The Interconnection Moat: Why Energy Is AI's New Barrier to Entry

Second Order Labs™ · June 2026

ABSTRACT

The grid interconnection queue has replaced chip allocation as AI's binding constraint. Microsoft, Google, and Amazon are becoming energy companies because whoever controls the electrons controls the frontier.

Keywords: AI infrastructure, energy, nuclear power, data centers

> *"I think before we're talking about [AGI], you're going to run into energy constraints. I don't think anyone's built a gigawatt single training cluster yet… building a gigawatt data center takes a really long time."*
>
> — Mark Zuckerberg, Dwarkesh Podcast

Microsoft just signed a deal to restart a nuclear reactor that most Americans associate with disaster, not innovation. In September 2024, Constellation Energy announced a 20-year agreement to bring Three Mile Island Unit 1 back online: 835 megawatts of carbon-free power, earmarked for Microsoft's data centers.[1] That same month, Oracle's Larry Ellison spent his earnings call talking about nuclear permits. "We've already got the building permits for three nuclear reactors," he told investors.[2]

This isn't PR or ESG window dressing. It's a structural shift in what it means to be a tech company. AI companies are morphing into energy companies because electricity access. Not chip supply. Now decides who can compete at the top. The grid interconnection queue, with years-long waits for new projects, has replaced TSMC chip allocation as the choke point for AI scaling.[3] The Interconnection Moat is real: the edge now comes from having the grid connection to run your hardware, not just the hardware itself.

Forget the myth of the garage AI startup at the frontier. Training a GPT-5 class model takes utility-scale infrastructure and capital on the scale of a nation-state. The story of two founders with laptops building the next breakthrough? That's now limited to app-layer tinkering, not foundation model development.

I. The Physics Constraint: From Silicon to Electrons

Chip designers have run into the "power wall." NVIDIA's Blackwell architecture claims up to 25x lower cost and energy use for generative AI compared to its predecessor.[4] Google's TPU v5e delivers up to 1.7x more performance per watt than the previous generation.[5] These aren't just upgrades. They're last-ditch efforts to keep up with physics.

The bottleneck has moved. When GPUs were rare, companies scrambled for TSMC allocation. Now that chips are available, power is the constraint. A single ChatGPT query burns about 10 times the electricity of a standard Google search.[6] Multiply that by hundreds of millions of daily queries, and it's obvious why tech executives now talk about gigawatt constraints instead of software margins.

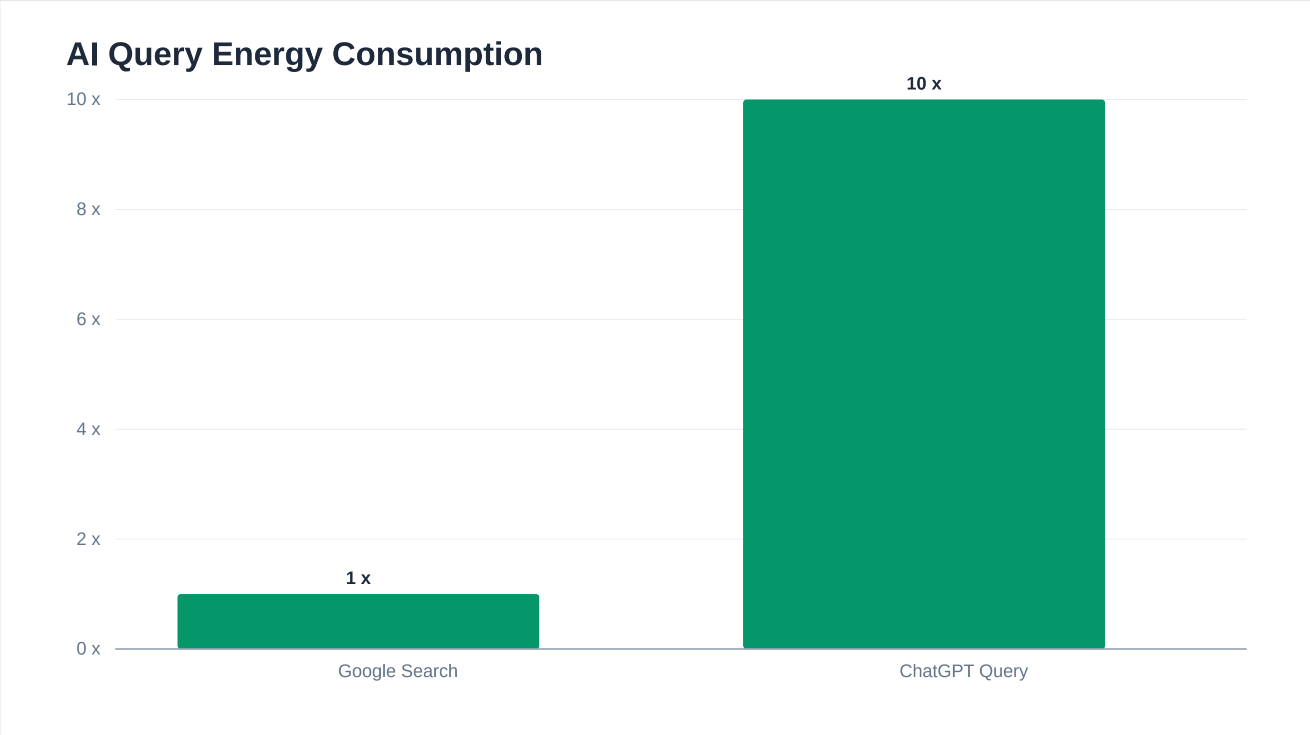

Figure 1: A single generative AI query consumes roughly 10x the energy of a traditional search, fundamentally altering baseline electricity demand at scale.

Hardware teams are scrambling. Liquid cooling, custom silicon tuned for FLOPs per watt, and data centers that look more like power plants than server farms. Meta's engineers now design facilities that are "liquid-cooled and optimized for our AI hardware."[7] Hardware-software co-design isn't about speed anymore. It's about thermal management and power density.

> *"I think before we're talking about [AGI], you're going to run into energy constraints. I don't think anyone's built a gigawatt single training cluster yet... building a gigawatt data center takes a really long time." — Mark Zuckerberg, Dwarkesh Podcast (April 2024)*

More than 2,600 gigawatts of generation and storage are waiting for grid connection in the US.[3] That's over twice the country's current generation capacity.

This backlog exists because the US grid was built for slow, predictable growth. Utilities planned decades ahead, building transmission for steady industrial expansion. No one planned for a single company demanding multiple gigawatts at one site in just a few years.

Bottleneck Era	Primary Constraint	Wait Time	Who Benefits
2020-2022	TSMC chip allocation	Months	Anyone with capital
2023-present	Grid interconnection	Years	Incumbents with existing queue positions

Figure 2: The constraint on AI scaling has moved from silicon to electrons. Having billions for GPUs is useless without the grid rights to power them.

Amazon's answer: buy what's already plugged in. In March 2024, AWS bought a 960-megawatt data center campus from Talen Energy for $650 million.[8] The site sits next to the 2.5-gigawatt Susquehanna nuclear plant, giving Amazon direct access to power and skipping the interconnection queue. That's the Interconnection Moat in practice. The real asset isn't the data center. It's the grid position.

Figure 3: When grid expansion timelines exceed AI competitive timelines, infrastructure migrates to the power source rather than waiting for power to reach the infrastructure.

II. Vertical Integration: Tech Companies as Power Producers

Google's 2024 deal with Kairos Power, billed as "the world's first corporate agreement to purchase nuclear energy from multiple small modular reactors," targets 500 megawatts by 2030.[9] But Google isn't just buying power. It's acting as an early-stage financier, underwriting the risk of unproven energy tech.

Utilities won't touch this risk. Their business depends on regulatory certainty and proven generation. When a utility proposes a new gas plant, it can model costs and timelines. SMRs have no commercial track record. Fusion is still theoretical. The math doesn't work for regulated utilities with ratepayer obligations.

Tech companies play by different rules. For Microsoft or Google, the value of reliable, carbon-free power for AI dwarfs the risk of backing unproven energy. They're becoming Independent Power Producers because the traditional energy market moves too slowly for AI.

Figure 4: The IEA projects data center electricity demand to more than double by 2026, driven primarily by AI workloads.

Look at the numbers in action. Microsoft announced a $10 billion investment in new data center infrastructure in 2024, including a major expansion in Iowa tied directly to new power purchase agreements. The International Energy Agency projects global data center electricity demand will more than double from 460 terawatt-hours in 2022 to over 1,000 by 2026.[10] Goldman Sachs expects AI to drive a 160% jump in data center power demand by 2030, requiring about $50 billion in new US generation.[11] By 2030, data centers could eat up 9% of US electricity, up from 4% today.[12]

Microsoft's 2024 Environmental Sustainability Report spells out the contradiction. Scope 3 emissions jumped 30.9%, "primarily driven by the construction of more datacenters and the associated embodied carbon in building materials, as well as hardware components."[13] Brad Smith, Microsoft's president, admitted in a May 2024 interview that "our cloud growth is outpacing our ability to reduce emissions in the short term." The company's pledge to be carbon-negative by 2030 collides with the reality that every new data center adds to the emissions tally.

Nuclear power is the only plausible escape, which is why Microsoft is reviving Three Mile Island. But even that's a slow fix. As Jesse Jenkins, an energy systems analyst at Princeton, put it: "Restarting a nuclear plant is a multi-year process, and new builds are a decade-long bet at best." The timelines for new nuclear capacity, even for restarts, stretch well beyond most corporate climate targets.

Efficiency arguments don't hold up under scrutiny. Google's 2023 environmental report shows that despite a 20% improvement in data center efficiency, total energy use still rose 17% year-over-year because of AI workload growth. The gains get swallowed by demand.

III. The Geopolitical Stakes and Death of the Frontier Startup

OpenAI's pitch to the White House in 2024 called for data center campuses that would each draw five gigawatts. About as much as a mid-sized US city.[15] The subtext: if the US grid can't deliver, AI infrastructure will move to where the power is. Mubadala, Abu Dhabi's sovereign wealth fund, has already approached Microsoft and Google about building hyperscale data centers in the Gulf, offering cheap natural gas and fast-track permitting. For these funds, it's simple: moving data is cheaper than moving electricity. The compute will follow the electrons.

This shift upends the definition of a data center. The old model. A building plugged into the grid like any other industrial customer. Breaks down at gigawatt scale. The new model is a power plant with servers attached. Amazon's Susquehanna deal and Microsoft's Three Mile Island play fit this mold. The data center isn't just grid-connected. It's wired straight to the generation asset.

The Interconnection Moat redraws the map of who can build at the AI frontier. Take a startup with $50 million in funding. It can rent cloud compute, but it can't lock in a 20-year nuclear power deal or wait out a multi-year interconnection queue. It can't bankroll an SMR pilot. The capital and timelines now exclude everyone but the hyperscalers and sovereign-backed giants.

Ask any founder trying to train a new foundation model. "We spent months negotiating for GPU access, but the real wall was power," said one startup CEO who abandoned plans for a custom LLM in 2024. Application-layer companies will keep building on top of hyperscaler models, and fine-tuning and inference remain open to smaller players. But the work that sets the industry's capability ceiling. Training the next generation of frontier models. Now requires infrastructure only a handful of organizations can touch.

Sam Altman didn't mince words at Davos 2024: "There's no way to get there without a breakthrough... It motivates us to go invest more in fusion."[16] The race for AGI is now a race for energy. Compute scaling laws hit physics. Chip efficiency buys time, but doesn't change the constraint. Whoever cracks energy generation will set the ceiling for AI.

Moats in tech shift. In the early 2000s, MySpace's lead in user numbers looked unassailable. Until Facebook's network effects and real-identity policy flipped the script. In mobile, Apple's platform control crushed Nokia's distribution advantage. Now, energy access is taking over as the moat for AI infrastructure.

History punishes those who miss these shifts. Kodak clung to film while digital ate its lunch. If you think Microsoft's nuclear deals or Google's SMR investments are just sustainability theater, you're missing the real play. These are not routine procurement moves. They're bets on a future where electrons, not algorithms, draw the competitive boundaries.

Will the interconnection queue clear? Not on any timeline that matters for AI. Grid expansion moves at a bureaucratic crawl. The next decade will see tech giants building private power plants, bypassing public infrastructure, and locking in control over the next wave of AI capability. The industry is heading for a world that looks less like the open software era and more like the old telecom monopolies, where owning the pipes meant owning the market. The question isn't who has the best model. It's who owns the switch.

KEY FINDINGS

Grid interconnection queues with five-year wait times have replaced chip allocation as the primary bottleneck for AI scaling.

Tech giants are becoming Independent Power Producers, signing 20-year nuclear deals and financing unproven SMR technology to bypass slow-moving utilities.

A single ChatGPT query requires nearly 10x the electricity of a Google search, fundamentally altering baseline global electricity demand.

The frontier AI startup era is ending—training next-generation models requires utility-scale infrastructure only hyperscalers can access.

Data centers could consume up to 9% of US electricity by 2030, forcing a collision between AI ambitions and corporate climate pledges.

REFERENCES

[1] Constellation Energy press release, "Constellation to Launch Crane Clean Energy Center," September 2024.

[2] Oracle Q1 FY25 Earnings Conference Call, October 2024.

[3] Lawrence Berkeley National Laboratory, "Queued Up: Characteristics of Power Plants Seeking Transmission Interconnection."

[4] NVIDIA Developer Blog, "Blackwell Architecture and GPUs."

[5] Google Cloud Blog, "Cloud TPU v5e: Accelerating AI with cost efficiency."

[6] de Vries, A. "The growing energy footprint of artificial intelligence," *Joule*, 2023.

[7] Meta Engineering Blog, "Reimagining our infrastructure for the AI age," May 2023.

[8] Talen Energy investor announcement, March 2024.

[9] Google Sustainability Blog, "Google and Kairos Power announce landmark advanced nuclear energy deal."

[10] International Energy Agency, "Electricity 2024: Analysis and forecast to 2026."

[11] Goldman Sachs, "AI is poised to drive 160% increase in data center power demand."

[12] Electric Power Research Institute, "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption."

[13] Microsoft 2024 Environmental Sustainability Report.

[14] Jensen Huang, GTC 2024 Keynote.

[15] Bloomberg, "OpenAI Pitched White House on Unprecedented Data Center Buildout," September 2024.

[16] Sam Altman, World Economic Forum Annual Meeting 2024, Davos.
