



ASCENDANCY ADVISORS LIMITED

The Physical Cost of Artificial Intelligence

Energy, Water and the Emerging Regulation of Compute Infrastructure

Research Paper

8 October 2026 | Version 1.2 | English authoritative text (UK English)

Prepared by Guy Duport for Ascendancy Advisors Limited

Core proposition

Artificial intelligence is increasingly discussed as a software industry while being built as a heavy industry. Its physical substrate—electricity, water, land, transmission, cooling, backup generation, semiconductor fabrication, grid hardware, fibre networks and increasingly dedicated generation assets—has become a material public-policy, industrial-strategy and national-security issue.

Scope

United States · China · European Union and Ireland · Australia · South Africa and wider Africa · Taiwan · Orbital compute

Abstract

The rapid scaling of artificial intelligence is changing the economics and politics of digital infrastructure. Data centres are no longer merely background facilities supporting internet services. AI training and inference are driving a new class of high-density, highly concentrated electrical loads whose requirements increasingly intersect with power-system planning, water allocation, industrial policy, land use and community consent. The

International Energy Agency projects global data-centre electricity consumption to rise from roughly 415–460 TWh in 2024 to around 945 TWh by 2030, with the United States and China accounting for the majority of incremental growth. Yet global percentages understate the policy challenge: data-centre demand is geographically concentrated, can arrive faster than generation and transmission infrastructure, and may create local competition for scarce capacity.

This paper examines the physical-resource implications of AI infrastructure through comparative case studies in the United States, China, the European Union and Ireland, Australia, South Africa and wider Africa, and Taiwan. It distinguishes direct facility water use from indirect water use embedded in electricity generation; cautions against simplistic per-query water claims; and proposes that data-centre policy should be assessed through three linked tests: additionality, infrastructure sovereignty and local benefit. The paper argues that governments should neither prohibit AI infrastructure reflexively nor subsidise it unconditionally. Instead, hyperscale compute should be treated as strategic heavy infrastructure: large projects should disclose their energy and water profiles, pay the incremental infrastructure costs they cause, demonstrate credible additional generation or storage where grids are constrained, protect essential household and industrial supply, and provide measurable domestic benefits.

The central conclusion is that the next phase of AI governance will not be limited to model behaviour, safety and liability. It will also govern the physical systems required to produce intelligence at scale. In many jurisdictions, the decisive question will be less whether AI is desirable than who receives scarce electricity and water, who pays for the supporting infrastructure, who controls the resulting compute capacity, and who captures the economic value.^[1]

Executive findings

- AI infrastructure is becoming an energy-system issue. The IEA expects global data-centre electricity consumption to more than double by 2030. The challenge is not simply the global share of electricity demand, but the speed and geographic concentration of new loads.
- Water claims require discipline. Direct cooling consumption, water withdrawals and indirect water consumption associated with electricity generation are different metrics. A low on-site WUE can coexist with a high system water footprint, and vice versa.
- The United States is the clearest example of compute demand moving upstream into generation. Berkeley Lab's 2026 update estimates data centres could account for 11.8% of US electricity use by 2030, with a scenario range of 9.5–15.3%.^[2]
- China is pursuing a more coordinated compute-energy geography through the 'East Data, West Computing' programme, national hub nodes, green-power targets and direct renewable supply. This model does not eliminate water or grid constraints, but it shows the strategic advantage of aligning compute siting with generation and transmission planning.
- Europe is moving from voluntary sustainability claims towards mandatory measurement. The EU reporting framework now requires qualifying data centres to report energy and water performance, while Ireland has gone further by linking new grid connections to local generation/storage and 80% additional renewable electricity.
- Australia has become a live test of social licence. A 2026 Senate inquiry is examining energy, water, environmental and regulatory impacts, while the withdrawal of Sydney's 90 MW Project Mars illustrates that technically viable projects can still fail politically when community consent and infrastructure impacts are poorly managed.^[3]
- South Africa demonstrates the distributional problem most starkly. More than 560 million people in sub-Saharan Africa still lack electricity access, while South Africa is simultaneously attracting hyperscale capacity. The correct policy response is not to reject data centres, but to make additionality, reliability, affordability and local benefit explicit conditions of development.^[4]
- Taiwan offers a useful energy-security case. Since 2025, hyperscale and colocation facilities at or above 5 MW have been required to submit energy utilisation manuals and adopt best available techniques; 2026 proposals add industrial-benefit review to the approval process.^[5]
- Orbital compute is no longer purely speculative. Google's Project Suncatcher prototype entered orbit on 1 October 2026, while SpaceX has sought US approval for an orbital data-centre constellation of up to one million satellites. Space shifts rather than eliminates externalities: terrestrial water and land constraints may fall, but launch demand, orbital congestion, debris, radiation and thermal-management challenges rise.^[6]
- The strongest regulatory principle is additionality: large new compute loads should, where systems are constrained, be paired with genuinely additional generation, storage,

transmission and—where relevant—water infrastructure rather than merely purchasing priority access to existing public capacity.

- AI infrastructure sovereignty should be understood as the capacity of a state to host and control strategically important compute without degrading essential services, imposing uncompensated costs on households, or becoming dependent on infrastructure it does not govern.

The physical bottleneck is broader than generation. Transformers, switchgear, fibre, subsea connectivity, semiconductors and construction materials can all become binding constraints; some grid components have procurement lead times measured in years.[29]

Climate resilience and distributional fairness should become explicit approval tests. Extreme heat, drought and network disruption can worsen facility impacts, while communities may bear noise, backup emissions, transmission corridors and water burdens without capturing proportionate economic value.

Compute nationalism is emerging alongside environmental regulation. Export controls, semiconductor sovereignty and differences in water/carbon rules can shift the geography—and therefore the physical footprint—of AI without necessarily reducing its global resource demand.[32]

Contents

1. Introduction: AI is becoming heavy industry.....	5
2. Methodology and analytical framework.....	6
3. The physical stack: electricity, water, land and heat.....	7
4. Comparative policy snapshot.....	10
5. United States: compute demand moves upstream into generation.....	10
6. China: infrastructure planning as compute policy.....	13
7. European Union and Ireland: from disclosure to conditional access.....	14
8. Australia: energy transition meets social licence.....	15
9. South Africa and wider Africa: the allocation dilemma.....	16
10. Taiwan: when compute sovereignty meets energy security.....	17
11. A regulatory framework for the physical cost of AI.....	18
12. Policy recommendations.....	21
13. Limitations and areas for further research.....	23
14. Conclusion.....	24
Bibliography.....	24
Source notes.....	26

1. Introduction: AI is becoming heavy industry

Artificial intelligence is usually described through models, algorithms, applications and software. That description is increasingly incomplete. The frontier of AI depends on a rapidly expanding physical system: advanced semiconductors, server racks, substations, high-voltage transmission, cooling equipment, water systems, backup generators, batteries, fibre routes, land and, increasingly, dedicated or contracted electricity generation. In policy terms, AI is acquiring characteristics more familiar from aluminium smelting, petrochemicals and other capital-intensive industries than from the conventional image of a software company.

The distinction matters because software scales differently from infrastructure. A model can be copied quickly; a transmission line, nuclear uprate, reservoir connection or new substation cannot. A data-centre campus may be planned and built within several years, while major grid reinforcements, power stations and water projects can require much longer. The resulting mismatch creates a governance problem: private demand can materialise faster than public infrastructure can respond.

The scale of the transition is visible in the IEA's global projections. Its 2025 Energy and AI analysis estimates that data-centre electricity consumption rises to about 945 TWh in 2030, just under 3% of global electricity demand, with consumption growing around 15% annually from 2024 to 2030. Accelerated servers, largely associated with AI, are projected to grow much faster than conventional server demand. A 3% global share can sound modest. But the average obscures concentration. Data centres cluster around fibre, power, tax incentives, cloud regions and existing technology ecosystems. A local grid therefore experiences the project as a 100 MW, 500 MW or even gigawatt-scale load, not as a fractional percentage of world electricity consumption.^[7]

The policy question is consequently not whether AI data centres use 'too much' electricity or water in the abstract. The relevant questions are more specific: Is the load additional to what the system was built to serve? Can generation, transmission and water infrastructure absorb it without reducing reliability? Who pays for reinforcement? Does the project create new supply or merely secure priority access to existing capacity? What economic value remains in the host jurisdiction? And are communities given adequate information and influence over siting decisions?

This paper uses the concept of infrastructure sovereignty to organise those questions. Infrastructure sovereignty does not imply autarky or state ownership. It means that a jurisdiction retains practical control over the allocation, resilience and strategic consequences of infrastructure on which the wider economy depends. A government may welcome foreign-owned hyperscale compute while still requiring that it strengthens rather than weakens the host country's energy and water systems.

2. Methodology and analytical framework

The paper is a comparative policy analysis based on publicly available government, regulator, international-agency, laboratory and industry evidence available through 7 October 2026. Priority is given to primary sources: the International Energy Agency; US Department of Energy and Lawrence Berkeley National Laboratory; European Commission; national regulators; parliamentary inquiries; and official Chinese, Taiwanese and South African sources. Civil-society and industry submissions are used where they illuminate contested policy positions, but their estimates are identified as advocacy or stakeholder claims rather than treated as neutral facts.

Three methodological cautions are important. First, data-centre capacity measured in megawatts is not equivalent to annual electricity consumption. A facility's nameplate or contracted load may exceed its realised average demand, particularly during phased build-out. Secondly, water use must distinguish withdrawal from consumption. Water withdrawn for cooling and returned to a source is environmentally different from water evaporated or otherwise removed from the local cycle. Thirdly, direct facility water use and indirect water embedded in electricity generation must not be conflated. They are both relevant, but they answer different policy questions.

The paper therefore applies eight recurring tests to each jurisdiction: (1) scale and speed of compute growth; (2) electricity-system capacity and reliability; (3) direct and indirect water exposure; (4) material, grid-hardware and network dependencies; (5) climate and operational resilience; (6) regulatory transparency; (7) additionality and cost allocation; and (8) local, distributional and strategic benefit. These tests are intended to be comparable across very different political and energy systems.

3. The physical stack: electricity, water, land and heat

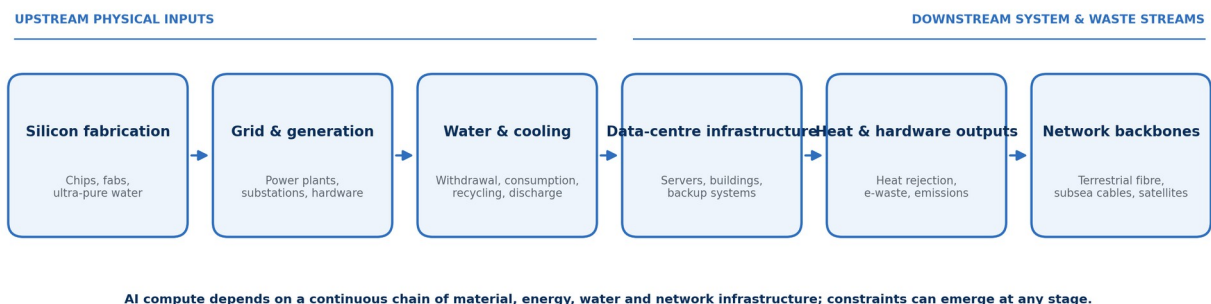


Figure 1. The physical stack of AI: an integrated industrial system rather than a weightless software service. Upstream elements include extraction, semiconductor fabrication, grid equipment and construction materials; downstream elements include heat rejection, water discharge, hardware end-of-life and network dependencies.

3.1 Electricity: the binding constraint

Electricity is the most immediate constraint because compute is continuous, concentrated and quality-sensitive. Modern AI facilities require not only large quantities of power but highly reliable power. The physical system must therefore support generation adequacy, transmission capacity, voltage stability, redundancy and emergency backup. The IEA expects the United States and China to account for nearly 80% of global growth in data-centre electricity consumption to 2030. US consumption is projected to increase by roughly 240 TWh from 2024 levels and China's by roughly 175 TWh.^[8]

The difference between average annual energy and peak or firm capacity is critical. A data centre can sign a renewable power purchase agreement equal to its annual consumption while still relying on the grid when wind or solar output falls. Annual matching is therefore not the same as physical additionality, hourly matching or system reliability. Policy frameworks increasingly recognise this distinction by requiring storage, firm generation, demand flexibility or geographically relevant new renewable projects rather than accepting certificates alone.

The physical bottleneck is often transmission rather than generation. A country can have substantial renewable resources and still be unable to deliver power to the location or on the timetable required. Large loads can also change the economics of network investment. If public tariffs socialise upgrades while a private hyperscaler receives most of the benefit, households and existing industry may effectively subsidise the project. Conversely, if the hyperscaler pays for additional infrastructure that strengthens the wider grid, the investment can create positive spillovers.

The bottleneck can also sit inside the grid supply chain itself. High-voltage transformers, switchgear, circuit breakers and power electronics are not interchangeable commodities that can always be procured on a data-centre construction timetable. US Department of Energy

material reports distribution-transformer lead times that rose from roughly three to six months in 2019 to 12–30 months in 2023, while large power transformers can require several years. This matters beyond the United States: rapid global electrification, ageing equipment and large-load growth create competition for the hardware needed to connect new compute. A country may therefore possess adequate generation potential yet still lack the physical equipment required to deliver it on time.[29] Self-provisioning does not remove regulatory friction. Behind-the-meter gas generation, microgrids and dedicated nuclear supply can still require utility interconnection studies, regional transmission operator approval, protection coordination and agreed rules for islanding, synchronisation and emergency backup. A project described as electrically self-sufficient may therefore remain operationally and legally dependent on the wider power system.

3.2 Water: a local resource, not a universal ratio

Water has become the most politically salient environmental issue because many cooling systems consume water through evaporation. Yet the headline numbers are unusually easy to misuse. Berkeley Lab's 2024 US assessment estimated direct data-centre water consumption of 66 billion litres in 2023. It also estimated a much larger indirect water footprint—nearly 800 billion litres—associated with the generation of electricity consumed by data centres. The same report emphasises that water intensity varies with cooling system, climate, power source and location.^[9]

That distinction makes simplistic statements such as 'one AI query uses X litres of water' unsuitable as a universal measure. A 2024 study modelling African data-centre water efficiency estimated that a ten-page report generated with GPT-4 could, under its assumptions, use up to about 60 litres of water, while an equivalent Llama-3-70B task was estimated at about 0.7 litres. The sixty-fold difference is itself evidence that these are modelled, architecture- and location-dependent results, not fixed physical constants. The study also found that eight of the eleven African countries examined had lower estimated water consumption than the global average for the same model, largely because of differences in electricity-generation water intensity.^[10]

Water policy should therefore focus on site-specific scarcity, drought resilience, cooling design, source quality and the opportunity cost of consumption. A litre used in a water-abundant basin is not equivalent to a litre consumed during drought in a stressed municipal system. For siting decisions, absolute consumption during drought conditions can be more important than a favourable Water Usage Effectiveness ratio.

Water impact also has a quality dimension. Cooling systems can generate concentrated blow-down, brine or chemically treated wastewater, while once-through or other discharge systems can alter receiving-water temperatures. The regulatory question is therefore three-dimensional: withdrawal, consumption and discharge quality/temperature. US environmental rules already illustrate the distinction by regulating cooling-water intakes and thermal discharges because heat, physical stress and treatment chemicals can affect aquatic

ecosystems. Closed-loop and reclaimed-water systems can reduce these impacts materially, but they do not eliminate the need for site-specific assessment.[30]

3.3 Heat, land and supporting systems

Nearly all electricity entering a data centre ultimately becomes heat. Cooling is therefore not an optional auxiliary function but an integral part of the compute system. Higher-density AI racks are accelerating liquid-cooling deployment, which can improve heat transfer and reduce compressor use but may alter water, plumbing and facility-design requirements. Waste-heat reuse offers a partial mitigation in cooler climates and dense urban systems with district heating, but it depends heavily on local heat demand and network economics.

Climate change tightens this relationship. Higher ambient temperatures can reduce the efficiency margin of cooling systems at the same time that wider electricity demand rises because of air-conditioning and heat stress. Drought can constrain water-based cooling; wildfire or flood can threaten transmission and fibre routes. For major campuses, planning should therefore move beyond annual-average efficiency towards climate-resilience stress testing under extreme heat, drought and network-contingency conditions.

Land is usually a smaller national constraint than electricity, but it can become decisive locally. Large campuses need secure sites, fibre access, substations, backup systems and sometimes buffer zones. The policy conflict is therefore cumulative: a project may be acceptable when assessed as a building yet much more consequential when its power, water, grid, land and backup-generation effects are assessed together.

The physical stack also extends beyond the data-centre fence. Semiconductor fabrication requires substantial electricity, ultra-pure water, chemicals and capital equipment; campuses require concrete, steel, copper and aluminium; and the supporting digital network depends on terrestrial fibre, landing stations and subsea cables. Grid equipment itself has critical-material and manufacturing dependencies. This paper does not attempt a full lifecycle assessment, but excluding those upstream systems from operational metrics should not be mistaken for absence of physical cost. Embodied emissions and material bottlenecks can be front-loaded years before a facility reaches full utilisation.

3.4 Beyond the fence: materials, grid hardware and networks

This broader boundary matters strategically even where it is not fully quantified. Constraints in semiconductor tools, advanced accelerators, transformers, copper, optical fibre or subsea connectivity can limit practical compute expansion as decisively as local megawatts. The result is an industrial supply chain rather than a self-contained digital service: compute capacity ultimately depends on manufacturing depth, logistics, energy systems and communications infrastructure.

4. Comparative policy snapshot

Jurisdiction	Primary constraint	Current policy direction	Additionality signal	Water governance	Strategic issue
United States	Grid capacity; local water; cost allocation	State/local controls plus federal energy response	Growing through PPAs, nuclear uprates and dedicated generation	Fragmented; facility and utility dependent	Can grid investment keep pace without shifting costs to households?
China	Regional imbalance; east-west latency; local water	State-directed hub siting and compute-power coordination	Strong policy emphasis on green-power supply and western resource regions	Varies by region; efficiency and siting increasingly central	Can national coordination align compute demand with power and water geography?
EU / Ireland	Transparency; constrained grids	Mandatory EU reporting; Ireland connection conditions	Ireland requires local generation/storage and 80% additional renewables	EU database includes water-footprint metrics	Can mandatory measurement become enforceable resource policy?
Australia	Grid transition; water; social licence	Senate review; emerging national standards	Strong policy debate around additional renewables and firming	Major transparency gap; proposed standards	Can growth coexist with energy transition and community consent?
South Africa / Africa	Access, reliability, affordability, drought	Human-rights and regulatory scrutiny emerging	Not yet systematic	Disclosure gaps; high local sensitivity	Should scarce infrastructure serve hyperscale loads before basic reliability is resolved?
Taiwan	Energy security; semiconductor/AI load growth	≥5 MW projects face energy-utilisation review and BAT requirements	Planning-stage scrutiny rather than blanket entitlement	Secondary to power in current framework	How should compute sovereignty be balanced against a constrained island energy system?

5. United States: compute demand moves upstream into generation

The United States is the most important near-term test because AI investment is colliding with a power system that had experienced comparatively slow load growth for much of the previous two decades. Berkeley Lab's 2026 update estimates that data centres could account for 11.8% of total US electricity use by 2030, with scenarios ranging from 9.5% to

15.3%. This is a materially larger burden than the laboratory's previous 2028 range and reflects the speed of planned AI deployment.^[11]

The US Energy Information Administration's October 2026 outlook similarly describes record electricity demand, with data-centre development and manufacturing driving commercial and industrial growth. This matters for regulation because the system is no longer allocating a largely static pool of electricity. It is entering a period in which transmission, generation and demand all need simultaneous expansion.^[12]

The market response increasingly resembles industrial self-provision. Technology companies are signing long-duration power contracts, supporting new renewable projects, exploring gas generation and batteries, and underwriting nuclear restarts, advanced reactors and uprates. On 6 October 2026, Google and Constellation announced a 20-year arrangement intended to add 890 MW of incremental nuclear capacity through uprates across eleven operating reactors in the PJM region. Google states that the arrangement brings the new nuclear capacity it has enabled through uprates and restarts to more than 1.5 GW. This is a useful illustration of additionality in practice: the anchor customer helps finance new firm capacity that is added to the wider grid rather than simply reserving existing output.^[13]

But the US governance problem remains fragmented. Siting, water, electricity tariffs and tax incentives are governed through different local, state and federal institutions. The result can be a project whose economic-development incentives are negotiated separately from the utility's grid-reinforcement costs and from the municipality's water obligations. Public opposition has therefore become a material project risk, particularly where communities believe that power bills, water scarcity or tax concessions are being socialised.

The speed of demand growth creates a second risk: near-term fossil-fuel lock-in. Where new firm low-carbon supply and transmission cannot arrive quickly enough, utilities and developers may preserve or add gas-fired capacity to meet large-load growth. Dedicated generation can improve reliability and cost causation, but it should not be treated as automatically clean merely because it sits behind the meter. Project review should therefore distinguish additional firm capacity from additional low-carbon capacity.

Small modular and advanced reactors are increasingly presented as a long-term answer for high-density AI demand. They may eventually provide useful firm, low-carbon power close to major loads, but policy should resist treating them as an immediate solution. US federal programmes in 2026 still describe major Gen III+ SMR deployment as a 2030s objective, while licensing, supply-chain, siting and first-of-a-kind execution risks remain material. AI campuses can often be built faster than new nuclear plants can be licensed and commissioned.^[31]

The correct regulatory response is not a national moratorium. The US benefits from rapid AI infrastructure development and has an obvious strategic interest in maintaining frontier compute capacity. The stronger principle is cost causation: projects that trigger identifiable

generation, transmission, distribution or water investment should bear those incremental costs unless a public authority can demonstrate a broader system benefit that justifies socialisation.

Physical infrastructure is also becoming part of strategic technology competition. US export controls seek to restrict Chinese access to advanced semiconductor manufacturing equipment, high-bandwidth memory and leading-edge AI capabilities. The infrastructure consequence is not mechanically predictable: constrained access to frontier hardware can increase the importance of software optimisation, domestic chip development and larger fleets of less-advanced equipment. The broader point is that controls on compute technology can shift energy, materials and capital requirements across jurisdictions. Compute policy and energy policy are therefore becoming increasingly difficult to separate. [32]

5.1 Water and geographic concentration

US national water totals are useful for scale but poor guides to local impact. Berkeley Lab estimates 66 billion litres of direct water consumption by data centres in 2023 and notes that hyperscale and colocation facilities accounted for 84% of that total. It projected hyperscale direct consumption alone at 60–124 billion litres in 2028 under its scenarios. The same report, however, shows why water-free cooling cannot automatically be treated as environmentally superior: air-cooled systems may reduce direct water consumption while using more electricity, potentially shifting water use upstream to the power system.^[14]

A sound US framework would therefore disclose both site and source water intensity, identify the basin or utility from which water is drawn, state the drought-case demand, and require operators to explain how cooling choices affect both electricity and water. Regulators should resist single-metric optimisation.

5.2 Beyond terrestrial compute

Resource constraints are already pushing the industry to consider an extreme alternative: moving part of the compute stack off-planet. Google's Project Suncatcher is a long-term research programme exploring solar-powered machine-learning infrastructure in low-Earth orbit. Google launched its first prototype satellite on 1 October 2026 to test TPU hardware under radiation, thermal and spaceflight conditions. The attraction is obvious: near-continuous solar exposure and the possibility of reducing terrestrial land and water constraints.^[15]

SpaceX has gone much further conceptually. In February 2026 the US Federal Communications Commission accepted for filing a SpaceX application for a non-geostationary 'Orbital Data Center' system of up to one million satellites. An application is not an operational system, and enormous engineering and economic questions remain. Orbital compute must deal with radiation, thermal rejection in vacuum, launch cost, failure replacement, network latency, collision avoidance and debris.^[16]

The policy significance is nevertheless real. Orbital compute demonstrates that AI's physical cost cannot be understood only as a terrestrial environmental problem. Moving computation into space would exchange some constraints for others. Water consumption and land competition could fall, while launch emissions, orbital congestion, debris risk, astronomy interference and space-traffic governance would rise. The externality is displaced, not abolished.

6. China: infrastructure planning as compute policy

China provides the strongest contrast with fragmented, project-by-project development. Its 'East Data, West Computing' strategy seeks to place more compute in western regions with greater land and energy availability while serving demand concentrated in the east. National policy explicitly coordinates compute, data, algorithms, electricity and network infrastructure, and directs large new data centres towards designated national hub nodes.^[17]

This coordination does not remove upstream constraints. China remains exposed to restrictions on leading-edge semiconductor equipment and advanced accelerators, even as domestic suppliers improve. A sustained gap in hardware efficiency could increase the physical resources required for a given level of computation, but that effect is neither automatic nor one-directional: model optimisation, system architecture, workload design and domestic innovation can offset part of the disadvantage. The policy significance is that semiconductor sovereignty and energy efficiency now interact.

This matters because it treats compute geography as a national infrastructure problem rather than merely a real-estate decision. The model is not frictionless. Latency-sensitive workloads cannot all be moved thousands of kilometres from users; western regions also face their own water constraints; and transmission, fibre and renewable integration remain engineering tasks rather than administrative declarations. Yet the strategic advantage is clear: compute demand is being planned alongside the power system rather than allowed to arrive independently of it.

The green-power dimension has also become more concrete. China's National Data Administration reported that national hub-node data-centre clusters had average PUE around 1.3, with leading facilities as low as 1.04, and set a target for new data centres in national hubs to reach more than 80% green electricity. In 2026, the National Development and Reform Commission highlighted a 500 MW photovoltaic project supplying the Zhongwei cloud base in Ningxia through a direct and virtual green-power arrangement. At full operation, the project is expected to generate 4.3 TWh annually and provide 2.29 TWh of clean electricity to the cloud base, enough to cover the core compute rooms cited by the NDRC.^[18]

The Chinese case therefore supports a broader argument for additionality. Instead of asking only whether a data centre can purchase enough electricity, policy asks where compute should be located, how it interacts with renewable resources and whether transmission and

dispatch can support it. Western jurisdictions need not copy China's governance model to learn from the infrastructure logic.

Water remains the harder comparative variable. China's northern and north-western regions can offer abundant solar and wind resources while also facing water scarcity. The optimal location for electricity is not necessarily the optimal location for evaporative cooling. The next stage of Chinese compute planning will therefore require even tighter integration of power, climate and water, including wider use of liquid cooling, dry cooling, recycled water and workload shifting where latency allows.

7. European Union and Ireland: from disclosure to conditional access

The European Union has taken the clearest step towards mandatory transparency. Under the recast Energy Efficiency Directive and Delegated Regulation (EU) 2024/1364, qualifying data-centre operators report energy-performance indicators to a European database, including information relevant to water footprint. The Commission states that it is also developing minimum performance standards, with a study launched in 2026.^[19]

Europe also provides an early model for extending physical-cost regulation into hardware lifecycle. EU ecodesign rules for servers and data-storage products include requirements relevant to disassembly, reuse, recycling, critical raw materials, secure data deletion and firmware availability. The European Commission is simultaneously developing data-centre sustainability ratings and minimum-performance standards. This suggests a future in which approval and procurement may increasingly consider not only operational PUE and WUE but also reparability, reuse and end-of-life management.^[33]

A stricter European regime also highlights the risk of regulatory leakage. If one jurisdiction prices water scarcity, carbon intensity and grid reinforcement while another does not disclose them, marginal workloads can migrate without reducing global resource use. Full international harmonisation is unrealistic, but a minimum common disclosure baseline for electricity, water, emissions and capacity would make such relocation visible and reduce incentives for a race to the bottom.

Transparency is necessary because the sector has historically relied heavily on voluntary corporate sustainability reporting. Those reports can be useful but are not a substitute for comparable, regulator-defined metrics. A common database creates the possibility of distinguishing efficient facilities from efficient claims.

Ireland illustrates the next stage: conditional access to constrained infrastructure. Data-centre electricity demand rose from 5% of national demand in 2015 to 22% in 2024. In December 2025, the Commission for Regulation of Utilities adopted a new connection policy requiring new data centres to provide generation and/or storage capacity at or near the site to match their requested maximum import demand. It also requires at least 80% of annual demand to be met by additional renewable projects in the Republic of Ireland, subject to a six-year development glide path.^[20]

Ireland is important because it moves beyond reporting. The regulator is effectively saying that a large new load has obligations to the system that go beyond paying a standard tariff. The policy also uses a serious definition of additionality: supported or already-contracted renewable projects cannot simply be relabelled to satisfy the requirement.

The Irish model will not transfer mechanically to every market. Requiring matching generation may be inefficient where the wider grid has ample capacity or where system-scale investment is cheaper. But as a principle, it provides a strong template for constrained jurisdictions: new strategic loads should arrive with a credible plan for the supply, flexibility and network capacity they require.

8. Australia: energy transition meets social licence

Australia's debate is unusually valuable because it is unfolding in public while the sector is still expanding. In May 2026 the Senate Environment and Communications References Committee opened an inquiry specifically into artificial intelligence and data centres, including the adequacy of existing regulation and impacts on communities, industry, the environment, water and energy. Public hearings were held in September and on 1 October, with the committee due to report in November.^[21]

Australia also illustrates the timing conflict between hyperscale growth and power-sector decarbonisation. Where transmission, storage and new firm supply lag, large new loads can strengthen the commercial case for keeping legacy thermal generation available longer than climate pathways assumed. That does not make data centres the sole cause of fossil lock-in, but it makes their marginal demand relevant to retirement schedules and capacity-market design.

Submissions reveal a sharp policy divide. Industry emphasises investment, digital sovereignty and the opportunity to use data centres as anchor customers for new power infrastructure. Environmental and civil-society groups emphasise the speed of load growth, limited water disclosure and the risk that data-centre demand could delay coal retirement or increase network costs. The Climate Council, for example, argues for binding additional renewable generation and firming, best-practice energy and water standards, demand flexibility, cost recovery and greater transparency. Those are advocacy recommendations rather than settled government policy, but they demonstrate how quickly additionality has moved from a technical concept to a political demand.^[22]

The most striking Australian development is social licence. Goodman Group withdrew its proposed A\$1.2 billion, 90 MW Project Mars data centre in Sydney's Lane Cove in late September after intense local opposition. The project was only about 20 metres from homes. Goodman cited the evolving policy and regulatory environment and community feedback, after acknowledging during the Senate inquiry that the project had not secured social licence.^[23]

The lesson is broader than Australia. Data-centre projects can no longer assume that classification as 'digital infrastructure' shields them from the politics faced by power stations, mines, warehouses or industrial plants. Noise, diesel backup, transmission works, visual impact, land use, water and electricity prices are local experiences. Community engagement must therefore occur before project design becomes effectively irreversible.

Australia also illustrates the danger of planning on annual averages. A power system simultaneously retiring ageing coal, adding variable renewables, electrifying transport and industry, and absorbing large data-centre loads must plan for firm capacity, transmission and system services. A data centre that contracts enough annual renewable energy may still worsen peak constraints unless accompanied by storage, dispatchable capacity or flexible demand.

9. South Africa and wider Africa: the allocation dilemma

The African case is where the ethical and economic trade-off is most visible. The latest SDG 7 assessment states that more than 560 million people in sub-Saharan Africa still lacked electricity access, and that the pace of electrification would need to accelerate dramatically to reach universal access. Even where households are connected, affordability and reliability remain separate problems.^[24]

It would nevertheless be a mistake to conclude that African states should reject data-centre investment until universal access is achieved. That would freeze the continent out of a strategic industry and ignore the possibility that large anchor loads can finance generation, fibre, transmission and skills. The stronger proposition is sequencing: where systems are constrained, a hyperscale project should demonstrate that it expands rather than merely reallocates scarce infrastructure.

South Africa is already confronting this issue. In May 2026 the South African Human Rights Commission sought public submissions on the human-rights implications of rapid data-centre expansion, explicitly identifying electricity demand, tariffs, water use, climate impacts, land, transparency and equitable access as issues for review. In August, five civil-society organisations submitted a joint report claiming that South Africa already hosted more than 60 known facilities with roughly 500 MW of disclosed capacity, while lacking a comprehensive public register or mandatory sector-wide disclosure of water and electricity use. They called for a public inquiry, mandatory disclosure, community-benefit obligations and consideration of a temporary pause on hyperscale approvals. These are advocacy claims and recommendations, not official findings, but the SAHRC process means the underlying governance issue is now institutional rather than rhetorical.^{[26][25]}

Water intensifies the distributional question. South Africa experiences recurrent municipal shortages and climate-related drought risk. In April 2026 the SAHRC separately initiated an inquiry into persistent water shortages in Gauteng, citing infrastructure failure and recurring service disruptions. A data centre can therefore be technically efficient yet politically

unacceptable if it receives reliable water and electricity while adjacent communities experience interruptions.^[27]

The appropriate policy test is not 'population first, data centres later'. It is whether the project delivers additional capacity and protects essential supply. A credible hyperscale approval in a constrained African market should therefore answer four questions before construction: What new generation or firming capacity is being added? What grid or distribution reinforcement is being financed? What water source and drought plan will be used? What measurable domestic benefit—tax revenue, skills, local compute access, research capacity or infrastructure—remains after the investment is operational?

The distributional question extends beyond household access. Communities can bear the substations, transmission corridors, diesel-backup emissions, noise, water withdrawals and land-use impacts that make hyperscale facilities possible even where tax revenue, ownership and digital value accrue elsewhere. Environmental justice therefore belongs inside the infrastructure balance sheet. The relevant comparison is not simply jobs versus megawatts, but durable local benefit versus the cumulative local burden of the entire supporting system.

This is the point at which infrastructure sovereignty becomes economically concrete. If the host state provides scarce electricity, water, land and tax concessions while ownership, high-value services and profits remain abroad, the transaction resembles an extractive infrastructure model. If, by contrast, the project finances new power and water systems, increases reliability for surrounding users, strengthens domestic cloud and research capacity and pays its full network costs, it can become developmental infrastructure.

10. Taiwan: when compute sovereignty meets energy security

Taiwan is an important case because it sits at the centre of the global AI hardware supply chain while operating an island power system with limited interconnection options. The growth of semiconductor fabrication and AI data centres therefore intersects directly with energy security. Taiwan cannot treat compute capacity as strategically optional, but neither can it assume that every large load has an automatic claim on the grid.

In late 2025 the Ministry of Economic Affairs amended subordinate energy regulations so that hyperscale and colocation data centres with contracted electricity capacity of at least 5 MW must submit Energy Utilisation Manuals at the planning stage and adopt Best Available Techniques. In July 2026 the ministry proposed adding an industrial-benefit assessment to that review so that data-centre projects are examined not only for technical efficiency and power feasibility but also for their contribution to domestic supply chains and industry.^[28]

This is a significant regulatory evolution. It moves the approval question from 'can the facility be powered?' towards 'is this the best use of scarce power capacity, and what strategic benefit does it deliver?' Taiwan's approach therefore provides a practical model for compute-sovereignty review in other constrained jurisdictions.

Concentration also creates a resilience issue. When a small number of very large campuses become material grid loads and are simultaneously embedded in cloud, semiconductor and critical-service ecosystems, their electrical and cyber systems become part of national infrastructure risk. Cyber compromise, cooling failure or abrupt load loss need not destabilise a grid in every case, but regulators should model such contingencies rather than assume that the facility boundary contains the consequence. The same logic applies to Ireland and large US clusters: compute concentration can become both an economic asset and a system dependency.

The principle should be applied carefully. Governments are poor at predicting which individual private investments will create the greatest long-term value. Industrial-benefit tests can become protectionist or politicised. The safeguard is transparency: criteria should be published, measurable and linked to genuinely scarce infrastructure rather than used as discretionary leverage.

11. A regulatory framework for the physical cost of AI

The comparison suggests that the regulatory problem is converging even though national systems differ. The most effective framework is not a single global cap on data-centre energy or water use. It is a set of conditions that become progressively stronger as local infrastructure becomes more constrained.

11.1 The Infrastructure Sovereignty Test

Test	Question
System adequacy	Can the electricity, water and communications systems serve the project under realistic peak, drought and contingency conditions without degrading essential services?
Additionality	Does the project cause genuinely new generation, storage, transmission, water recycling or other infrastructure to be built, or does it merely contract existing capacity?
Cost causation	Are network, generation, water and resilience costs caused by the project borne by the project and its customers rather than automatically socialised?
Climate and resource resilience	Can the facility and its supporting systems operate through extreme heat, drought,

	flood, wildfire, supply-chain disruption and plausible network contingencies?
Transparency	Are facility-level electricity, water, cooling, backup-generation, emissions and outage metrics disclosed in a comparable form, subject to legitimate security protections?
Local benefit	Does the host jurisdiction retain measurable and durable value through jobs, tax, skills, research capacity, local compute access, infrastructure or supply-chain development?
Distributional fairness	Are local burdens—water use, noise, backup emissions, land, substations and transmission corridors—proportionate to the durable local value retained? Where burdens are material, are there tangible offsets such as district-heating integration, local ratepayer-protection covenants, municipal utility support, water-reuse investment or other enforceable community benefits?
Resilience and strategic sovereignty	Does the project strengthen or weaken the state's ability to maintain critical digital and physical services during energy, cyber, climate, geopolitical or supply-chain disruption?

11.2 Additionality should be the central principle

Additionality is the strongest common principle because it separates productive investment from resource capture. A project that causes a new solar farm, nuclear uprate, battery installation, transmission line or recycled-water system to be built is economically different from a project that signs contracts for capacity that already serves households and existing industry.

Additionality should not be reduced to renewable certificates. The relevant unit is physical system capability. In a constrained grid, regulators should ask whether the project adds dependable megawatts at the location and hours when they are needed. In a water-stressed region, the equivalent question is whether it adds resilient non-potable supply,

recycling or drought-proof capacity. This is the logic behind 24/7 Carbon-Free Energy (CFE) and hourly matching: claims should increasingly be tested against when and where electricity is physically supplied, rather than against annual certificate balances alone.

The requirement should also be proportionate. A small enterprise data centre should not face the same obligations as a 500 MW hyperscale campus. Thresholds can be based on contracted import capacity, expected annual consumption and local system stress. Taiwan's 5 MW review threshold and Australia's discussion of standards for very large facilities show how a differentiated approach can work.

11.3 Access, reliability and affordability are separate

Infrastructure policy frequently treats connection as the objective. For AI-era planning, three distinct outcomes should be measured: access, reliability and affordability. A household connected to an unreliable grid does not receive the same economic service as a hyperscale facility backed by redundant substations, contractual reliability and diesel or battery backup. Regulators should therefore test whether new compute demand worsens outage risk, tariff pressure or fuel dependence for other users.

This is particularly important in emerging markets, but not unique to them. Ireland's grid constraints, Australia's transition and US transmission queues show that advanced economies also face scarcity. The difference is degree, not category.

11.4 Interdependence and cyber-physical resilience

Electricity, compute, communications and public infrastructure are becoming mutually dependent. Power systems enable AI and cloud services; those services increasingly support grid forecasting, logistics, water management and critical operations; fibre networks connect the control layer. A prolonged power interruption can disable compute, while a major compute or communications disruption can increasingly impair systems used to manage physical infrastructure. Regulation should therefore test circular dependencies, recovery times and failure modes rather than assessing each asset in isolation.

11.5 Compute nationalism, regulatory leakage and the international baseline

The physical cost of AI is also becoming geopolitical. States can influence rivals through controls on advanced chips, semiconductor equipment, energy technology, grid hardware and access to strategic infrastructure. At the same time, firms can arbitrage differences in environmental and disclosure rules by moving workloads or new campuses. A single global licensing regime is neither realistic nor desirable, but a minimum international reporting baseline would improve comparability and make carbon, water and infrastructure leakage visible.

12. Policy recommendations

12.1 Transparency and reporting

Mandate facility-level reporting for large data centres covering contracted and actual electricity demand, annual consumption, direct water withdrawal and consumption, water source, cooling technology, discharge characteristics where relevant, backup generation, emissions and major outages. Legitimate security information can be protected without accepting commercial opacity.

Require dual water accounting: direct site withdrawal/consumption and indirect water associated with electricity generation. Where facilities discharge process or cooling water, regulators should also assess temperature and water quality rather than volume alone.

Create a periodic national compute-infrastructure balance sheet tracking operational and planned MW, annual electricity use, water consumption, generation additions, network investment, domestic ownership, cloud concentration and major strategic dependencies.

Work towards a minimum international disclosure baseline. Comparable reporting should make carbon, water and infrastructure leakage visible even where substantive environmental standards remain national or local.

12.2 Grid integrity, additionality and flexibility

Apply an additionality requirement in constrained systems. Large new loads should present a credible package of new generation, storage, transmission, demand flexibility or water infrastructure proportionate to the capacity they require.

Adopt cost-causation rules. Where a project triggers identifiable network, generation or water-system upgrades, developers and anchor customers should bear those incremental costs unless an independent regulator determines that broader users receive commensurate benefits.

Reward temporal, spatial and computational flexibility. Non-urgent training and batch workloads should be encouraged to shift across hours or regions, or reduce intensity, when this lowers peak demand, renewable curtailment or water stress, provided cybersecurity, data-sovereignty and service requirements permit it. Demand flexibility is increasingly important to systems with high shares of variable renewables.[34] Regulators and utilities can operationalise this through dynamic tariffs, interruptible-load contracts, demand-response payments and pre-agreed curtailment arrangements that make flexibility financially valuable rather than merely aspirational.

Require cumulative infrastructure assessment for major clusters. Individual project approval is insufficient where several facilities depend on the same substation, transformer supply chain, aquifer, reservoir, fibre route or municipal system. For very large projects whose connection triggers long-lived public infrastructure, approval should also consider decommissioning bonds, minimum-use commitments or appropriately structured asset-

stranding indemnities so that ratepayers are not left carrying unrecovered costs if a proposed campus is cancelled, materially downsized or rapidly rendered unnecessary by changing compute economics.

12.3 Planning, resilience and community rights

Replace average-case siting tests with climate and resource resilience tests covering extreme heat, drought, flood, wildfire, cooling degradation, backup-fuel availability and plausible grid/fibre contingencies.

Treat community consent and environmental justice as project-risk variables from the beginning. Early disclosure should cover noise, water, backup-generation emissions, substations, transmission corridors, land use and cumulative impacts before design choices become locked in.

Link public incentives to public value. Tax concessions, accelerated approvals and public infrastructure support should be conditional on measurable outcomes such as additional power capacity, local research access, workforce development, heat reuse, water recycling or community investment.

Require large operators to maintain verified hardware lifecycle plans covering secure reuse, refurbishment, take-back and recycling of servers and cooling equipment. EU ecodesign rules already demonstrate that circularity can be incorporated into server policy.[33]

12.4 Industrial and geopolitical strategy

Integrate compute infrastructure into national security and industrial policy. Governments should map where strategic compute resides, who owns it, what power and network systems support it, and how quickly capacity could be restored after grid, cyber, climate or geopolitical disruption.

Apply the Infrastructure Sovereignty Test before allocating scarce public infrastructure or major incentives. Hyperscale projects should strengthen the host system rather than merely consume it.

Do not regulate through a single global efficiency number. PUE and WUE are useful operational metrics, but absolute resource use, scarcity-adjusted impacts, embodied infrastructure and system resilience also matter.

Treat dedicated nuclear and other firm generation as strategic options, not timetable assumptions. SMRs and advanced reactors may become important for compute, but first-of-a-kind licensing and deployment schedules remain materially longer and less certain than many data-centre build cycles.[31]

13. Limitations and areas for further research

The largest limitation is data quality. Public reporting often mixes planned, contracted and operational megawatts. Water disclosure is less standardised than electricity disclosure, and indirect water estimates depend heavily on assumptions about the electricity mix. Cross-country comparisons should therefore be interpreted as policy comparisons rather than precise rankings of environmental performance.

AI efficiency is also changing quickly. Improvements in chips, model architectures, quantisation, scheduling and cooling can reduce energy per unit of computation, while falling costs can simultaneously increase total demand through rebound effects. Forecasts to 2030 are therefore scenario-dependent rather than deterministic.

The economic benefits of data centres are also difficult to compare. Construction employment can be substantial while permanent on-site employment is relatively modest; however, cloud availability, lower latency, research access and digital sovereignty can generate benefits not captured by direct job counts. Future work should develop a common framework for measuring the domestic value retained per megawatt of strategic compute.

Finally, this paper does not attempt a full lifecycle assessment of AI hardware.

Semiconductor fabrication, mining, equipment manufacture, construction materials and e-waste add water, energy, carbon and material impacts outside the operational data-centre boundary. These supply-chain effects are material and have been brought forward in Section 3 only to define the wider physical stack; a full lifecycle quantification would require separate treatment.

Financial and stranded-asset risk also warrants further work. Generation, substations and water systems are long-lived investments, while model efficiency, workload location and AI demand can change rapidly. Overbuild is therefore possible, particularly where public infrastructure is sized around highly uncertain future campuses. Lenders, insurers and regulators will increasingly need to test utilisation, water availability, climate exposure and technology obsolescence together rather than treating data centres as conventional real estate.[35] The risk is not purely theoretical: public utilities may commit to substations, transmission or water capacity years before the associated compute load is fully operational, which strengthens the case for financial security mechanisms tied to project delivery and sustained utilisation.

A future research programme should also examine whether a multidimensional physical-cost dashboard can be standardised across jurisdictions. A single composite score would create arbitrary weighting problems and risk false precision; a better starting point is a common panel of electricity, firm-capacity, water, land, embodied-carbon, hardware-lifecycle and resilience indicators per clearly defined unit of service or facility capacity.

Orbital compute similarly remains a research frontier rather than a near-term substitute for terrestrial hyperscale training. Its governance would sit across fragmented regimes for

spectrum, launch licensing, debris mitigation and state responsibility under existing space law. Any eventual environmental comparison would need to account for launch emissions, replacement cadence, orbital congestion and end-of-life disposal as well as terrestrial water and land savings.[36]

14. Conclusion

Artificial intelligence is becoming one of the defining infrastructure demands of the 2020s. The relevant policy choice is not between embracing AI and protecting electricity or water. That framing is too crude. The real task is to design institutions that allow strategic compute to grow while forcing its physical costs into the investment decision rather than externalising them to households, existing industry or future public budgets.

The comparative evidence already points towards convergence. The EU is mandating disclosure. Ireland is conditioning grid access. China is coordinating compute geography with energy geography. Taiwan is reviewing large projects at the planning stage. Australia is testing social licence through parliamentary scrutiny and community resistance. South Africa is beginning to treat data-centre expansion as a human-rights and infrastructure question. In the United States, technology companies are increasingly financing the generation assets required to support their own growth.

These approaches differ politically, but they share an emerging recognition: compute is not weightless. It has a location, a power source, a cooling system, a water footprint, a grid connection, a hardware supply chain, a communications backbone and a set of opportunity costs. As those dependencies deepen, AI infrastructure also becomes part of national resilience and geopolitical competition.

The strongest policy principle is therefore straightforward. Where infrastructure is abundant, governments should enable growth. Where it is constrained, hyperscale compute should arrive with additional infrastructure, transparent resource accounting and a credible local-value proposition. Strategic AI capacity should strengthen the host system, not merely consume it.

The next era of AI governance will consequently extend far beyond algorithms. It will govern substations, reservoirs, reactors, transmission corridors, cooling systems, industrial incentives—and potentially orbital infrastructure. Artificial intelligence may remain digital in output, but its strategic competition is increasingly physical.

Bibliography

- International Energy Agency (IEA), Energy and AI (2025), especially sections on energy demand and energy supply for AI.
- Lawrence Berkeley National Laboratory, 2024 United States Data Center Energy Usage Report (December 2024).

- Lawrence Berkeley National Laboratory, United States Data Center Energy Usage Report: 2025 Update (June 2026).
- European Commission, Energy performance of data centres; Energy Efficiency Directive reporting framework and Delegated Regulation (EU) 2024/1364.
- Commission for Regulation of Utilities (Ireland), Decision on New Electricity Connection Policy for Data Centres (12 December 2025).
- National Development and Reform Commission (China), materials on 'East Data, West Computing' and green compute-power coordination, including July 2026 update.
- National Data Administration (China), green electricity and data-centre efficiency policy materials.
- Australian Senate Environment and Communications References Committee, Inquiry into Artificial Intelligence and Data Centres (2026).
- Climate Council of Australia, Submission to Senate Inquiry into AI and Data Centres (June 2026).
- South African Human Rights Commission, Call for Public Submissions on Data Centres and Human Rights in South Africa (22 May 2026).
- Open Secrets et al., Joint submission to the South African Human Rights Commission on data centres (August 2026).
- International Energy Agency and partner organisations, Tracking SDG 7 / electricity access data (2026).
- Ministry of Economic Affairs, Taiwan, Energy Administration materials on data-centre energy-utilisation review (2025–2026).
- Google, Project Suncatcher research updates (2025–2026).
- US Federal Communications Commission, Public Notice DA 26-113, SpaceX Orbital Data Centers application (4 February 2026).
- Shumba, N. et al., 'A Water Efficiency Dataset for African Data Centers', arXiv:2412.03716 (2024).
- US Energy Information Administration, Short-Term Energy Outlook, October 2026.
- Google / Constellation Energy, 890 MW nuclear uprate announcement, 6 October 2026.
- US Department of Energy, Office of Electricity, Supply Chain and Market Analysis; Large Power Transformer Resilience Report; 2026 grid-hardware materials.
- US Environmental Protection Agency, Cooling Water Intakes; Thermal Discharges in NPDES Permits; Water Reuse Case Study: Quincy, Washington.
- US Department of Energy / Nuclear Regulatory Commission, 2026 advanced-reactor and SMR deployment and licensing materials.
- US Department of Commerce, Bureau of Industry and Security, advanced semiconductor export-control materials.
- European Commission, Ecodesign Regulation (EU) 2019/424 and data-centre sustainability / circular-economy policy materials.
- International Energy Agency, Electricity 2026, chapter on flexibility and demand response.
- S&P Global, data-centre insurance and infrastructure-risk materials (2026).

- UNOOSA / NASA / FCC materials on space debris, spectrum and orbital infrastructure governance.

Publication note

This paper is an independent strategic-policy assessment by Ascendancy Advisors Limited. It is not legal, engineering, investment or environmental advice. Quantitative forecasts are scenario-dependent and should not be read as guarantees of future consumption, capacity or environmental impact. Where stakeholder submissions are cited, their claims are identified as such and should be distinguished from findings of regulators or public authorities.

Source notes

1. International Energy Agency, Energy and AI (2025), energy-demand and energy-supply chapters. <https://www.iea.org/reports/energy-and-ai/>
2. Lawrence Berkeley National Laboratory, United States Data Center Energy Usage Report: 2025 Update (June 2026). <https://datacenters.lbl.gov/publications/united-states-data-center-energy-2025>
3. Australian Senate Environment and Communications References Committee, Artificial intelligence and data centres inquiry (2026); Reuters, 28 September 2026, on the withdrawal of Goodman Project Mars.
4. International Energy Agency et al., Tracking SDG 7 update, 24 June 2026.
5. Taiwan Ministry of Economic Affairs, Energy Administration, data-centre measures, 16 December 2025 and 1 July 2026.
6. Google, Project Suncatcher update, 1 October 2026; US Federal Communications Commission Public Notice DA 26-113, 4 February 2026.
7. International Energy Agency, Energy and AI (2025), “Energy demand from AI”.
8. International Energy Agency, Energy and AI (2025), regional outlook.
9. Lawrence Berkeley National Laboratory, 2024 United States Data Center Energy Usage Report, pp. 56–58.
10. Noah Shumba et al., “A Water Efficiency Dataset for African Data Centers”, arXiv:2412.03716 (2024).
11. Lawrence Berkeley National Laboratory, United States Data Center Energy Usage Report: 2025 Update (June 2026).
12. US Energy Information Administration, Short-Term Energy Outlook, 6 October 2026.
13. Google and Constellation Energy, 890 MW nuclear-uprate announcements, 6 October 2026.
14. Lawrence Berkeley National Laboratory, 2024 United States Data Center Energy Usage Report, Figure 5.9 and pp. 56–58.
15. Google, “Our Project Suncatcher prototype satellite is in orbit”, 1 October 2026.
16. US Federal Communications Commission, DA 26-113, 4 February 2026.
17. National Development and Reform Commission and National Data Administration, national integrated computing network / “East Data, West Computing” policy materials (2024).
18. National Development and Reform Commission, green-compute update, 27 July 2026; National Data Administration, 18 March 2025.
19. European Commission, “Energy performance of data centres”; Delegated Regulation (EU) 2024/1364.
20. Commission for Regulation of Utilities (Ireland), new electricity connection policy for data centres, 12 December 2025.
21. Australian Senate Environment and Communications References Committee, Artificial intelligence and data centres inquiry (2026).
22. Climate Council, Submission to Senate Inquiry into AI and Data Centres, 9 June 2026.
23. Reuters, “Australian developer axes Sydney suburban data centre plan after local opposition”, 28 September 2026.
24. International Energy Agency et al., Tracking SDG 7 update, 24 June 2026; IEA, Financing Electricity Access in Africa (2025).
25. South African Human Rights Commission, call for submissions on data centres and human rights, 22 May 2026.
26. Open Secrets et al., joint SAHRC submission / press release, August 2026. Figures cited are stakeholder estimates.
27. South African Human Rights Commission, Gauteng water-crisis inquiry, 13 April 2026.
28. Taiwan Ministry of Economic Affairs, Energy Administration, data-centre energy review measures, 2025–2026.

29. US Department of Energy, Office of Electricity, "Supply Chain and Market Analysis" (2026) and Large Power Transformer Resilience Report (2024); DOE 2026 materials report distribution-transformer lead times of 12–30 months in the latest underlying data and large-transformer lead times that can extend to several years. <https://www.energy.gov/oe/supply-chain-and-market-analysis>
30. US Environmental Protection Agency, "Cooling Water Intakes" and "Thermal Discharges in NPDES Permits"; EPA, Water Reuse Case Study: Quincy, Washington. <https://www.epa.gov/cooling-water-intakes> ; <https://www.epa.gov/waterreuse/water-reuse-case-study-quincy-washington>
31. US Department of Energy, "Energy Department Awards \$94 Million... Small Modular Reactors," 14 May 2026; US Nuclear Regulatory Commission, 2026 advanced-reactor licensing materials. DOE describes supported Gen III+ SMR deployments as contributing to new nuclear generation in the 2030s. <https://www.energy.gov/articles/energy-department-awards-94-million-american-companies-help-expedite-deployments-small>
32. US Department of Commerce, Bureau of Industry and Security, 2 December 2024 advanced-semiconductor export-control package and related controls on AI/advanced-computing technology. <https://www.bis.gov/press-release/commerce-strengthens-export-controls-restrict-chinas-capability-produce-advanced-semiconductors-military>
33. European Commission, "Servers and Data Storage Products" (Ecodesign Regulation (EU) 2019/424) and "Energy performance of data centres." Circular-economy provisions include component extraction, critical-material information, secure data deletion and firmware availability. https://energy-efficient-products.ec.europa.eu/product-list/servers-and-data-storage-products_en
34. International Energy Agency, Electricity 2026, "Flexibility": demand-side flexibility becomes increasingly important as variable renewable penetration rises. <https://www.iea.org/reports/electricity-2026/flexibility>
35. S&P Global, "Underwriting the Digital Backbone: Navigating Risk and Growth in Data Center Insurance" (2026), identifying location/aggregation, supply-chain, power infrastructure, project-finance and credit risks. <https://www.spglobal.com/energy/en/events/webinars/underwriting-the-digital-backbone-navigating-risk-and-growth-in-data-center-insurance>
36. US FCC materials on orbital data-centre applications; NASA/UNOOSA materials on orbital debris and space sustainability. This paper treats orbital compute as an emerging governance question rather than an established hyperscale alternative.