



\$325

PJM 2028/29 CLEARING PRICE
PER MW-DAY; THIRD STRAIGHT AUCTION
AT CAP

6.8 GW

PJM SHORTFALL VS. ITS 156 GW
RELIABILITY REQUIREMENT

2.8T

KIMI K3 PARAMETERS; 16 OR
896 EXPERTS ACTIVE

277 GW

AURORA TRACKED U.S. DATA
CENTER DEVELOPMENT QUEUE

WEEK IN REVIEW

Cheaper Intelligence, Pricier Infrastructure

The week's apparent contradiction is the point. At the model layer, capability became cheaper and more portable: Moonshot's Kimi K3 arrived as a 2.8-trillion-parameter sparse model with aggressive API pricing, near-frontier benchmarks and a promised open-weights release, while sub-1-bit and ternary quantization pushed server-class checkpoints toward phones, vehicles and industrial endpoints. Neither is yet a kilowatt-hour measurement. Cheaper tokens redistribute inference among hyperscalers, neoclouds, sovereign clusters and enterprises; on the available evidence, they do not reduce aggregate electricity demand.

The endpoint shift does not reduce data center demand; it creates a hierarchical edge-and-cloud system that still requires frontier training, simulation, fleet learning, orchestration and difficult-case escalation. The unit of demand moves from human prompts to intelligent endpoints and continuous inference cycles. The missing variables remain workload growth, compute per completed task and actual utilization.

Against that software deflation, the physical market became more discriminating. Aurora Energy Research's 277 GW development queue includes 145 GW from developers with no operating capacity; stakes worth tens of billions are being marketed, lenders are showing saturation, Oracle's credit sits one notch above junk and two-thirds of industry respondents expect more distress. The buildout remains enormous; the quality of the load book is separating rapidly.

Meta supplied the counter-example: a Louisiana campus moving toward 5 GW of compute and more than \$250 billion of total investment including chips, a potential \$10 billion Anthropic lease, and senior AWS infrastructure hires. Overbuild is becoming an option — capacity can serve internal models, hedge scarcity through third-party leases or become a merchant compute business.

The most expensive power signal came from Valley Forge. PJM's 2028/29 Base Residual Auction, released July 14, cleared at the FERC-approved cap of \$325/MW-day (UCAP) across the entire footprint — the third consecutive auction to hit the collar negotiated with the region's governors, and roughly 11 times the price level of the 2024/25 delivery year. The auction procured 138,318 MW of unforced capacity through RPM; Fixed Resource Requirement entities committed another 10,864 MW, bringing total supply to 149,182 MW against a reliability requirement of 156,013 MW. The RTO is 6,831 MW short of its one-in-ten-year standard for the second consecutive delivery year, operating on a 14.7% reserve margin against a roughly 20% target.

The composition is as instructive as the shortfall. Only 525 MW of new generation and uprates cleared, while nearly 3 GW of coal exited through retirement or conversion; the resulting mix is 46% gas, 20% nuclear and 18% coal. Forecast peak load rose roughly 2,000 MW year over year, driven largely by data centers, and cleared supply times price implies a \$16.4 billion capacity bill. FERC Chair Laura Swett called the result another alarm bell. With an expanding shortfall, the consensus view is that PJM's 'Shapiro Collar' in the mid-\$300s is less than half the price necessary to unlock new builds, and their upcoming Reliability Backstop Procurement is going to try and introduce higher prices without burdening ratepayers overall. If it is not successful, then the Price will continue to be administered while quantity fails to arrive: the definition of a market that is rationing, not clearing.

The electorate is tightening the boundary from the other side. DOE's draft National Transmission Needs Study treats data center growth as a national transmission driver, New York imposed the first statewide moratorium and nationally organized data center opponents staged 142 protests across 42 states this past weekend. Power, credit and social license are converging into one underwriting test.

BOTTOM LINE

Intelligence is getting cheaper, but the megawatt is not — PJM is now administratively pricing scarcity it cannot procure. Value continues to migrate toward firm power, credible tenants, flexible compute architectures and projects with the local legitimacy to reach energization.

In this Issue

02 The Queue is not
the Load

04 Meta Becomes a
Compute Utility

06 Weekly Roundup

★ CHART OF THE WEEK | PJM 2028/29 BASE RESIDUAL AUCTION



Sources: PJM Interconnection, 2028/29 Base Residual Auction results (July 14, 2026) and prior BRA reports. Clearing prices shown on a UCAP basis; the last three auctions cleared at the FERC-approved price collar agreed with the 13 PJM-state governors.

The auction's message for developers is asymmetric: capacity revenue is capped, but scarcity is not. With 525 MW of new supply clearing against a 6,831 MW shortfall, gas carrying 46% of the accredited stack and the next auction — May 2027 — uncapped, the premium accrues to projects that can bring firm, accredited megawatts to market ahead of the 2029/30 cycle, and to load that arrives with its own generation.

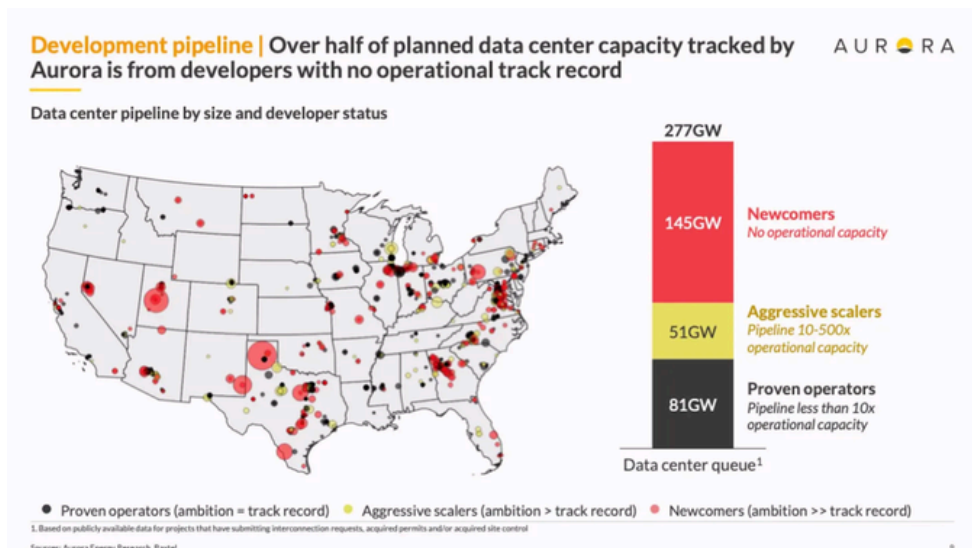
★ MARKET SPOTLIGHT | FEATURED ANALYSES

THE QUEUE IS NOT THE LOAD

Aurora's 277 GW map is a development pipeline, not an electricity forecast

Aurora Energy Research's development map is the most useful chart of the week because it puts a denominator under the AI infrastructure boom. The firm tracks 277 GW of proposed U.S. data center capacity tied to interconnection filings, permits or site control. But 145 GW — more than half — comes from developers with no operating capacity, while another 51 GW comes from aggressive scalars whose pipelines are 10 to 500 times their existing footprints.

The commoditization thread connects directly to the inference-cost data above and to the supply chain. AWS is raising prices on Nvidia-based EC2 Capacity Blocks by 20% effective July 1 — its second increase since January — explicitly citing supply and demand, even as open-weight routing accelerates. The simultaneous truth is higher compute prices at the cloud layer and falling effective prices per unit of intelligence at the model layer. Both point to the same conclusion: margin accrues to whoever controls the scarce physical inputs.



Sources: Aurora Energy Research / Battel; projects with interconnection requests, permits and/or site control. Bubble size represents proposed capacity; color reflects developer operating status.

The map does not say the demand is false. It says the conversion risk is concentrated. Site control can be purchased; queue positions can be filled; press releases can be issued. Firm power, air and water permits, construction labor, equipment, tenant credit and financing must still arrive in the correct sequence. The distinction matters because utilities, regulators and developers often treat announced capacity as though it were a homogeneous load forecast.

The more accurate framework is a probability-weighted pipeline. Proven operators with contracted tenants and repeatable capital structures deserve a materially higher conversion factor than first-time sponsors assembling land around speculative load. Aurora's segmentation is therefore not a judgment about ambition; it is a map of execution risk.

THE CAPITAL STACK IS ALREADY SORTING THE PIPELINE

The first evidence is not in interconnection data but in the capital markets. U.S. data center owners are marketing majority stakes worth tens of billions of dollars, including a potential DataBank transaction of as much as \$25 billion. Buyers are explicitly valuing guaranteed power and considering milestone-based payments for projects exposed to cost and schedule risk. That is infrastructure underwriting replacing thematic AI exposure.

\$334.5B AI infrastructure bonds and loans issued in 2026 year-to-date	68% Industry respondents expecting more distress within 12-18 months	BBB- Oracle rating, one notch above junk	\$30B U.S. high-yield data center issuance in 2026
--	--	---	---

Sources: Bloomberg; AlixPartners survey; S&P Global Ratings; Wall Street Journal reporting on data center stake sales.

DISTRESS CAN COEXIST WITH STRONG END DEMAND

AlixPartners found that 68% of more than 400 industry respondents expect distressed situations to increase over the next 12 to 18 months. The causes are cumulative: energy costs, thinner cash flow, compute-price pressure, labor shortages, equipment logjams, delayed permits and tightening capital markets. As one adviser put it, this is not necessarily a demand problem; it is a margin problem.

Prime Data Centers provides an interesting market signal. It floated a \$600 million Nordic bond, downsized it to \$500 million and then postponed the sale when investors rejected the proposed terms. That happened after the sector had already raised roughly \$30 billion in U.S. high-yield debt this year. The constraint is not the absence of capital; it is saturation and selectivity. Fully leased, hyperscaler-backed assets can still finance. Corporate-level growth capital and weaker tenant stories are repricing first.

ORACLE SHOWS HOW EXECUTION RISK REACHES THE BALANCE SHEET

Oracle is spending into the gap between contracted future revenue and cash required today. S&P cut the company to BBB-, Moody's retains a negative outlook and Oracle has burned more than \$20 billion after capital expenditure over the last four quarters. Credit protection reached a record close near 198 basis points. S&P estimates the free-cash-flow deficit could widen to \$42 billion, while roughly half of Oracle's \$638 billion of remaining performance obligations is tied to OpenAI.

Oracle's Project Jupiter turns the abstract credit issue into a physical one. To placate environmental concerns, the New Mexico campus pivoted from turbines to a roughly 2.45 GW Bloom fuel-cell microgrid estimated near \$8 billion, several billion dollars more than the initial generation concept. In spite of this, last week the state rejected the required 17-mile, 400 MMcf/d gas pipeline route for a second time. Every engineering workaround preserves schedule optionality but consumes balance-sheet capacity.

THE BUILDOUT IS ALSO CROWDING THE EQUIPMENT MARKET

Data center construction spending rose 23% year over year in May even as manufacturing construction fell 22%. Nonresidential materials costs are more than 55% above early-2020 levels, transformer prices are up about 70% in five years and some electrical-equipment backlogs extend beyond two years. AI campuses are not merely absorbing capital; they are competing for the same steel, copper, switchgear, contractors and skilled labor needed by the rest of the industrial economy.

That creates a feedback loop. Higher project costs force sponsors to seek deeper-pocketed equity, guaranteed power and milestone-based financing. Those requirements favor proven operators and hyperscaler-backed assets, while speculative queue positions become harder to finance. The physical shortage therefore amplifies the credit sorting already visible in the bond market.

META BECOMES A COMPUTE UTILITY

The 5 GW Louisiana build turns overcapacity from a risk into a merchant option

Meta's Louisiana campus is no longer simply a data center project. The physical build is moving toward 5 GW of compute at a cost above \$50 billion, while expected investment including chips is estimated above \$250 billion (according to Meta's recent announcements). Blue Owl owns an 80% stake in the project vehicle and is raising billions from Wall Street; Entergy is building 10 gas-fired plants to serve this new load and related transmission. More than 2 GW of additional electrical capacity is expected for the campus beyond the 5 GW compute load.

META COMPUTE: BUILD, BUY AND SELL



The option is the ability to allocate scarce compute among internal models, leased supply and external customers.

Sources: Bloomberg, New York Times and Wall Street Journal reporting; figures are announced or reported transaction values and may change.

The strategic novelty is the third box. Meta is discussing a potential two-year, \$10 billion compute lease with Anthropic while simultaneously buying \$21 billion of capacity from CoreWeave and \$27 billion from Nebius. That is not inconsistency. It is evidence that compute has become a traded strategic resource. The same company can be structurally short capacity in one period, long in another, and willing to monetize the spread.

Zuckerberg has described a cloud business as 'definitely on the table,' and companies reportedly approach Meta almost every week seeking models or spare compute. Hiring senior AWS executive Dave Brown to focus on infrastructure makes the organizational signal match the commercial one.

THE CAPITAL STRUCTURE IS PART OF THE PRODUCT

Blue Owl's 80% project stake and Entergy's dedicated generation build turn a corporate capex program into an infrastructure platform. Outside equity absorbs construction exposure; the utility owns and operates the power assets; Meta retains the compute capacity and the right to decide how it is used. The financing structure therefore preserves strategic control while reducing the amount of physical infrastructure that must sit directly on Meta's balance sheet.

That structure also creates a credible external product. A customer is not renting speculative GPU inventory from a thinly capitalized reseller; it is accessing capacity anchored by Meta's captive demand, dedicated power and long-duration infrastructure. The difference should show up in lease tenor, deposits, refinancing spreads and the willingness of lenders to underwrite expansion before every rack is assigned

OVERBUILD BECOMES AN OPTION, NOT A WRITE-OFF

Traditional data center underwriting treats unused capacity as a utilization problem. Meta can treat it as a portfolio option. Internal demand retains first call on the fleet; external leases monetize temporary slack; and merchant compute offers a price-discovery mechanism for the marginal accelerator-hour. The result is a hedge against both underbuilding and model-layer uncertainty.

\$250B+	\$125-145B	\$48B	\$10B
Estimated Louisiana site investment including chips	Meta 2026 capital-expenditure guidance	CoreWeave and Nebius compute purchases	Potential Anthropic lease over two years

Sources: Bloomberg; Wall Street Journal; New York Times; Meta guidance.

THE CREDIT DIFFERENCE MATTERS

Meta's merchant-compute option is not the same credit as a leveraged neocloud. It is backed by one of the world's largest free-cash-flow engines, a diversified advertising business and a physical fleet the company can consume itself. That changes lease bankability, equipment residual assumptions and refinancing risk. For developers, a Meta-backed compute platform can look more like investment-grade infrastructure than technology resale.

It also creates competitive tension. Meta can be a customer of CoreWeave, Nebius and data center developers while becoming a supplier to the same market. Neoclouds retain speed, specialization and customer neutrality; Meta brings balance sheet, captive demand and a lower cost of capital.

WHAT THE NEOCLOUDS CANNOT COPY

The hardest advantage to reproduce is not hardware access. It is the combination of captive utilization and investment-grade credit. Meta can redirect unsold capacity to ads, recommendation, model training or internal inference; a standalone neocloud must keep third-party customers on the machines. Meta can also finance through corporate cash flow, project equity and utility infrastructure rather than relying primarily on GPU-backed debt and repeated refinancing.

The counterargument is organizational. AWS, Azure and specialized neoclouds have years of experience selling service-level agreements, developer tools and enterprise support. Meta must prove it can operate a customer-facing cloud without compromising its internal priority queue or becoming an unwanted competitor to the partners from which it still buys capacity. Merchant compute is a valuable option before it is a mature business.

THE POWER ARCHITECTURE IS THE BUSINESS MODEL

A merchant-compute strategy works only if the underlying megawatts are controllable. Louisiana's structure links Meta-funded infrastructure, a dedicated utility generation build, outside project equity and long-duration compute demand. The physical campus is therefore both a production asset and a capacity inventory. Power is not merely an input; it is the commodity that makes the resale option possible.

The community economics are part of the same architecture. Meta has agreed to bear energy, water and related infrastructure costs, and construction-linked tax revenue has produced teacher bonuses of as much as \$50,000 in Richland Parish. Those benefits do not eliminate ratepayer and environmental scrutiny, but they show how a project can convert private infrastructure spending into a local political asset.

FOUR IMPLICATIONS

- 1. Hyperscalers become counterparties on both sides.** A developer may lease land or powered shell to Meta, compete with Meta for power and later sell capacity into a Meta cloud channel.
- 2. Excess capacity gains a market value.** The ability to lease compute reduces the penalty for building ahead of internal demand, which can support larger initial campus phases.
- 3. Credit migrates toward the balance sheet.** Projects supported by diversified hyperscalers should refinance better than capacity dependent on a single non-investment-grade reseller.
- 4. Power optionality becomes product optionality.** Sites with firm supply, modular expansion and flexible workload allocation can capture both internal AI growth and merchant-serving demand.

SCENARIO READ-THROUGH

SCENARIO	MARKET SIGNAL	INFRASTRUCTURE EFFECT
Internal demand outruns supply	Meta remains a net buyer; third-party leases stay expensive	Developers retain pricing power and Meta keeps building ahead
Balanced fleet with merchant sales	External customers absorb temporary slack	Meta cloud becomes a bankable new tenant category
Persistent overbuild	Lease prices fall and utilization disappoints	Campus phasing slows; weaker neoclouds reprice first
Platform expansion	Meta backstops third-party sites and sells capacity broadly	Compute finance converges with the Google TPU model

WHAT TO WATCH

The decisive evidence will be an executed Anthropic lease, a formal Meta cloud organization, disclosed pricing and service-level terms, and proof that external sales do not cannibalize higher-value internal workloads. Also watch whether Meta begins to finance third-party campuses with its credit signature, following Google's TPU backstop model.

BOTTOM LINE. Meta is building more than an AI fleet. It is assembling a two-sided capacity platform in which power, data centers and accelerators can be allocated dynamically between captive demand and external customers. The company that feared overbuilding is creating a business that can monetize it.

Sources: Bloomberg; Wall Street Journal; New York Times; Meta and Entergy disclosures.

WEEKLY ROUNDUP

NINETEEN STORIES, ONE PHYSICAL STACK

The week's top stories, itemized

1. PJM clears below reliability target: PJM's 2028/29 Base Residual Auction cleared at the \$325/MW-day FERC-approved cap and fell 6,831 MW below the reliability requirement — the second consecutive shortfall — while attracting only 525 MW of new generation and uprates. FERC Chair Laura Swett called the result another alarm bell and paired it with criticism of a stakeholder process that is too slow and opaque for the region's load-growth challenge. PJM will ask FERC to approve a special Reliability Backstop Procurement in September. The signal is not simply higher prices; it is that procurement is failing to convert price into timely new capacity. (Utility Dive, July 17)

2. DOE puts data centers into the national transmission case: The draft 2026 National Transmission Needs Study identifies accelerating data center construction, aging infrastructure and a shifting generation mix as simultaneous grid challenges. DOE estimates 2023 congestion costs at \$11 billion, finds the highest marginal value across regional seams and cites data center electricity projections exceeding 400 TWh by 2030. Transmission is becoming an AI industrial-policy asset, not a background utility function. (U.S. Department of Energy, July 2026)

3. New York imposes the first statewide pause: Gov. Kathy Hochul ordered a moratorium of up to one year on new data centers of at least 50 MW while the state develops rules for grid, environmental and community impacts. Projects still awaiting permits are exposed, and the administration also wants to examine sales-tax subsidies. The immediate load may move elsewhere; the larger effect is a template other states can adapt during an election cycle. (Wall Street Journal, July 14)

4. Opposition goes national: National organizers staged 142 protests across 42 states in the first coordinated nationwide action against data center development. A Reuters/Ipsos poll cited in the coverage found only 14% support for an AI data center in respondents' own communities. The coalition is ideologically mixed and focused on transparency, ratepayer protection, water, health and enforceable community benefits - making social license a bipartisan, project-level credit variable. (Reuters, July 18)

5. Palm Beach rejects Project Tango: After a 12-hour hearing, Palm Beach County commissioners voted 5-1 against the proposed Project Tango digital-infrastructure hub. Residents focused on water, power bills, noise and proximity to homes and a school, despite staff recommending approval with conditions. Florida law now gives local governments greater zoning authority and bars large data centers from shifting utility costs to customers. Local siting power is becoming decisive. (Bloomberg, July 15)

6. Pennsylvania requires load transparency: Pennsylvania's new budget requires data centers to report annual power and water use and directs PJM to give state regulators more insight into utility load forecasts. The law explicitly challenges the opacity of the process by which utilities submit demand inputs. For developers, this raises the evidentiary standard around peak load, phasing and project credibility; for PJM, it inserts state oversight into a forecasting function with major capacity consequences. (Utility Dive, July 14)

7. Oracle reaches the edge of investment grade: S&P cut Oracle to BBB-, one notch above junk, as the company funds a \$250 billion data center expansion while burning more cash than it generates. Credit protection reached a record high, and S&P estimates a possible \$42 billion free-cash-flow deficit. Project Jupiter adds physical execution risk: a costlier 2.45 GW fuel-cell design and a second rejection of its gas-pipeline route. The capex race now has a visible credit boundary. (Bloomberg / The Information, July 16-17)

8. Google takes the TPU war outside Google Cloud: Google is pitching TPUs to Nvidia-centric neoclouds, discussing financial backstops and offering to rent chips back from third-party operators. The strategic shift turns TPU efficiency from an internal Gemini cost advantage into a merchant product. Open-weight models such as Kimi K3 strengthen the logic because hosts can choose silicon based on delivered-token economics. The power-market read-through is a new class of TPU tenant behind someone else's meter. (The Information, July 13)

9. TSMC and ASML say the hardware cycle is still tight: TSMC reported a 36% quarterly sales increase and plans close to \$56 billion of capital spending, while SK Hynix expects memory shortages beyond 2030. ASML raised its annual sales forecast to EUR43-EUR45 billion as chipmakers expand leading-edge capacity. Equity markets may question AI returns, but the upstream equipment chain is still reporting scarcity. The physical buildout has not yet received the demand-destruction memo. (Bloomberg, July 13-15)

10. Mitsubishi buys the Haynesville optionality: Mitsubishi closed its \$7.5 billion acquisition of Aethon's U.S. gas and pipeline assets, its largest purchase ever and the biggest Japanese entry into American shale. The portfolio is concentrated in the Haynesville, positioned between Gulf Coast LNG exports and Louisiana/Texas data center demand. The transaction is a long-duration bet that gas molecules serving both sovereign energy security and behind-the-meter AI power will command strategic value. (Fortune, July 14)

11. SLB and Liberty package oilfield speed for data centers: SLB will provide prefabricated infrastructure modules and Liberty Energy will supply gas-fired generation under a global data center partnership. SLB expects deliveries for data center projects to exceed 2 GW by year-end, while Liberty targets roughly 3 GW of power projects by 2029. The oilfield-services proposition is deployment speed, modularity and integrated project management - capabilities developed in a volatile industry now being redirected toward AI load. (Bloomberg, July 14)

12. Valar Atomics tests venture-scale nuclear finance: Small-reactor developer Valar Atomics is reportedly seeking \$1 billion at a roughly \$5 billion pre-money valuation, with Sequoia in talks to lead a debt-and-equity round. The valuation follows a power milestone and reflects capital's willingness to price late-decade firm-power optionality aggressively. The caveat remains timing: nuclear can become strategically important to data centers before it becomes a material source of near-term energized capacity. (The Information, July 16)

WEEKLY ROUNDUP

13. Finland wins a 550 MW AI campus: Oaktree-backed Pure Data Centres began a EUR1.5 billion, 110 MW first phase of a Finnish campus that can exceed 550 MW and EUR7.5 billion of investment. The initial phase is fully leased and Microsoft is expected to be a customer. The project captures Europe's sovereign-compute push and the movement toward remote, power-rich sites, while showing that political engagement, land, energy and financing can require years even in supportive markets. (Bloomberg, July 14)

14. U.A.E. chip access becomes a geopolitical reward: The United States moved the U.A.E. into a more favorable export category, allowing G42 to buy advanced chips more freely for at least nine months and easing licensing constraints on planned Microsoft and OpenAI data centers. The decision links military alignment, energy security and AI infrastructure directly. Advanced accelerators are now diplomatic currency, and sovereign compute locations can change quickly with trade classification. (Wall Street Journal, July 14)

15. Washington considers a FINRA-like AI watchdog: The Trump administration is weighing an independent body under the SEC to vet leading AI models with industry participation, modeled loosely on FINRA. The proposal tries to give labs more predictable safety standards after ad hoc restrictions while addressing Wall Street's cybersecurity concerns. Model regulation matters to power because release delays, access controls and compliance rules can change the geography and timing of inference-serving demand. (Bloomberg, July 17)

16. Pennsylvania land becomes the new mineral right: Ninety-six Salem Township families sold roughly 1,700 acres to QTS for \$586 million, averaging about \$330,000 per acre and \$5.5 million per family. The value was not in the soil but in adjacent transmission, substations and nearby gas and nuclear generation. Another 200 sellers are pursuing a \$1.3 billion transaction next door. Power adjacency is capitalizing rural land at infrastructure-platform values. (Wall Street Journal, July 13)

17. Orbital data centers remain a late-decade option: Bain estimates space-based data centers will account for almost none of global compute in 2030, about 3.5% in 2035 and 5% in 2040. A 100 MW orbital network would currently cost roughly 50% more per kilowatt than a terrestrial facility, largely because of launch and infrastructure requirements. Economics may improve with Starship-class reuse, but terrestrial power and permitting remain the main battlefield this decade. (Bloomberg / Bain, July 17)

18. Green debt becomes a social-license instrument: Data center entities have issued \$186 billion of sustainable debt since late 2022 as QTS, Compass, AirTrunk and others use green bonds and loans to document efficiency and renewable-power commitments. The financing benefit is modest; the reputational function is larger. With at least 48 projects worth \$156 billion blocked or stalled in 2025, lenders increasingly need evidence that environmental and community promises are measurable and enforceable. (Bloomberg, July 15)

19. Thinking Machines joins the open-weight race: Mira Murati's Thinking Machines Lab released Inkling, a 975-billion-parameter open-weight model with only 41 billion parameters active per query. The company explicitly optimized cost against performance and paired the model with its Tinker fine-tuning platform. The release reinforces the week's central silicon theme: more capable models are becoming portable, customizable and price-sensitive, expanding the addressable market for non-Nvidia serving architectures. (Wall Street Journal, July 15)