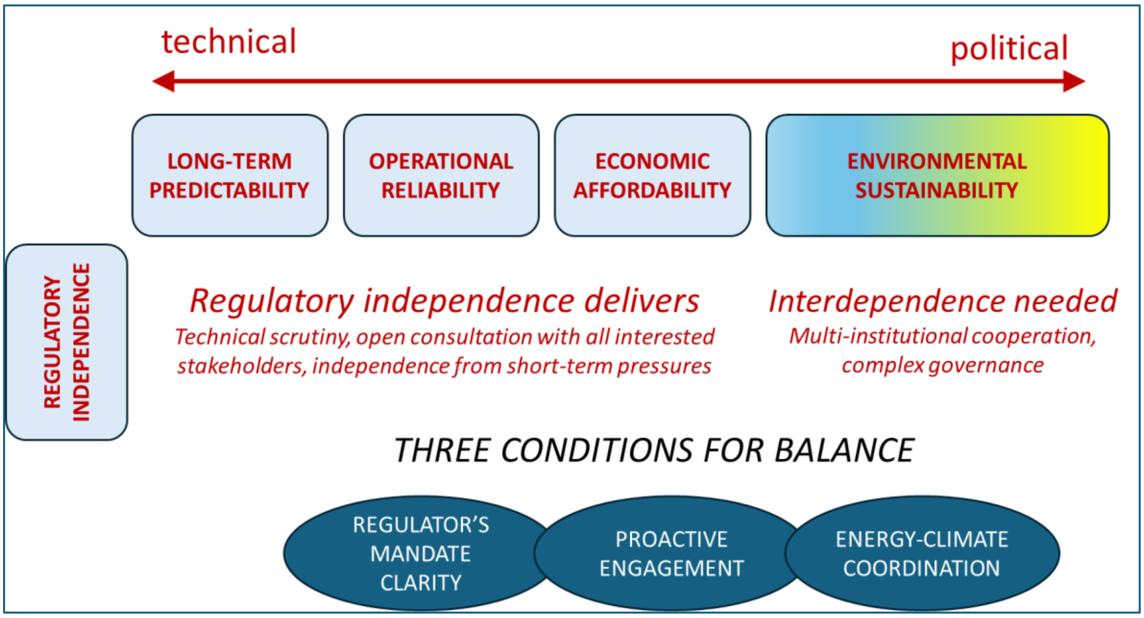
[73] "*Sgravi energivori ai data center: l'Italia ci prova*", QualEnergia.it, 10 December 2025 https://www.qualenergia.it/pro/articoli-pro/sgravi-energivori-data-center-italia-ci-prova/ (paywall).
[74] Jorge Vasconcelos, interview conducted for the 25th anniversary of ERRA, 2025 https://erranet.org/erra-25th-anniversary-podcast-transformation-in-regulating-energy-over-the-last-25-years/. Jorge Vasconcelos is currently a member of ERRA's Strategic Advisory Board.

produce a low-carbon grid. The governance challenge that AI-driven data center growth poses to energy regulators is, in its deepest form, the challenge of managing that tension at a scale and speed for which existing frameworks were not designed.

This article has examined that challenge across five dimensions and three continents. The analysis supports a conclusion that is both specific and general: **the regulatory tools needed to govern AI infrastructure already exist, they are unevenly exercised, and the gap between what regulation could achieve and what it is currently achieving is primarily an institutional and political problem — not a technical one.**

The Spectrum from Technical to Political

The first four dimensions examined in this article form a spectrum that impact on the fifth (regulatory independence).



At one end sit long-term predictability, operational reliability, and economic affordability — domains where the quality of regulatory decisions depends primarily on technical expertise, analytical rigour, and open consultation with all stakeholders that is democratic way to ensure protection from short-term pressures. The core questions in these domains — whether a demand forecast reflects structural incentive bias or genuine customer data, whether a power-electronic load poses voltage stability risks, who bears the cost of network upgrades triggered by a specific interconnection — are questions that trained regulators working with good data are better positioned to answer than political actors. They are questions of analysis.

At the other end sits environmental sustainability — and, specifically, the distributional and planning dimensions that sustainability entails. Decisions about whether a jurisdiction's carbon budget should constrain data center growth, how to balance the fiscal benefits of investment attraction against the energy burden placed on existing consumers, or where to site the renewable infrastructure that large loads require: these are choices about trade-offs between legitimate competing interests. They require democratic legitimation that technical regulators cannot provide alone.

Regulatory independence is not equally the right answer across this entire spectrum. Understanding where it is decisive, and where it must be complemented by deliberate interdependence with other institutional actors, is the organising lesson of the evidence assembled here.

Where Independence Delivers

On predictability, reliability, and affordability, the comparative evidence is consistent: independent regulatory and system operation institutions have demonstrated a clear capacity to manage the challenges posed by AI-driven load growth — when they exercise their mandates proactively. The AER's scrutiny of utility demand projections, NERC's characterisation of power-electronic load risks before those risks crystallised into incidents, the CRU's construction of a non-firm connection framework from first principles when the existing policy produced a de facto moratorium: these are not exceptional achievements. They are what independent regulation, functioning as designed, is supposed to produce.

The contrast cases are equally instructive. Where cost-of-service incentives systematically bias utilities toward overestimating demand — as documented in the US — the absence of a well-resourced independent regulator willing to challenge those projections means the bias enters the planning process uncorrected, and the costs are borne by consumers. Where tariff suppression substitutes for cost-reflective pricing — as in Korea[75]— the suppressed costs accumulate on public balance sheets and require eventual corrections more abrupt than any transparent regulatory process would produce. The technical case for independence on these dimensions is not a matter of institutional preference; it is a matter of demonstrated outcome.

Where Interdependence Becomes Necessary

The sustainability dimension is where the limits of conventional energy regulation become most visible — and where the argument of this article turns. No energy regulator, however independent and however competent, can govern alone the full set of questions that data center sustainability raises. Land use planning, water allocation, district heating integration, and the translation of national climate commitments into binding operational constraints on individual connection decisions: these require collaboration across institutional boundaries that the energy regulator can signal but not cross unilaterally.

The CRU identified this boundary with unusual precision and honesty in its December 2025 decision, explicitly acknowledging that the medium-term development of data center infrastructure in Ireland requires "a State-led approach encompassing spatial planning, targets-based and plan-led infrastructure development, which drives synergies between energy, environmental and enterprise policy." This is not a regulatory failure. It is the correct identification of where the mandate runs out — and the appropriate institutional response: acknowledge the boundary, exercise what authority exists within it rigorously, and pass the baton explicitly to the actors who hold the remaining instruments.

---

[75] South Korea has no independent energy regulator in the European or North American sense: the Korea Electricity Regulatory Commission operates under a ministry that is simultaneously a major shareholder of KEPCO, the vertically integrated incumbent. The IEA's *Korea 2025 Energy Policy Review* (October 2025) recommended that "*consumers and the energy sector would be better served by an independent market regulator*" (https://www.iea.org/reports/korea-2025/executive-summary). The consequences are documented: tariff suppression calibrated to anti-inflationary politics produced cumulative KEPCO losses of approximately 43 trillion won between 2021 and 2023, and data center connection confirmation times extended from two-to-three months to approximately twelve months — not from physical capacity constraints but from institutional gridlock.

Spain's *concurso de demanda* illustrates what the productive form of that collaboration can produce: a government-designed competitive mechanism that embeds greenhouse gas emissions avoided as a selection criterion for grid access, aligning the energy connection process with climate policy objectives that no energy regulator could impose alone on its own statutory authority. The mechanism works because it sits at the intersection of energy regulation and industrial policy — precisely the intersection that the CRU identified as requiring State engagement.

The contrast cases are equally instructive in the other direction. FERC operates within an administration that has withdrawn from IPCC and IRENA,[76] whose Section 403 proposal contains no sustainability dimension across fourteen principles, and whose new Chair has declared the priority of connecting data centers "as quickly and as durably as possible." Formal independence is intact; the sustainability dimension has been excluded from the regulatory agenda by political framing rather than statutory limitation. Italy's ARERA, by contrast, has the mandate — explicitly covering energy and environment — but has been excluded from the data center policy process by government choices that directed consequential decisions through industrial policy channels. In neither case is there a failure of independence in the conventional sense. In both, the outcome is the same: sustainability is not governed.

Three Conditions for Balance

The comparative evidence points to three institutional conditions under which the balance between independence and interdependence can be maintained.

The first is **mandate clarity**: regulators need to know, and to be publicly explicit about, where their authority runs and where it does not. The CRU's honesty about its statutory limits is a governance virtue, not a weakness. Regulators who paper over gaps with expansive interpretations of uncertain authority produce decisions that are legally fragile and institutionally costly when challenged. Regulators who identify the gaps precisely create the pressure for legislative response.

The second is **proactive engagement**: the difference between regulatory systems managing the AI challenge effectively and those failing to do so is less a matter of statutory authority than of institutional posture. The evidence from Australia, Ireland, and Germany consistently shows that regulators who engage before crises — who build frameworks in anticipation of the problem rather than in response to it — achieve outcomes that reactive systems cannot. Proactivity is not a legal question; it is a choice.

The third is **institutional coordination**: the governance of AI infrastructure requires that energy regulators, climate authorities, planning agencies, water regulators, and governments operate with shared situational awareness and explicit coordination mechanisms. The model is not merger but structured dialogue[77] — each institution

[76] President Trump's Memorandum for the Heads of Executive Departments and Agencies, 7 January 2026, points (xiv) and (xxiv) https://www.whitehouse.gov/presidential-actions/2026/01/withdrawing-the-united-states-from-international-organizations-conventions-and-treaties-that-are-contrary-to-the-interests-of-the-united-states/

[77] For more details and practical recommendations on inter-institutional collaboration on data centers issues, see: Regulatory Assistance Project, "*Building Resilient Foundations for Large Loads*", by L. Eberle and C. Kadoch, February 2026, https://www.raponline.org/knowledge-center/building-resilient-foundations-large-loads/

exercising its mandate within its domain, with transparent handoffs at the boundaries. Italy's data center governance gap is not primarily a failure of ARERA; it is a failure of the coordination architecture that should connect ARERA's technical competence to the government's industrial strategy.

An Open Moment, but Unstable and Risky

This article was written at a genuinely open moment in the governance of AI infrastructure. The financial assumptions underlying that infrastructure are themselves under scrutiny: the capital expenditure cycle that has driven data center demand forecasts to historic highs rests on revenue projections for AI services that remain, as yet, largely unvalidated by market returns.[78] At the same time, the energy markets on which both data center operations and grid planning depend are exposed to geopolitical shocks of a kind that no regulatory framework can absorb alone: the disruption of liquefied natural gas supply chains and the oil price volatility produced by conflict in the Gulf are reminders that the transition assumptions embedded in long-term grid plans rest on a very fragile geopolitical substrate.[79]

The regulatory challenge examined in this article does not disappear if the underlying business case proves weaker than currently assumed— but the scale and urgency of the governance problem may look different in two years than it does today.

The European Commission's AI and Energy Roadmap, expected in 2026, will test whether the EU's ex ante regulatory architecture can maintain coherence as data center growth scales. FERC's trajectory — formal independence within an agenda that has structurally excluded sustainability — will test whether procedural rigour is sufficient to protect the distributional dimension of large load governance when the framing has been defined by political choice. AEMO's transition planning is proceeding in parallel with a large-scale renewable build-out that is itself reshaping the technical conditions under which data centers connect.

The regulatory institutions examined here — CRU, CNMC, ARERA, AEMO, AER, NERC, FERC, ENTSO-e, and ACER — did not design the AI transition. They are managing it, with varying degrees of effectiveness, within institutional structures built for a different era. The institutions doing this most effectively share a recognizable profile: they engage early with the policy choices that frame their technical mandate; they apply consistent, analytically rigorous standards that do not vary with the political salience of the decision; and they communicate those standards transparently enough that investors, consumers, and governments can predict regulatory outcomes and build them into their strategies.

These characteristics are not jurisdiction-specific. They do not require a particular constitutional architecture or market structure. They require independence, capacity, and the institutional will to exercise both — combined, when the problem reaches the boundary of the technical mandate, with the willingness to say so clearly and to build the collaborative frameworks that the boundary requires.

---

[78] Matteo Wong and Charlie Warzel, "Welcome to a Multidimensional Economic Disaster," *The Atlantic*, 26 March 2026.

[79] "Even the Best-Case Scenario for Energy Markets is Disastrous", *The Economist,* 22 March 2026 (updated 23 March 2026)

For regulators in ERRA's member community, operating across diverse governance traditions from Eastern Europe to Asia to the Middle East and Africa, that is the actionable conclusion. The toolkit is available — the challenge of AI-driven load growth is not, at its core, a problem that requires new regulatory instruments so much as the disciplined and proactive exercise of those that already exist.[80] The question is whether the institutional conditions — mandate clarity, engagement, coordination — are in place to put that toolkit at work in practice. Where they are, the evidence suggests that balance, in the sense the theme of this World Forum of Energy Regulation invokes, is achievable. Where they are not, the evidence is equally clear about the costs.

[80] The toolkit is composed of: Prohibition Regulation; Process Regulation; Input-based Regulation; Output-based Regulation; see ERRA working paper, "*Regulatory and Policy Frameworks for AI Data Centers in EU and US* 2025", p. 5 and Tables 1a and 1b, p. 15 https://erranet.org/regulatory-and-policy-frameworks-for-ai-data-centers-in-eu-and-us/. See also *"Resetting Antidiscrimination Law in the Age of AI"*, Harvard Law Review, Volume 138, Issue 6, April 2025, https://harvardlawreview.org/print/vol-138/resetting-antidiscrimination-law-in-the age-of-ai/.