

# Time-to-Power: The Hidden Economics of AI Infrastructure

The Economy Research Editorial<sup>1,2</sup>

<sup>1</sup>The Economy Research, 71 Lower Baggot Street, Dublin 2, Co. Dublin, D02 P593, Ireland

<sup>2</sup>Swiss Institute of Artificial Intelligence, Chaltenbodenstrasse 26, 8834 Schindellegi, Schwyz, Switzerland

## Abstract

This article argues that in the global AI infrastructure race, the decisive economic variable is not simply the price of electricity, chips, or tax advantages. Instead, the decisive variable is time-to-power: how quickly capital, land, servers and electricity can be converted into working compute. Using a 100 MW AI data-center model, the article shows how a small one-year delay can have a bigger negative impact on the lifecycle value of a given project than large variations in power or tax variables. The article connects project finance to the national competitiveness argument and explains how efficiencies of grid-queue, permit and energy constraint management are now zero-level factors in determining where in the world AI capability expands first. As data-center electricity demand and frontier-model training become ever more capital-intensive, countries that can compute faster will capture earlier revenue, enable faster iteration cycles and build stronger local innovation ecosystems. The essential policy takeaway is fairly direct: AI deployment speed must be a central strategic priority for national AI policy.

# 1 Introduction - Current Investments in AI Compute Infrastructure

An unprecedented global surge of investment is underway to establish the compute backbone necessary for cutting-edge AI.<sup>[1]</sup> In 2026 America's largest technology companies are expected to spend roughly \$670 billion, equivalent to about 2 percent of U.S. GDP, on compute clusters and AI-capable data center infrastructure; over the same period, the combined efforts of institutions around the world will total nearly \$1 trillion.<sup>[2]</sup> Such investment, one study suggests, surpasses other peacetime efforts in US history. This investment surge underscores a recognition that high-end computing capacity has accrued a new dimension of strategic importance. More than two years of data show that while the United States accounts for nearly 75 percent of the world's high-end computing clusters, other countries are moving to reduce the gap.<sup>[3]</sup> Strained grids, slow permitting and local opposition to new data centers are tightening the pipeline of domestic capacity, even as other countries are throwing massive institutional support behind data-center development. China's government has infused massive funding into achieving chip dominance and establishing new data centers and Middle Eastern energy exporters such as Saudi Arabia and the United Arab Emirates are using abundant power and capital to attract AI infrastructure projects.<sup>[4]</sup> Many US allies are at risk of falling behind; the largest European AI clusters in terms of compute footprint have less power than a single hyperscale campus in the American midwestern states and a number of countries, such as Australia, South Korea and Italy, currently operate virtually no AI data centers.<sup>[5]</sup>

All of these trends have far-reaching consequences. Countries that host large-scale compute infrastructure will capture not only economic returns but also influence over the direction, accessibility and governance of advanced AI systems. Since AI is essentially dual-use, with civilian and military applications, democracies with a large block of computing power have an opportunity to direct safety research, accountability and liberal values into the future of AI technology, whereas authoritarian regimes have the potential to pursue the development of AI under conditions of secrecy, repression and increased risk of malevolent use and catastrophic failure.<sup>[6]</sup> Compute infrastructure in this regard is just as geopolitically critical as oil or steel was in the last century, as one recent analysis explains: if the 20th century ran on oil and steel, the 21st century runs on compute.<sup>[7]</sup>

However, despite the stakes, many existing narratives of AI still concentrate on algorithms, data, or employment impacts and treat computation as a secondary or neutral factor. This familiar narrative misses the largest constraint: AI development is entering a physical bottleneck of hardware, energy and grid capacity.<sup>[8]</sup> Recent research suggests that the AI boom is not merely a software phenomenon but an infrastructure shock dictated by electricity availability, location planning of data centers, chip manufacturing and interstate competition for computing power.<sup>[9]</sup> AI workloads are already using high proportions of the world's electricity: in 2023, U.S. data centers consumed around 176 TWh, equivalent to roughly 4.4% of national electricity use and this could double or more by 2028.<sup>[10]</sup> In Ireland, power for data centres rose from 5% of overall electricity to 21% within a few years, causing grid emergencies and regulations.<sup>[11]</sup> These pressures – on land use, water for cooling and supply chains of chips - mean decisions over AI compute involve second-order trade-offs far beyond those of a typical productivity gain.

This article advances a corrective analytical framing: AI is not just another piece of software, compute is not

just another commodity. The central claim is that time-to-power, defined as the speed at which a planned AI data center comes online, is the hidden financial leverage in determining who wins and who loses in the global AI race.<sup>[12]</sup> Standard policy recommendations on energy policy have called for putting a cap on the cost of energy, or offering tax holidays for data centers; against this, the paper shows that reducing delays and ramping up project speed generate much higher financial gains. Recent evidence from 2023 to 2026 suggests that grid bottlenecks are becoming more binding, debates on energy policy are heating up and international competition for digital capacities is thickening. As advanced large language models get increasingly expensive to train, the value of arriving first is rising. The article's central finding is that AI infrastructure competitiveness depends less on marginal cost differences than on the speed with which capital, electricity, chips and land are converted into working compute.

The article develops this argument in three steps. The article first describes a prototype financial model used to compare the economics of a 100 MW AI data center across ten countries.<sup>[13]</sup> It then introduces the model's quantitative takeaways and why time delays trump all other cost factors. From there, the analysis then turns to a broader treatment of how computing availability not only influences innovative incentives but also broader geopolitical repercussions. In all, the quantitative findings are then connected to the policy question of which corporate and governmental reforms would best shape and preserve democratic influence over an AI future increasingly shaped by compute access? In short, the article concludes by showing that the eventual crux of the matter when it comes to AI is the time dimension of computing and democracies must act on this constraint before they are left behind.

## 2 Financial Model – Time to Power: Why Earlier Is Far Better

To ground these arguments, the article uses an abstracted financial model for a large AI data center. The case study is a 100 MW facility roughly equivalent to a hyperscale deployment, with a scale on par with 100,000-150,000 high-end GPU boxes and tens of thousands of servers.<sup>[14]</sup> This AI cluster is treated as the core unit of analysis, using industry-based assumptions and a series of expert interviews to determine case study cost, revenue and time-to-market benchmarks. This model is then applied to cover 10 different countries (the US, UAE, Finland, Canada, India, France, South Korea, Australia, the UK and Germany), chosen to exemplify a wide range of different regulatory and electricity grid environments. For each country, the model assumes 12 years of operation (the typical useful life of IT in cloud data centers), an average annual revenue-per-MW based on existing calculations of the GPU-hour market rate (roughly \$200 million per month at full capacity) and a set of different assumptions for capital expenditure (building, cooling and power infrastructure) and operating expense (especially, power prices).<sup>[15]</sup> As a key difference between scenarios, the model considers in each case the impact of "time-to-power" (the period that begins at project approval and ends with chips running), to isolate the effect delays have on rollout.

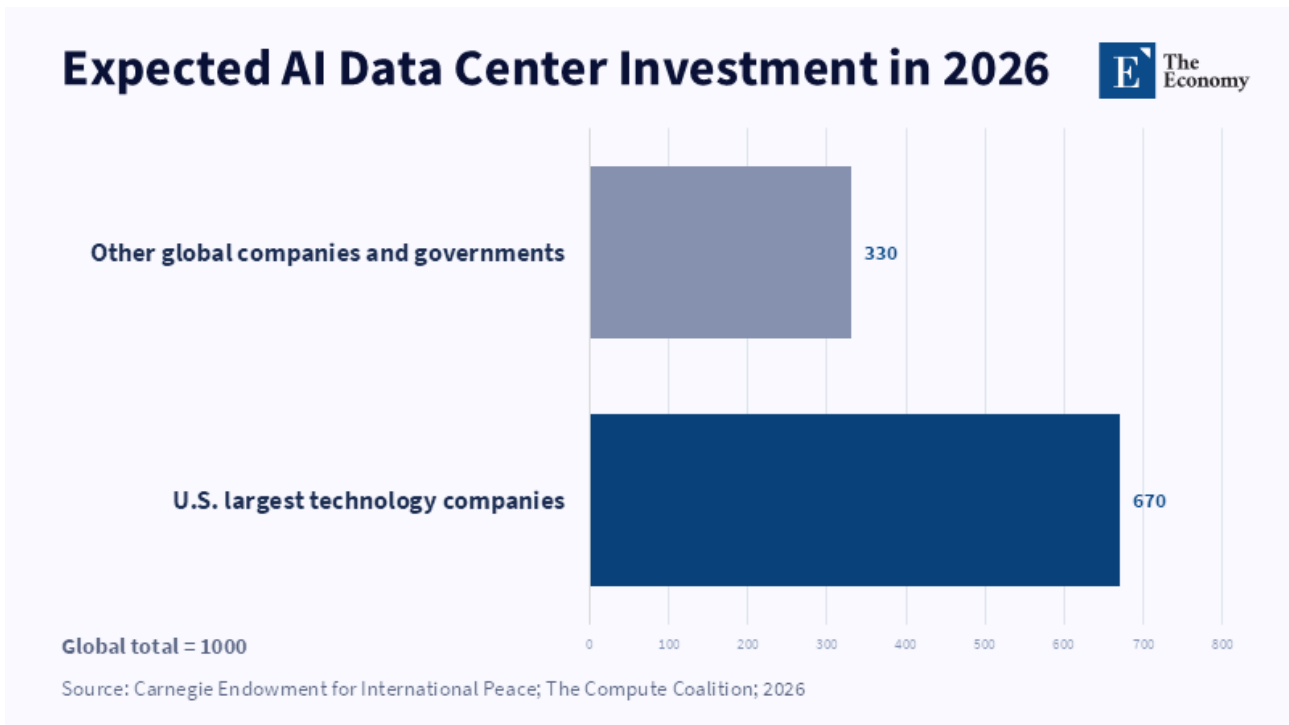


Figure 1: AI infrastructure investment has reached a scale that makes compute timing a macroeconomic variable.

Because any model of this kind depends on contestable assumptions, the figures should be treated as estimates of direction and magnitude rather than precise forecasts. Several benchmarks from published sources lend perspective on magnitude: one industry breakdown pegs upfront costs for construction and IT equipment in a 100 MW AI data center at roughly \$3.4-5.5 billion of today’s dollars.<sup>[16]</sup> Of that, the physical construction and utility infrastructure might account for less than \$1.5 billion (approximately \$9-15 million per MW) while the thousands of specialized graphics processing units and servers drive another \$2.5-4+ billion higher. Power use is equally daunting: at 100 MW constant power, they would use roughly 876,000 MWh (876 GWh) per year.<sup>[17]</sup> Given a common industrial rate (say 0.05-0.10/kWh), the annual energy bill would run around \$40-80 million, a sizable sum in absolute terms. However, energy costs are a secondary concern given projected revenues and capital costs. With these rough figures, a standard US project might contain on the order of \$10 billion in discounted lifetime value (revenue minus costs) in the baseline case.<sup>[18]</sup>

Afterward, the model considers counterfactuals: what if the 100 MW cluster were postponed by 3, 6, 12, or 18 months? What if electricity costs increased by 50% or 100%? What if a moderate tariff were placed on imported AI servers? What if some tax subsidies were switched off? Each potential scenario is then compared to the base case to estimate the net present value (NPV), showing which variable erodes value most sharply. The most important result is also the most policy-relevant: delay erodes value more sharply than the other variables tested: the timing to power. This proved to be by far the most effective mechanism to erode value among all tested factors.

To give a concrete example, consider the U.S. case. The Carnegie figure should be used here to show the relative lifecycle-value impact of delay, electricity prices, tax incentives and server tariffs (an adaptation of Carnegie’s model) and keeping the default case of full operation at time zero, which should net about \$10 billion NPV. Moving to an entire year of delay, moving every revenue and cost stream 12 months to the future,

removes about \$550 million, over 5% of the entire venture. By comparison, doubling the assumed electricity price would reduce NPV by about \$441 million; removing state-level tax incentives, including bonus depreciation and related incentives, would cost about \$338 million; a moderate tariff on imported servers would cost about \$172 million; and even a three-month delay would reduce value by about \$178 million.<sup>[19]</sup> In other words, in the U.S. context, each additional year of latency destroys more value than doubling the electricity bill.<sup>[20]</sup>

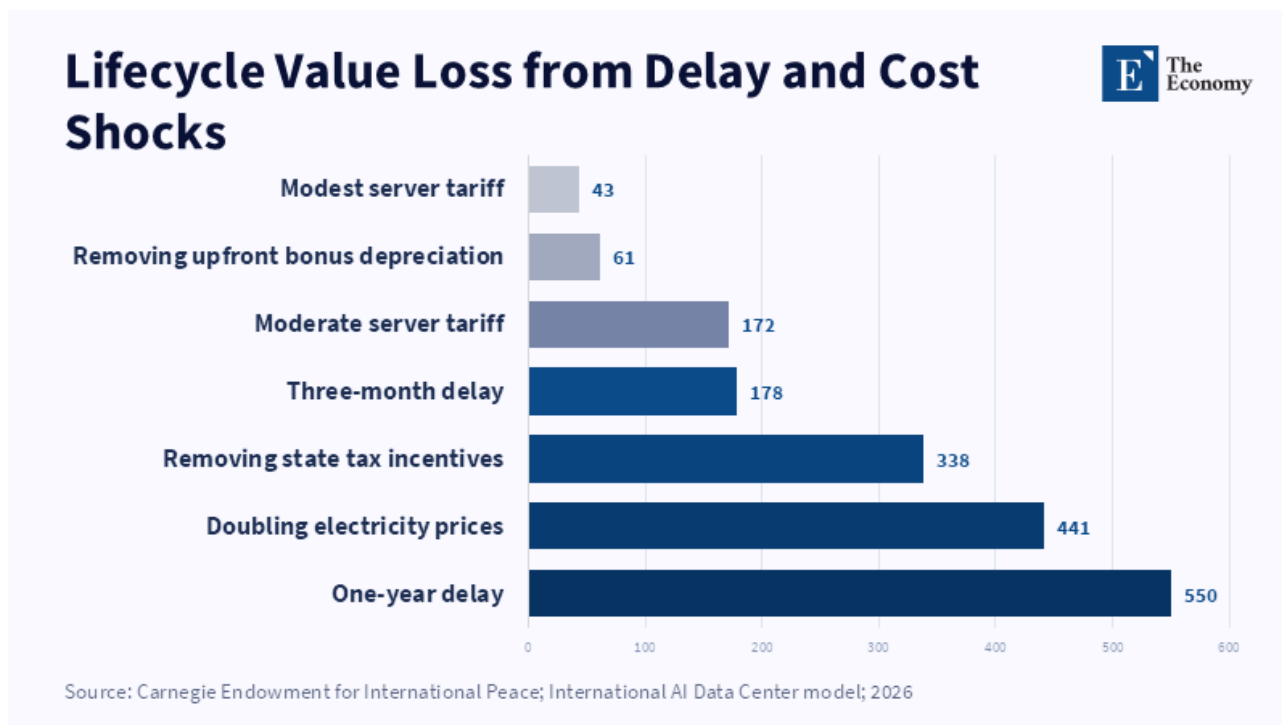


Figure 2: Delay is the largest value destroyer in the model, exceeding power-price, tax-incentive and tariff shocks.

The reason is basic finance: earlier flows are more valuable than cash flows pushed several years into the future. A one-year delay postpones hundreds of millions of dollars in GPU-hour gross revenue.<sup>[21]</sup> With existing GPU prices and utilization, a 100 MW Blackwell data center operating at parallel efficiency can generate \$200 million per month gross revenue. A twelve-month delay postpones roughly \$2.4 billion in gross revenue before costs.<sup>[22]</sup> Nearly any reasonable discount rate makes this decline in regular earnings well worth watching for. In practice, most chip inventory arrives near the end of the build, which limits the period during which the balance sheet is exposed to idle chip inventory.<sup>[23]</sup> Even so, some capital remains tied up in construction, land and financing commitments, and financing terms may tighten if market conditions shift during a prolonged delay

The effect compounds in yet another measure. Faster deployment also creates a compounding research-and-development advantage.<sup>[24]</sup> Each month earlier online is a multiplying factor in the number of GPU-hours spent training the future waves of models; developing innovations for new uses and applications. The drive toward frontier AI is self-sustaining: the better the models, the better the trade press will be at suggesting new applications, attracting more customers, extracting more data and so on, with accelerating feedback. A lagging project, by contrast, is ceding ground. One of the more poetic passages of the Carnegie report states that time swamps every other factor.<sup>[25]</sup> For the world’s largest AI companies, for which a single month can translate into

billions in strategic competitive benefit, being an early mover is equally meaningful.

The headline conclusion from this model turns out to be that time-to-power is the key economic factor. Across markets, the project's average rank in project competitiveness closely follows build time.<sup>[26]</sup> Even for the U.S., the benefits of the 4-month speed-up in approval and connection outstrip the playing field of the 4% power-price subsidies or 4-point tax cuts.<sup>[27]</sup> Both the chart and model results (see above) prove this: delays are steeper as the electricity or cost parameters. This single insight blows away the conventional wisdom: the obvious focus on low power prices will be offset by the unseen priority: speed.

The evidence is fairly robust to alternative assumptions as well. Sensitivity checks using different discount rates and revenue-growth assumptions do not change the hierarchy of results. At a corporate discount rate of about 8-10%, shifting profits by a year (say shifting a 12-year stream by a year at 10% discount) typically reduces NPV by roughly 5-9%, depending on the discount rate (on the order of 5-9%) (roughly a 9.1% drop at 10% discount for a 1-year shift of the 12-year stream; roughly a 5.5% drop for 6% discount; close in size to the estimated loss described above). Similarly, the revenue assumption about revenue scaling at \$200 million per month can be adjusted a bit without changing the core relationship: delay remains the most damaging variable. Additional checks by changing other assumptions like capital intensity (more or fewer GPUs per MW) and the bottom line remains: the penalty for lost time (weeks, not months) is so large that only a fundamental change like reducing a mid-range delay to weeks rather than months can materially change competitiveness.

A simple back-of-the-envelope calculation on these results as well, to confirm that time really is the most important factor here.<sup>[28]</sup> If a 100 MW AI data center has a discounted lifecycle value of roughly \$10 billion, delaying the revenue stream by one year reduces that value by approximately  $r/(1+r)$ , where  $r$  is the discount rate. At a 6 percent real discount rate, the loss is about 5.7 percent, or roughly \$570 million. At 10 percent, it is about 9.1 percent, or roughly \$910 million. Even allowing for model-specific timing effects, this calculation shows why delay can outweigh much larger-looking changes in electricity prices.

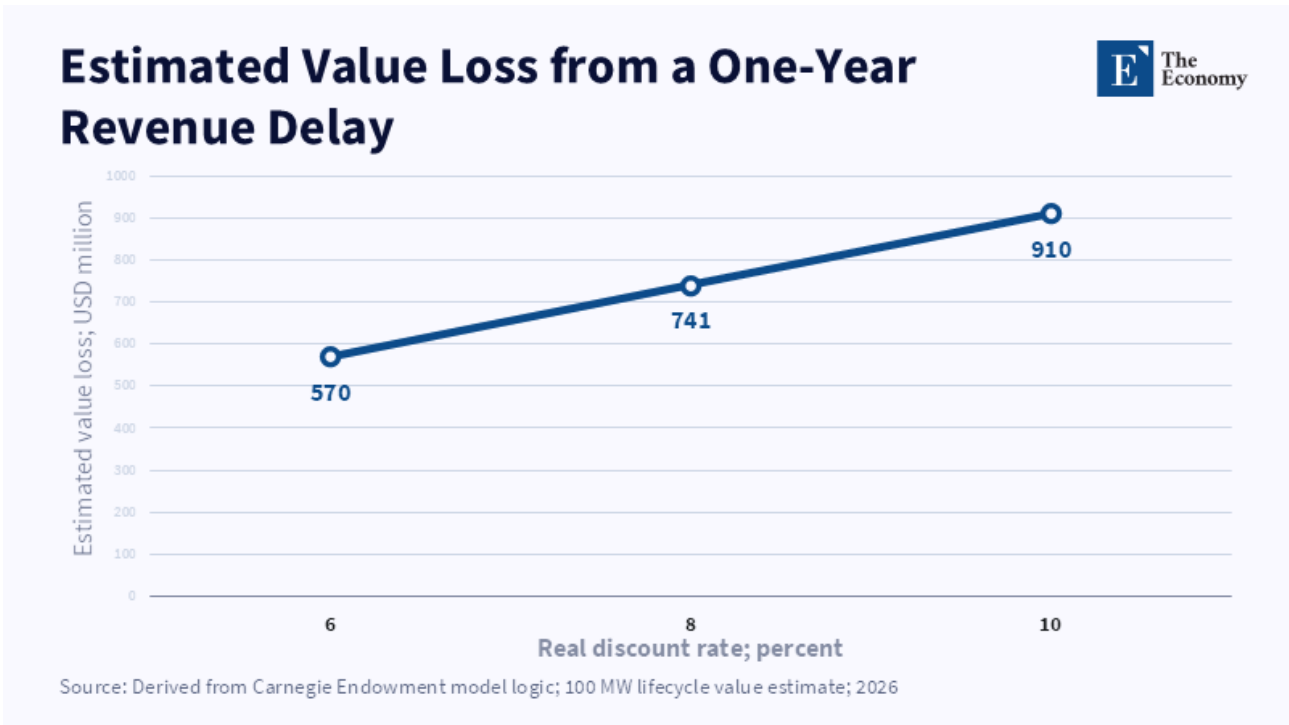


Figure 3: The discount-rate logic shows why delayed revenue can erase hundreds of millions in lifecycle value.

But there are also wider, national effects to this analysis. Nations capable of bringing their cluster online in 2 years instead of 3 or 4 will have a clear, durable advantage in the race to build AI capabilities. They will have more experiments, iterate quickly and bring AI products to market sooner. Conversely, permitting delays, grid-connection queues and equipment shortages may seem insignificant relative to a \$3-5 B budget, but they function to a massive, multi-fold drain relative to a classic subsidy.

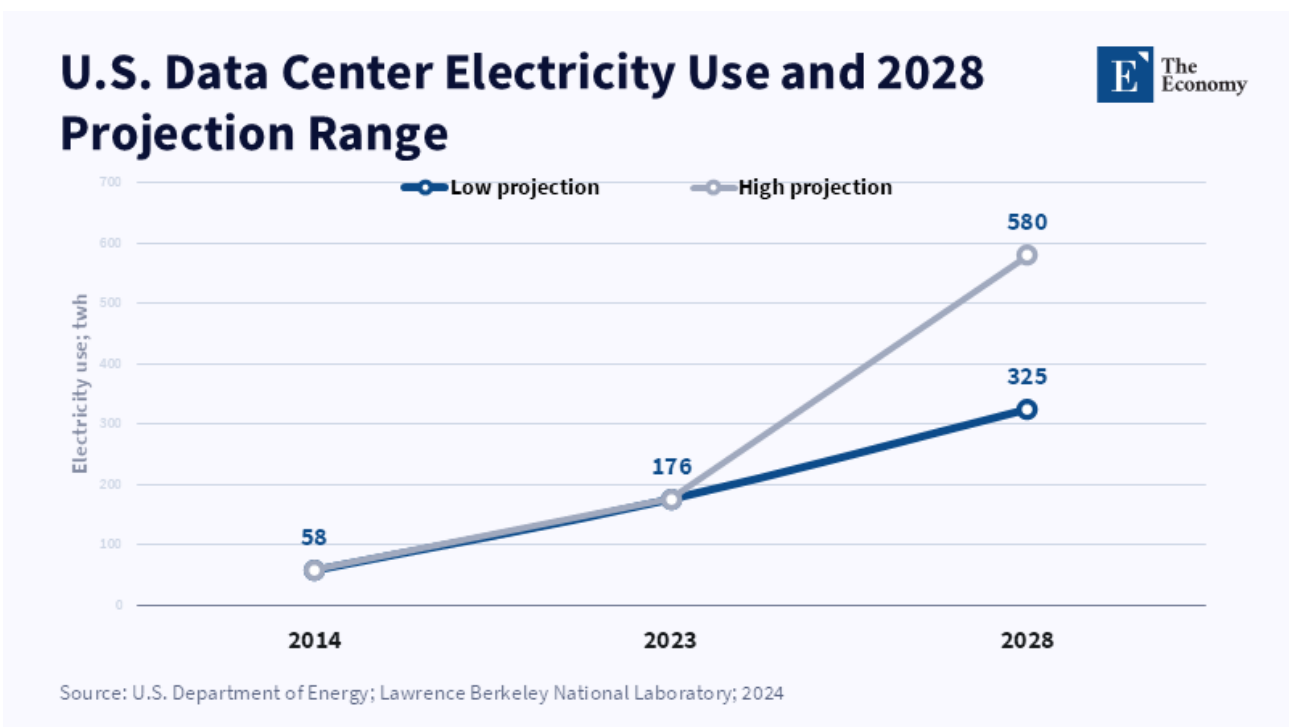


Figure 4: Rising U.S. data-center electricity demand makes grid connection speed central to AI competitiveness.

In sum, the exercise demonstrates a fundamental strategic conclusion: the priority should be to get to

compute as soon as possible, by far. Even if this costs additional, substantial power bills or tariffs, the answer should still be to make this your priority, even if that means beating the competition to market by a few months. This logic needs to be adopted in judging projects and allocating policy time and money: the dominant criterion should be time to deliver compute, not the immediate electricity price. And the single message it delivers to policy makers and investors: any reform that shortens the timeline by even a few months will have a life-cycle economic impact at least as great as large power-cost subsidies.

### **3 AI Infrastructure Economics – Time-to-Power as the Hidden Financial Variable Behind the Data Center Race**

Synthesizing the above, the realization that time-to-power overrides project returns requires a wider reinterpretation of the economics of AI. It implies a dynamic focus away from marginal costs alone; to what does it translate in the longer term at the country, firm and societal levels?

Now, how does this timeline impact firms, countries and innovation ecosystems. For a tech company, a delay in data center availability might just mean some additional interest payments or leasing capacity for interim periods. In a competition for corporate profits that might just be a redistribution to new investments (cloud credits, slower growth), a manageable cost. But for an entire country, the aggregate effect is different: a delay in domestic compute infrastructure translates into a slower overall AI development cycle that slows the wider AI innovation cycle. A slower rate of AI innovation means perhaps fewer local servers available to rent for training, higher training costs and delayed launches of new products and services. This could, over time, lead to a talent drain as AI startups and university-based researchers move abroad to countries with better infrastructure. Simply put, a lagging country in infrastructure cedes parts of its AI ecosystem to other countries that are moving at a faster rate.

Suppose further that the two cooperating nations, A and B, have the same available pool of tech workers and R&D capacity. Suppose further that Country A's regulators effectively fast-tracked permits and its grid operators organized the early build-out of the energy infrastructure, providing rapid grid connection and new data centers opened in 24 months. Country B's regulators pushed through the slower approval processes, faced grid resistance and the data centers only received power in 36-48 months. What does this mean? Over three years' time, firms in Country A may have been able to deploy two generations of large language models. These models trained twice will be more capable earlier. The more capable early models from A will be able to offer better services and tooling earlier. Users, firms, even military planners in B will benefit from foreign-hosted models, meaning that firms, citizens and public institutions in Country B may rely on systems developed in Country A or somewhere that A dominates. This transfers value-added gain (subscription fees, advertising, transactions) to A while giving individuals and firms a worse-off informational footing. If the population of B is dependent on foreign hosting of AI models, they are dependent upon the design choices, privacy policies and externally imposed data access policies of the foreign host.

Some analysts have started to cast this as a compute sovereignty question. Having local (domestic) AI

capacity grants a state more control over a strategic technology.<sup>[29]</sup> It also entails trade-offs. On a foundational level, hosting massive data centers ensures a local source of compute power (like local energy production), which can spark local innovation. But it also means the thing powering that computer comes from somewhere else; those compute-intensive data centers need large amounts of electricity (which requires water to cool and uses up energy resources in the process) and those chips need rare metals that are mined elsewhere. As recent sovereign-compute research suggests, countries that open up to hosting data centers improve their access to key computational resources, but also the region’s demand for electric power and water, affecting other public priorities.<sup>[30]</sup> Practically, hosting large AI data centers can impose local infrastructure costs in potentially growing electric rates and the need to develop additional infrastructure; those costs are often passed on to other taxpayers or consumers.

The notion of computing race can be viewed as an analogy to that of natural resources from the Geopolitical perspective in the sense that a country or a coordinate bloc that controls a significant amount of global AI infrastructure might be able to hold the same kind of driving power as the members of the Organization of the Petroleum Exporting Countries (OPEC) currently can in today’s energy economy, as one of the comments suggests. Even a relatively lower percentage share of global compute capacity and control would present a source of leverage. This helps explain why China is investing heavily in chips and data centers to reduce dependence on U.S. tools and why slower U.S. or allied buildout could leave dependent states exposed to the pricing and access decisions of foreign providers.<sup>[31]</sup>

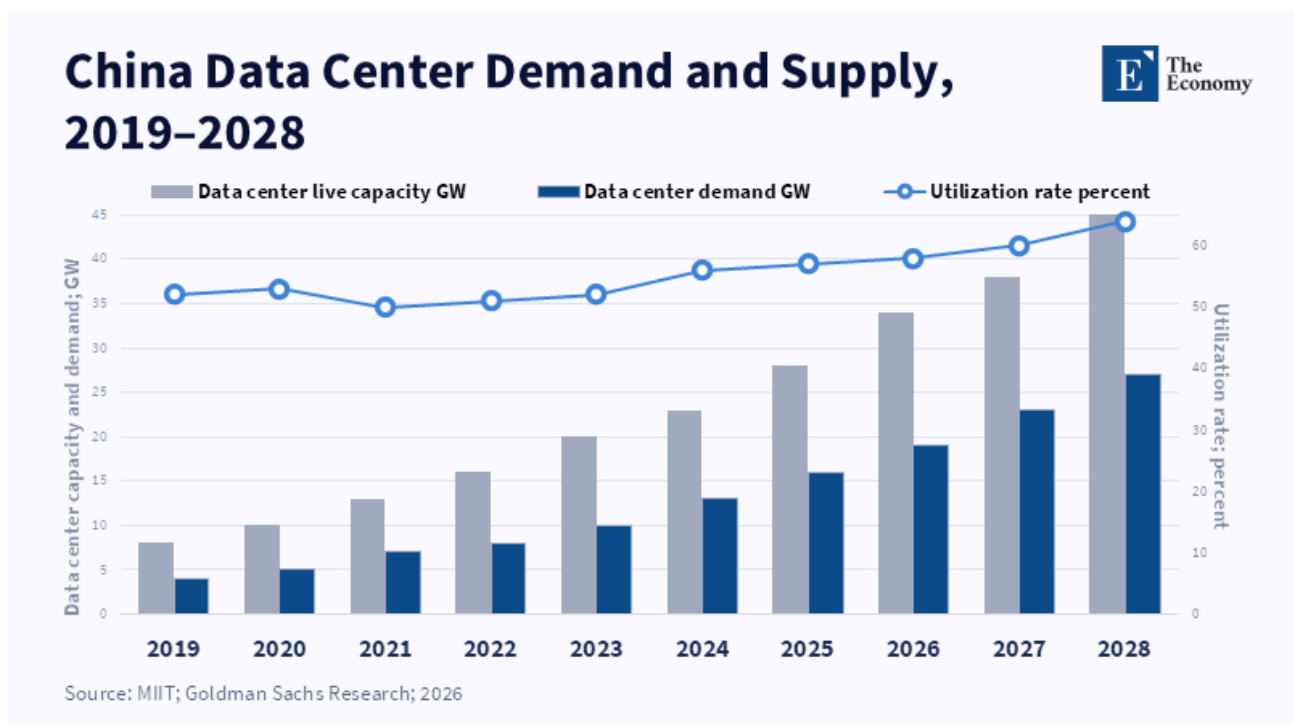


Figure 5: China’s rising utilization shows how quickly demand can absorb new capacity, turning buildout speed into strategic leverage.

The same logic is relevant for the inequality and labor trends. Calls for rapid adoption of AI tools have led to comparisons between generative AI and calculators or search engine tools that merely increase individual productivity. But history indicates this paradigm is incomplete. Early calculators did increase productivity, but

they redefined education and skill development pathways, with returns going to those better at math and logic-a pattern that may be amplified at a scale far larger than before. Recent research highlights that widespread deployment of advanced AI systems relies on the availability of hardware, much more than mere interface. If some parts of the world lack the necessary hardware, firms and students may be slower to access the productivity improvements from AI, generating new global technology rifts. Existing OECD research also shows AI is rapidly evolving in the trend of electricity consumption, likely concentrated near centers of AI development and local environments lacking the robust electrical infrastructure may decline to host new installations altogether. A cautionary tale from Europe comes to mind: recent grid-straining waves of AI prompt an urgent freeze on new data center projects in Ireland, due not to AI technology failure but to ill-equipped access to electricity. Similar issues could arise elsewhere, with ad hoc AI development exacerbating existing inequalities unless coordinated guidelines exist.

Evidently, then, investing in AI infrastructure could magnify existing inequality, if unmanaged. Those positioned to build fastest (so far, the U.S. and Gulf states, subsequently the utility-friendly Nordic countries and Canada) will extend their lead, leaving the other laggards even further behind. Democracies need not fear such clouds cast long shadows even in terms of high-rank inequality: a country like Germany or the UK with a 3.5-year cluster roll-out period produces a research industry and entrepreneurs that function at a disadvantage compared to those whose nations need only two years, as in the United States or the UAE. A shift of only one year can alter competitive rankings; the Carnegie model demonstrates that the United States drops from number one to number five if an extra year's holdup occurs.<sup>[32]</sup>

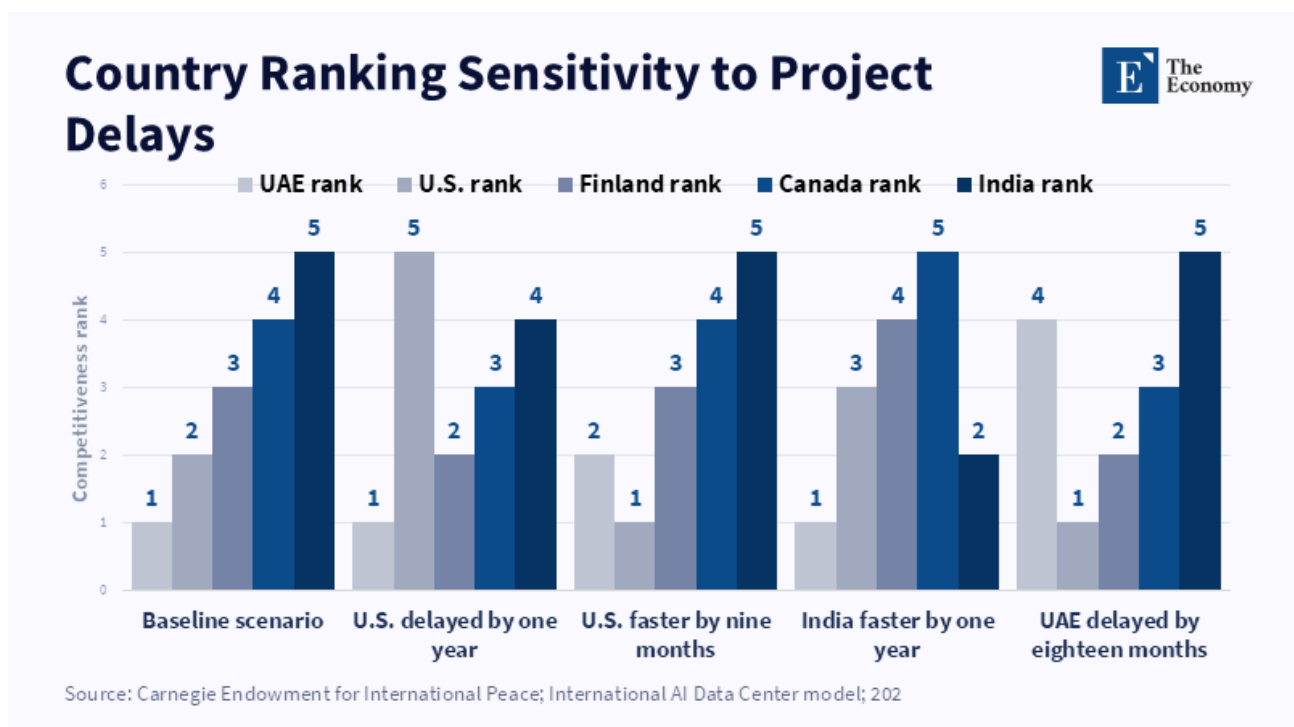


Figure 6: A one-year delay can materially reshape country-level competitiveness in AI infrastructure deployment.

A common counterargument is that AI will simply be a productivity boost or an information equalizer, much like earlier waves of technology. Perhaps AI could indeed democratize access to certain types of knowledge at the user level (some chatbots are widely available) and there may be cases where firms have use AI to increase

output. But this view neglects the systemic second-order effects. No prior tech boom has relied on a rare, geographically fixed asset so much as the fundamental fuel of AI-compute. Unlike calculators or search, training a new state-of-the-art AI model is so compute- and energy-heavy that only a handful of actors could afford it.<sup>[33]</sup> And while inference may one day be inexpensive, even planning for research is a significant first stage that is asset-intensive, points out one expert, which will limit progress until third-party providers scale up. The point is that although consumer prices will likely decrease, the wide adoption of AI relies on physical factors that are precious and difficult to acquire (semiconductors, electricity, datacenters). In essence, technological progress will be limited by infrastructure.

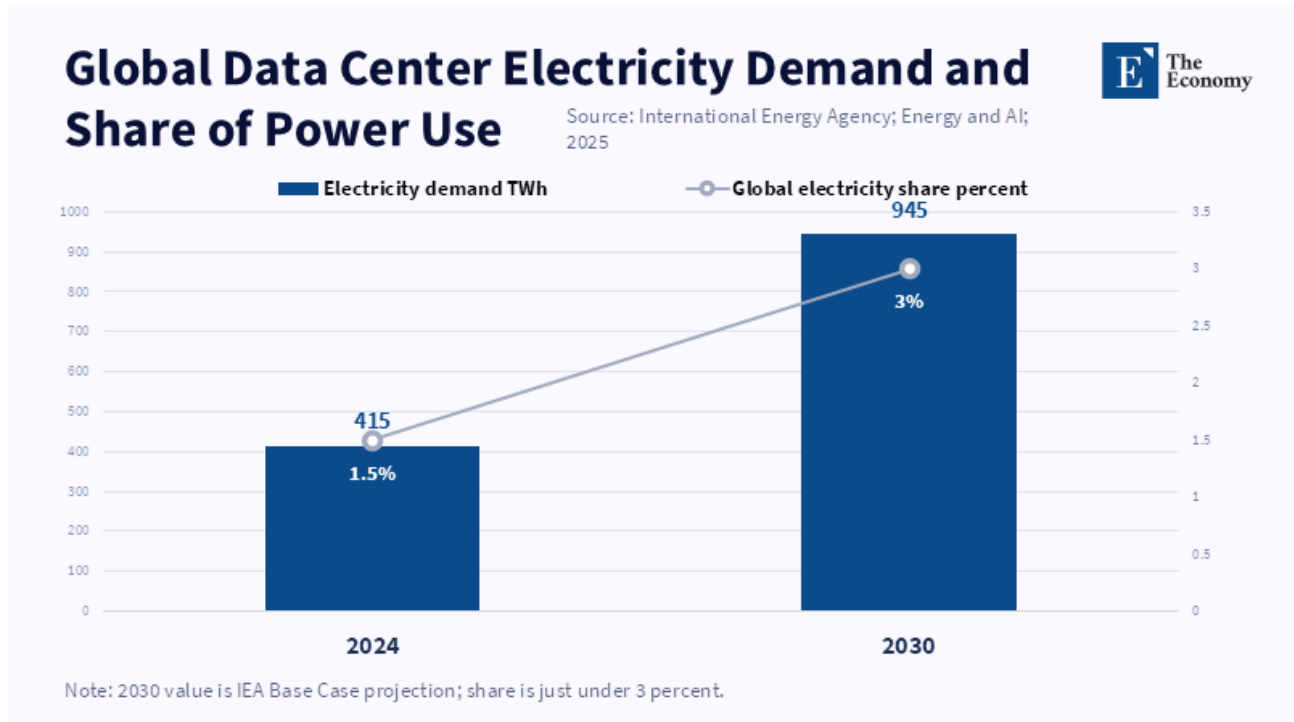


Figure 7: Global data-center electricity demand is projected to rise sharply, reinforcing the link between AI growth and power availability.

Others might also contend that the market will take care of this: more demand and profit will generate more investment (as data centers sprang up in Virginia or Ireland, in earlier cloud cycles). There is an aspect of truth here, indeed, one can see the market response in the rapidly rising capex levels for the four largest cloud providers (more than \$200 billion in 2024). The evidence indicates that markets by themselves are much slower and less geographically balanced than expected. Private incentives push firms toward those regions where they get the quickest return, which may be an obstacle to service coverage of a very large part of the globe. The results suggest that policy coordination is needed even at the market-design level: a handful of commentators advocate creating an international compute coalition of democracies that would adopt standard arrangements and fast-tracking rules, so that allied states collectively can run enough AI infrastructure to remain at pace. Without coordination, the power and queuing physics may give rise to a digital divide.<sup>[34]</sup>

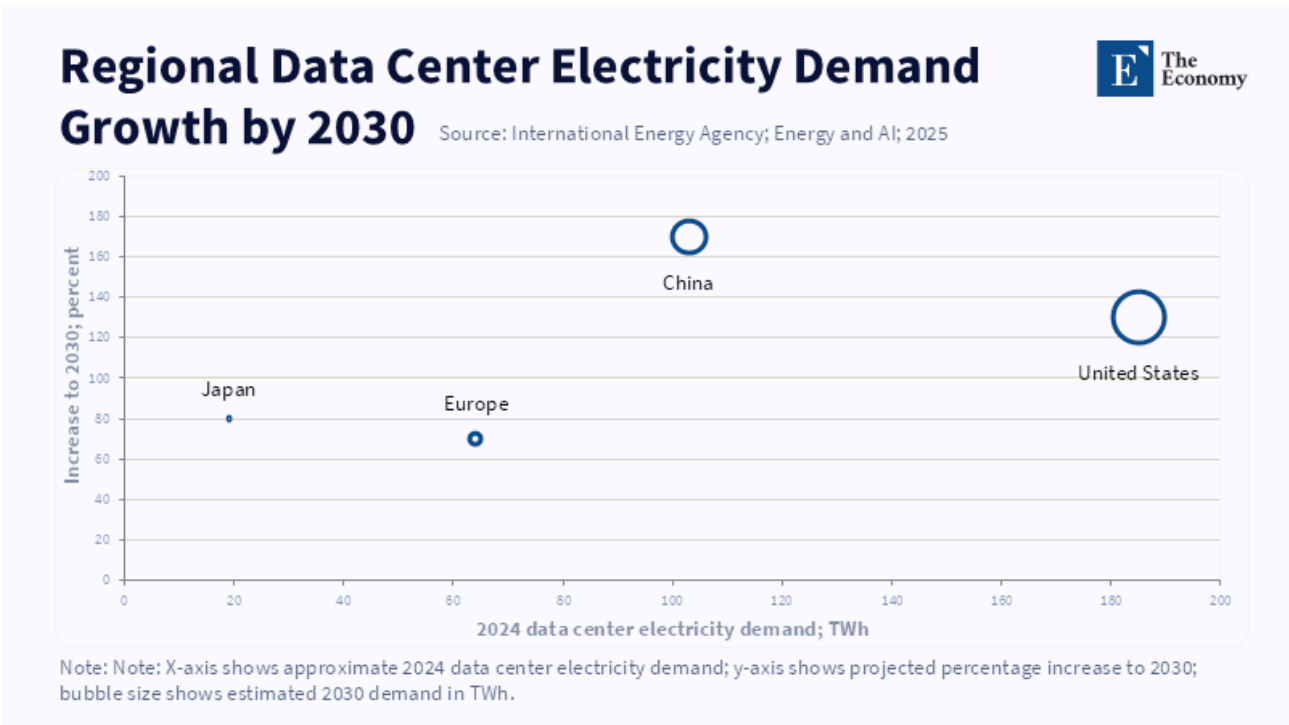


Figure 8: Regional growth patterns show that compute capacity will expand unevenly across the global power system.

In conclusion, the economic logic indicates that the AI infrastructure is not a footnote but a structural economic variable, reshaping the economics of large compute to map incentive structures. Reduce the wait for available compute by just a little and you find new forms of intelligence. Incentives are mismatched; everyone wants the newest, cheapest AI. Yet those who produce it, whether nations or corporations, must contend with grid limits, public opposition and regulatory headwinds. The allure of short-run cost-minimization, to minimize electricity bills, obscures the larger issues at play. As one observer notes, the story must shift from code to kilowatt hours, from algorithms to raw minerals. The central constraint is speed: the faster capital, chips, land and electricity become working compute, the stronger the economic position of the firms and countries that control it.

#### 4 Conclusion – Speed as the Core Economic Logic of Compute Strategy

When it comes to the infrastructure race, being faster is being better. Data center expenditures have surged in the last few years. Yet considerable policy debate still centers around subsidies, energy prices, or chip exports. However, the analysis shows that they are less important to the advantage gained through being among the first: a delay of just a few months can remove hundreds of millions from a project’s worth, delay a nation’s AI advancements for more than a year and hand advantage to a quicker adversary. Larger reductions in energy prices or tax breaks, on the other hand, make only modest improvements.

The strategic implication is that AI strategy in the future must treat large-scale compute capacity as strategic infrastructure, not a mere commodity. Rather than only manufacturing the latest semiconductor chips,

democratic governments will want to remove the bottlenecks now preventing projects from receiving power: streamline permitting procedures, upgrade weak power grids, promote "behind-the-meter" options such as dedicated clean-energy centers and other types of power plants and enable them to deliver power more quickly. In the same spirit, democracies will want to coordinate in the international arena.

The article's thesis is practical as much as strategic: broad AI adoption depends on broad and timely access to working compute. Policymakers who keep in mind the centrality of "time-to-power" will be empowered to design more effective approaches. Those who don't may see their institutions struggling over the coming decade to stay front-running, while still playing catch-up to an increasingly faster-moving AI revolution. The data and arguments presented herein support the conclusion that by committing resources and planning to move faster and maintain coordination, policy actors in educational systems, administrations and political systems can better ensure that future advanced AI systems are developed in ways that support domestic innovation, institutional resilience and broad economic use rather than deepening scarcity, dependence and strategic vulnerability. There is little time left: getting the immediate next wave of AI online first may become the most profound policy challenge of this century.

## References

- [1, 3, 10, 15] International Energy Agency (2025) *Energy and AI*. International Energy Agency.
- [2, 14, 17, 19, 22, 23, 25, 32] Phillips-Robins, A., Tawil, T. and Winter-Levy, S. (2026) *The Compute Coalition: How to Build the Future of AI in the Free World*. Carnegie Endowment for International Peace.
- [4, 7, 9] Pilz, K.F., Sanders, J., Rahman, R. and Heim, L. (2025) *Trends in AI Supercomputers*. arXiv.
- [5, 11] Chen, X., Wang, X., Colacelli, A., Lee, M. and Xie, L. (2025) *Electricity Demand and Grid Impacts of AI Data Centers: Challenges and Prospects*. arXiv.
- [6] Goldman Sachs Research (2026) *The Outlook for Data Centers in Asia*. Goldman Sachs; Reuters (2025) *Stargate UAE AI Datacenter to Begin Operation in 2026*. Reuters.
- [8] Trusilo, D. and Danks, D. (2024) *Commercial AI, Conflict, and Moral Responsibility: A Theoretical Analysis and Practical Approach to the Moral Responsibilities Associated with Dual-Use AI Technology*. arXiv.
- [12] Shehabi, A., Smith, S.J., Hubbard, A., Lei, N., Siddik, M.A.B., Holecek, B., Koomey, J., Masanet, E. and Sartor, D. (2024) *2024 United States Data Center Energy Usage Report*. Lawrence Berkeley National Laboratory and U.S. Department of Energy.
- [13] Central Statistics Office Ireland (2025) *Data Centres Metered Electricity Consumption 2024*. Central Statistics Office.
- [16, 20, 33] Cottier, B., Rahman, R., Fattorini, L., Maslej, N., Besiroglu, T. and Owen, D. (2024) *The Rising Costs of Training Frontier AI Models*. arXiv.

- [18, 20, 27] Ardestani, E.K., Piga, L., Stojkovic, J., Balaji, P., Ozdal, M., Jimenez Fernandez, M., Dimovska, M., Tadic, L., Shen, H., Vishwanath, D., Mishra, R., Mihret, M., Andrei, V., Cespedes, M., Prigent, J., Monahan, J., Graf, T., Li, B., Marquez, C., Kanaujia, S., Veeraraghavan, K. and Tang, C. (2026) *Provisioning to Runtime Optimization of a 100 MW-Scale AI Cluster*. arXiv.
- [21] U.S. Energy Information Administration (2024) *Units and Calculators Explained: Kilowatts, Megawatts and Electricity Use*. U.S. Energy Information Administration.
- [24, 26, 28] Brealey, R.A., Myers, S.C., Allen, F. and Edmans, A. (2022) *Principles of Corporate Finance*. 14th edn. New York: McGraw-Hill.
- [29] Cruzes, S. (2026) *AI Infrastructure Sovereignty*. arXiv.
- [30] Richardson, A., Yi, H., Nie, M., Wisdom, S., Price, C., Weijers, R., Veld, S. and Baker, M. (2025) *How Sovereign Is Sovereign Compute? A Review of 775 Non-U.S. Data Centers*. arXiv.
- [31] Goldman Sachs Research (2026) *The Outlook for Data Centers in Asia*. Goldman Sachs.
- [34] Chen, D., Zhou, Z., Cai, Y., Qin, J., Katchova, A. and Chen, L. (2026) *Concentrated Siting of AI Data Centers Drives Regional Power-System Stress Under Rising Global Compute Demand*. arXiv.