

AI infrastructure is becoming a national security asset

Data centers, chips, and the power that feeds them have moved from the balance sheet to the situation room. For CXOs, operators, and investors, that shift changes who gets to build, who gets to buy, and who gets to sell.

COMPUTE FORECAST · JULY 2026 · AI INFRASTRUCTURE / NATIONAL SECURITY

IN BRIEF

- In 17 months, US policy on advanced AI chip exports to China has reversed direction roughly eight times — turning chip export policy into a live variable that AI infrastructure operators must underwrite for, not a fixed constraint.
- The Trump administration's July 2025 AI Action Plan and its companion executive order on data center permitting formally recast data centers as strategic national assets, opening federal land and expedited review to “Qualifying Projects” above 100 megawatts.
- Sovereign AI deals with the UAE and Saudi Arabia — a 5 GW campus anchored by a 1 GW Stargate cluster in Abu Dhabi, and Humain's push toward 6.6 GW in Saudi Arabia — now carry explicit security conditions, including divestment from Chinese technology partners.
- The Pentagon has become a top-tier compute buyer, with one federal-spending analysis estimating more than \$32 billion committed to AI, cloud, and cybersecurity programs in the first half of fiscal year 2026 alone, while Congress debates a standalone critical-infrastructure designation for data centers.

The new premise

For most of the cloud computing era, a data center was a commercial asset like any other: sited for power, land, and tax incentives, financed like a warehouse, and regulated mostly as a local zoning and utility question. That premise no longer fully holds. Over the past eighteen months, the United States government has moved — through export controls, executive orders, defense contracts, and foreign investment

review — to treat the physical infrastructure of artificial intelligence as something closer to a strategic resource: comparable, in the language now used by officials and executives alike, to energy pipelines, telecommunications backbones, or the defense industrial base.

This is not a single policy. It is a pattern across at least six distinct channels: semiconductor export controls, federal land and permitting policy, sovereign investment screening, defense procurement, critical-infrastructure classification, and diplomatic deal-making around who gets access to advanced compute. Each channel has moved independently, on its own timeline, often driven by different agencies with different — at times competing — mandates. But taken together, they describe a single, unmistakable direction: the US government now sees who builds, owns, and operates AI infrastructure as a matter of national security, not just industrial policy.

For CXOs, founders, and investors in data centers, neoclouds, and the power infrastructure that feeds them, this creates a genuinely new operating environment. Capacity planning, capital structure, and even choice of customer now carry a policy-risk dimension that did not meaningfully exist three years ago. This analysis maps the six channels and sets out what the shift means for anyone allocating capital into AI infrastructure through the rest of this decade.

Export controls: the fastest-moving lever

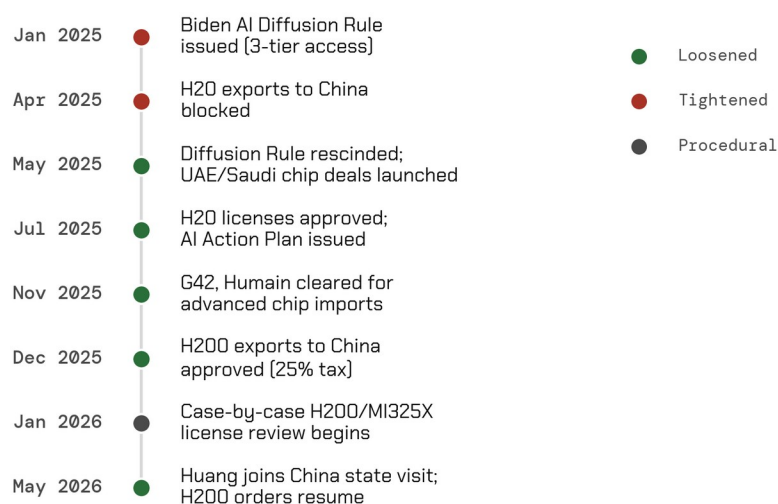
Nowhere is the volatility more visible than in semiconductor export policy. The Biden administration's approach culminated in the AI Diffusion Rule, issued in January 2025, which sorted the world into three tiers of access to advanced US chips based on political alignment with Washington. The Trump administration moved to rescind that rule on May 13, 2025 — two days before it was set to take effect — and replaced tiered access with a more transactional, deal-by-deal approach.

What followed was a sequence of reversals that policy trackers now treat as a defining feature of the environment rather than an aberration. In April 2025, the Commerce Department's Bureau of Industry and Security (BIS) determined that Nvidia's China-tailored H20 chip — engineered specifically to comply with the prior round of controls — required a license to export, effectively halting shipments and forcing Nvidia to take a multibillion-dollar charge. Three months later, in July 2025, BIS reversed course again, confirming that H20 export licenses would be approved; President Trump publicly defended the reversal in August 2025 by describing the H20 as “obsolete” relative to Nvidia's newer Blackwell architecture. In practice, few if any H20 units were ultimately sold into China: Chinese authorities discouraged domestic AI firms from buying them on security grounds, and Nvidia had already wound down production.

The next inflection came in December 2025, when President Trump announced a one-year waiver allowing export of Nvidia's H200 chip — a more capable, Hopper-generation part the original Diffusion Rule would not have permitted China to access at all — subject to a 25% tax on transaction value. The Commerce Department subsequently cleared roughly ten Chinese customers, including Alibaba, Tencent, and ByteDance, to purchase H200 units, capped at 75,000 units per customer. BIS instituted case-by-case license review for H200 and AMD's competing MI325X part effective January 15, 2026. In practice, the licenses have not translated into meaningful trade: Beijing's State Council ordered a supply-chain security review and told domestic firms to hold off on purchases, and by mid-July 2026 a senior Commerce official testified to Congress that actual H200 shipments against approved licenses remained “trivial.” Nvidia CEO Jensen Huang joined President Trump's state visit to China in May 2026 in an effort to help break that deadlock — a sign of how stalled, rather than resumed, the trade had become.

Export policy has moved eight times in 17 months

US advanced-chip export control actions toward China, January 2025–May 2026



COMPUTE FORECAST

US ADVANCED-CHIP EXPORT CONTROL ACTIONS TOWARD CHINA, JANUARY 2025–MAY 2026

Two things are worth separating clearly for planning purposes. What is confirmed: the Diffusion Rule is rescinded; H20 exports are formally approved (if commercially moot); H200 exports are licensed subject to a 25% tax and case-by-case review, though actual shipments remain minimal; and Nvidia's most advanced Blackwell-generation parts remain restricted from China. What is contested or in flux: whether the H200 waiver and an associated licensing carve-out — both due to expire in late 2026 — will be renewed, allowed to lapse, or replaced; and whether a bipartisan Chip Security Act, introduced in May 2025 and since advanced out of House committee, will be enacted and impose statutory guardrails that survive changes in administration posture.

For infrastructure operators, the practical takeaway is that chip export status is no longer a fixed input to a five-year model. It is a variable that has moved materially, in both directions, an average of roughly once every two months over the period examined. Any capacity plan, customer contract, or capital raise premised on a specific chip's export status remaining fixed for the life of the asset should now carry an explicit sensitivity case for policy reversal.

Data centers as strategic federal infrastructure

While chip policy has been volatile, a second and more durable shift has taken shape: the US government's own physical infrastructure posture toward data centers. On July 23, 2025, the White House released its AI Action Plan — organized around three pillars: accelerating innovation, building American AI infrastructure, and leading in international AI diplomacy and security — alongside three executive orders. One of them, “Accelerating Federal Permitting of Data Center Infrastructure” (Executive Order 14318), is the clearest statement to date that Washington now treats large-scale AI data centers as a category of national infrastructure warranting direct federal facilitation.

The order defines “Data Center Projects” as facilities requiring more than 100 megawatts of new electrical load for AI training, inference, simulation, or synthetic data generation, and directs the Departments of Energy, Defense, and the Interior to identify suitable federal land for both data centers and the power generation needed to run them — including authorizing the Department of Defense to competitively lease land on military installations for qualifying projects. It also directs the Secretary of Commerce to establish financial support covering loans, grants, tax incentives, and offtake agreements, and instructs agencies to identify environmental-review shortcuts within ten days of signing.

Implementation has moved unevenly across agencies but is real. One day after the order, the Department of Energy announced it had selected four federal sites — Idaho National Laboratory, Oak Ridge Reservation in Tennessee, Paducah Gaseous Diffusion Plant in Kentucky, and Savannah River Site in South Carolina — to move toward private-sector solicitations for co-located AI data centers and energy generation, including nuclear. By the fall of 2025, several DOE program offices had issued formal requests for proposals at three of the four sites, and the US Air Force had begun accepting proposals on its own land. The Department of the Interior, by contrast, had not publicly identified any sites on its federal land holdings as of late 2025.

DOE Secretary Chris Wright described the effort as a step to “accelerate the next Manhattan Project” and secure US AI and energy leadership by putting federal land to work, language that treats compute capacity and energy capacity as a single strategic asset. For operators, the practical opportunity is real — federal land access, expedited

permitting, and new financing tools are genuinely on offer — but so is the reality that “Qualifying Project” status increasingly comes bundled with federal visibility into who a project's ultimate customers and capital sources are.

Sovereign AI deals: access to compute as diplomacy

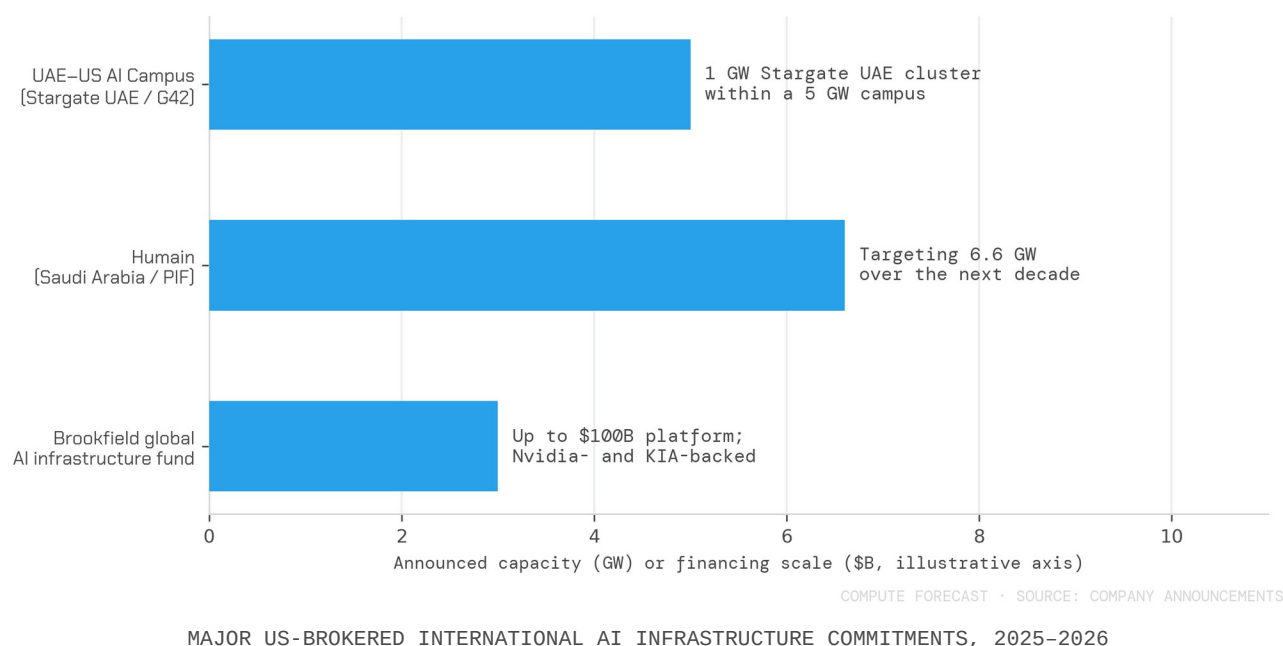
If domestic policy has reframed data centers as federal infrastructure, US diplomacy has simultaneously reframed access to American AI chips as a foreign-policy instrument — most visibly in the Gulf. During President Trump's state visit to the UAE in May 2025, the US and UAE governments announced the US-UAE AI Acceleration Partnership, anchored by a 5-gigawatt “UAE-US AI Campus” in Abu Dhabi. Within that campus, G42 is building Stargate UAE, a 1-gigawatt AI compute cluster developed with OpenAI, Oracle, Nvidia, Cisco, and SoftBank. The first 200-megawatt phase was targeted for 2026; G42 confirmed in October 2025 that construction was underway on that timeline.

The deal did not move smoothly. Reporting in mid-2025 indicated the Commerce Department had withheld chip-shipment approvals over G42's historical ties to Chinese technology firms. The resolution came via explicit security commitments: G42 agreed to divest from Chinese partners, including Huawei — building on the divestment terms already attached to Microsoft's own \$1.5 billion investment in G42 in 2024 — and, per Microsoft President Brad Smith's description of the “very strict conditions” attached to the November 2025 chip-export approvals, to build a matching US data center for every one it operates in the UAE. By November 2025, BIS had formally authorized G42 to import advanced Nvidia chips, and the US approved roughly \$1 billion in combined Nvidia exports to G42 and Saudi Arabia's Humain.

A parallel deal took shape in Saudi Arabia. Humain, established by the Public Investment Fund in May 2025, is pursuing 6.6 gigawatts of capacity over the next decade — including a 500-megawatt facility with xAI, an 18,000-unit initial Nvidia GB300 order, a \$10 billion collaboration with AMD, and a \$5 billion “AI Zone” partnership with AWS. Separately, Brookfield Asset Management announced plans in November 2025 to finance up to \$100 billion in global AI infrastructure, backed by Nvidia and the Kuwait Investment Authority.

Sovereign AI deals are now measured in gigawatts

Major US-brokered international AI infrastructure commitments, 2025-2026



Both Gulf programs share a common design: American chips, deployed on foreign soil, wrapped in security conditions that the parties involved have taken to calling a “Gold Standard” for securing AI models, chips, data, and access. What is not in dispute is that the deals establish a template — American compute exported abroad only in exchange for verifiable security commitments — likely to recur as more countries seek sovereign AI capacity.

The Gulf is not the only theater. OpenAI announced Stargate Norway in July 2025 — a considerably smaller, lower-friction deal than its Gulf counterparts, consistent with Norway's status as a NATO ally requiring none of the security-conditioning apparatus built around G42 or Humain. Operators evaluating cross-border compute partnerships should expect compliance burden to track that same logic: allied-nation partnerships moving quickly, while partnerships with nations perceived as closer to strategic competitors carry heavier security conditions and longer timelines.

CFIUS and the new scrutiny of foreign capital

The same logic is now working in reverse, applied to foreign capital flowing into US data centers. The Committee on Foreign Investment in the United States (CFIUS) has both expanded its jurisdiction and, under the Trump administration's “America First Investment Policy,” begun sorting foreign capital by country of origin rather than reviewing all foreign investment uniformly.

The National Defense Authorization Act for Fiscal Year 2026 authorized CFIUS to build a running list of “national security-sensitive sites” that would fall under its real estate jurisdiction. The Treasury Department separately opened comment, closing March 18, 2026, on a proposed “Known Investor Program” intended to fast-track review for allied-nation investors who agree to avoid partnering with designated “foreign adversaries.” Data centers fall within CFIUS's remit particularly where they host sensitive data or support government or defense customers — a description that covers a growing share of the neocloud sector.

The result is a bifurcated regime: capital from the UAE, Saudi Arabia, and similar partner nations is being courted and eased through review, provided it comes with security commitments; capital with traceable links to China or other designated adversary nations faces a rising, increasingly automatic bar. For data center operators seeking capital, the passport attached to a term sheet is now a material underwriting question, not an afterthought.

China's parallel build-out

US policy is only one half of the picture; the other is what China is doing with the compute it can access domestically. Huawei's Ascend line remains the centerpiece of China's effort to substitute for restricted Nvidia hardware. Huawei shipped an estimated 812,000 Ascend chips in 2025 and is targeting roughly 600,000 units of its Ascend 910C chip alone in 2026 — roughly double 2025 output — with total Ascend die output potentially reaching 1.6 million units in 2026. Huawei's roadmap for its next-generation Ascend 950PR and 950DT chips, expected in 2026, projects lower total processing performance than the current 910C — which the Council on Foreign Relations reads as evidence that manufacturing partner SMIC is struggling to scale advanced-node production.

The Council on Foreign Relations calculated that the best current US AI chips outperform Huawei's best offerings roughly fivefold on aggregate processing performance, a gap CFR projects could widen to seventeenfold by 2027. China's leading domestic DRAM producer, CXMT, is projected to produce enough high-bandwidth memory in 2026 to support only 250,000–300,000 Ascend 910C-equivalent packages — well short of SMIC's die-production capacity.

None of this has stopped Beijing from committing substantial state capital: Chinese planners are reportedly drafting a roughly \$295 billion national AI computing grid initiative mandating 80% domestic-chip content, a guaranteed-demand structure that gives Huawei a scale of committed customer base competitive markets alone would be unlikely to provide. Reasonable analysts differ sharply on how to weigh these facts — some read China's supply constraints as a durable structural disadvantage; others argue state-directed demand will compound into meaningful capability over a longer

horizon regardless of near-term performance gaps. Both readings are represented in the current policy debate.

Grid security and the critical-infrastructure question

A third front concerns whether data centers and the grid infrastructure feeding them should be formally classified as critical infrastructure in their own right. Today, data center security in the US largely falls under the existing “Information Technology” critical infrastructure sector designated by CISA — one of sixteen sectors under a framework that predates the current AI-driven data center boom by more than a decade.

That gap became explicit in 2026: on April 29, the House Homeland Security Subcommittee on Cybersecurity and Infrastructure Protection held a hearing examining whether AI data centers warrant their own standalone sector designation. Witness Mark Montgomery pointed to the United Kingdom, which has already created a separate designation for data centers, noting current US policy “does not provide a clear, unified approach to data center security.”

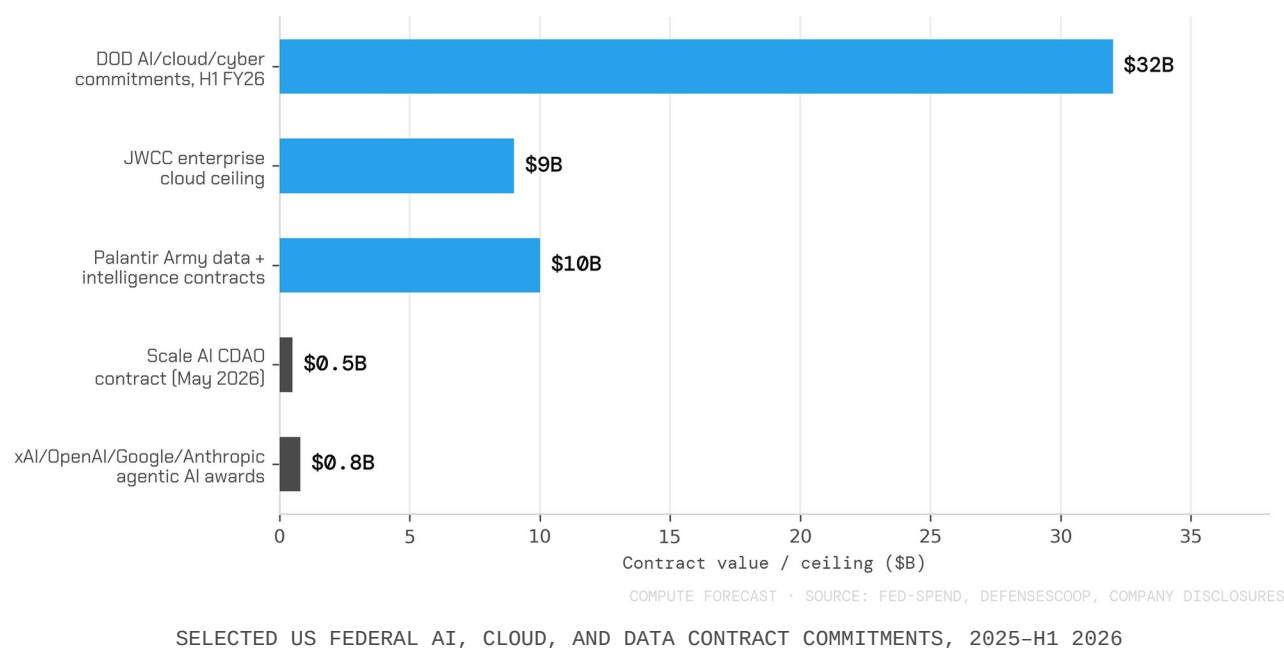
CISA's own posture reinforces the urgency: on May 5, 2026, the agency released its “CI Fortify” initiative, instructing utilities to plan on the assumption that adversaries will already have some access to operational technology networks in a crisis, naming Iran-linked units as willing to deploy data-wiping malware and describing China and Russia as having invested in long-term “pre-positioning” inside Western grid infrastructure. No formal decision to create a standalone data center sector has been made, but the direction of travel suggests data center-specific security obligations will become more concrete over the next 12 to 24 months.

The Pentagon as compute customer

The US Department of Defense has become one of the largest single buyers of AI compute and cloud services in the American market. The Pentagon's Chief Digital and Artificial Intelligence Office (CDAO) has emerged as the primary vehicle: in May 2026, Microsoft, Amazon, and Google each signed expanded classified military AI agreements the same week Scale AI's CDAO contract grew fivefold, from \$100 million in September 2025 to \$500 million. CDAO has now approved AI systems from eight firms, including the major hyperscalers, for classified networks.

The Pentagon is becoming a top-tier compute buyer

Selected US federal AI, cloud, and data contract commitments, 2025–H1 2026



One federal-spending analysis estimates DoD committed more than \$32 billion in contract ceiling to AI, cloud, cybersecurity, and data analytics programs in just the first half of fiscal year 2026 — production deployment, not experimental R&D. The Pentagon's Joint Warfighting Cloud Capability program, which awarded Google, Oracle, AWS, and Microsoft a \$9 billion cloud contract in 2022, is moving toward a follow-on “JWCC Unified Cloud Marketplace,” with awards expected in early 2027. Palantir holds more than \$10 billion in contract ceiling across Army and intelligence-community programs on the same analysis, and in mid-2025 the Pentagon awarded \$800 million combined to xAI, OpenAI, Google, and Anthropic for agentic AI capabilities — a program Google's access to was separately expanded in April 2026.

This buildout has not gone unchallenged: the bipartisan Protecting AI and Cloud Competition in Defense Act argues DoD's reliance on a small number of large technology companies risks vendor lock-in, and would require contracts structured to preserve competition. That effort remains pending but signals defense AI spending is now significant enough to generate its own competition-policy debate.

Market and investment implications

Taken together, these six channels point to a durable structural shift rather than a temporary policy fashion, even though individual rules — chip export status above all — remain genuinely volatile.

First, “trusted compute supply chain” status is becoming a commercial asset in its own right, distinct from raw capacity or price. Operators who can demonstrate clean chip provenance and security-vetted ownership structures are positioned to win government and government-adjacent contracts growing faster than the commercial hyperscaler market in percentage terms.

Second, foreign capital sourcing now carries jurisdiction-specific underwriting risk. Gulf sovereign capital is being courted but arrives bundled with security conditions; capital with Chinese ownership links faces a rising, increasingly automatic CFIUS bar.

Third, federal land and financing tools under the Accelerating Federal Permitting order represent genuine opportunity — but likely one that favors larger, better-resourced developers able to navigate multi-agency coordination and meet the >100 MW threshold.

Fourth, chip export volatility should be modeled, not assumed away. An operator building a financial model on the assumption that current export status will remain fixed for the life of an asset is building on a variable that has moved materially roughly eight times in the period examined.

Fifth, the still-undecided critical-infrastructure classification question is worth monitoring for its compliance-cost implications. Operators who voluntarily adopt CI Fortify-consistent security practices now are better positioned than those waiting for a mandate to formalize.

What this means for decision-makers

For CXOs and founders, the operating assumption should shift from “compliance is a fixed cost of doing business” to “policy positioning is a competitive input,” on par with power procurement or chip supply. For investors, geopolitical and export-control risk deserves the same explicit line-item treatment in diligence that construction risk or power availability already receive. For operators and vendors serving both commercial and government customers, the security and provenance standards increasingly demanded by the Pentagon, CFIUS, and sovereign-deal frameworks are converging; building to the highest common standard now is likely cheaper than retrofitting later.

The throughline across export controls, federal land policy, sovereign deals, foreign investment review, critical-infrastructure debate, and defense procurement is the same: the United States government has concluded that whoever controls the physical and diplomatic infrastructure of AI compute holds a form of strategic leverage it intends to actively manage, not leave to market forces alone. For the industry Compute Forecast covers, the practical consequence is straightforward: the balance sheet and the geopolitical map are no longer separate documents. They increasingly have to be read together.

Compute Forecast tracks the infrastructure, capital, and policy decisions shaping the AI build-out across data centers, neoclouds, and the global energy grid.

COMPUTE FORECAST — TRACKING THE COMPUTE ECONOMY