Goldman Sachs Investment Banking

Powering the AI Era

Contents

Dan Dees
Co-Head of Global
Banking and Markets

Economic progress is rarely linear—throughout history, it's been punctuated by technology-driven inflections.

The inexorable forces of finance and technology have shaped our world for centuries—dismantling barriers and forging new industries.

In 19th century America, railroads catalyzed commercial activity by unlocking new markets, spawning industries like mail-order retail, and enabling Midwestern farmers to transport crops overseas. A century later, the internet ushered in a digital economy: lowering the cost of hosting and transmitting data, linking global markets, and giving rise to e-commerce. Today, we stand at the dawn of another defining era, driven by artificial intelligence—and the proliferation of foundation models begetting new industries altogether.

But while innovation sparks these new eras, capital and infrastructure are necessary to power them. Steam engines marked a historic invention, but thousands of miles of track ultimately unlocked their economic potential. During the dot-com era, $800B+[1] invested in critical internet infrastructure like fiber-optic cables, broadband, and servers enabled a digital economy through faster and cheaper transmission of information. In turn, financial markets evolved to meet these unprecedented demands. Investment banks first emerged during the railway age to mobilize global sources of capital. In the dot-com era, a combination of venture capital funds, active IPO and M&A markets, and broadening credit markets unlocked the requisite growth capital for internet infrastructure.

Similarly, the future of AI will not be forged in code and large language models alone. It will be built with concrete, steel, and silicon. The average cost to bring a typical 250 MW AI data center online is roughly $12B[2] inclusive of the equipment inside—requiring innovative financing solutions to fuel this growth.

That said, a lack of capital is not the most pressing bottleneck for AI progress—it's the power needed to fuel it. After a decade of flat demand growth, global data center power demand is expected to surge +160% by 2030[3] as a result of AI workloads that run on energy-intensive graphics processing units (GPUs). And the current grid was not designed for this future. Transmission and permitting timelines for natural gas plants stretch 5–7 years, renewable sources like wind and solar can only provide intermittent power as it stands, and nuclear is a longer-term solution. Soaring power demand is currently being met with marginal increases in power supply—stifling AI development by limiting data center activity. Unlike previous infrastructure buildouts, the accelerated pace of innovation requires immediate solutions.

It's impossible to accurately predict how the next decade will unfold, but AI is an economic force that will permeate every industry and geography. This convergence of compute and power is creating new urgency. Corporate leadership should be thoughtful, strategic, and opportunistic. Goldman Sachs has been at the epicenter of technological inflection points since our founding—innovating and iterating to provide the capital solutions necessary for fueling progress. This moment is no different, and we are energized to help today's innovators leave their mark.

Dan Dees
Co-Head of Global Banking and Markets

19TH CENTURY

Railroads

80%

share of US stock market[4]

215K

miles of track by 1900[5]

Railroads revolutionized commerce by turning localized markets into a national economy. The first investment banks were formed to channel global capital into railway infrastructure through a nascent corporate bond market.

20TH CENTURY

Electrification

$295B

capital raised by utilities (1920–1930)[6]

50 GW

capacity added to power grid (1920–1930)[7]

Electricity transformed productivity through innovations like Henry Ford's moving assembly line. To meet soaring power demand, utilities pioneered the holding company structure to access cheaper debt financing and unlock economies of scale.

20TH/21ST CENTURY

Telecom & Dot Com

$800B+

investment in infrastructure (ex: fiber-optic cables)[8]

39M

miles of fiber-optic cables laid by 2001[9]

The infrastructure laid to support the internet birthed a digital economy and still pays dividends today. To drive progress, a combination of venture capital funds, active IPO and M&A markets, and broadening credit markets unlocked the requisite growth capital.

PRESENT

Artificial Intelligence

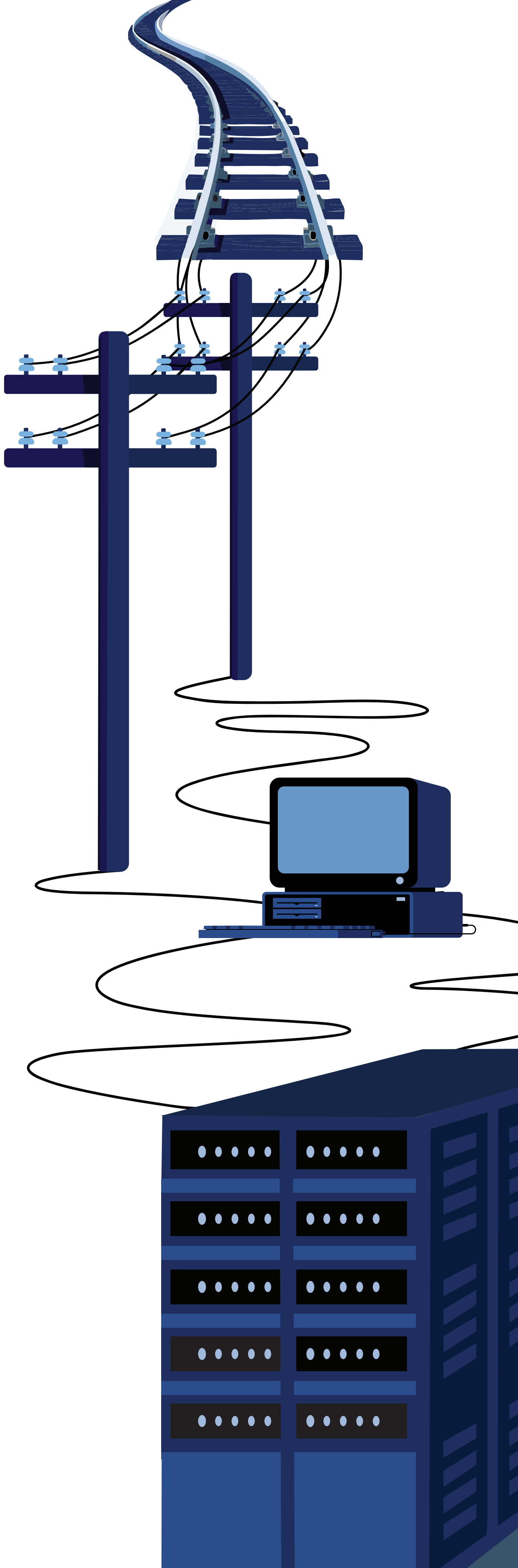
+160%

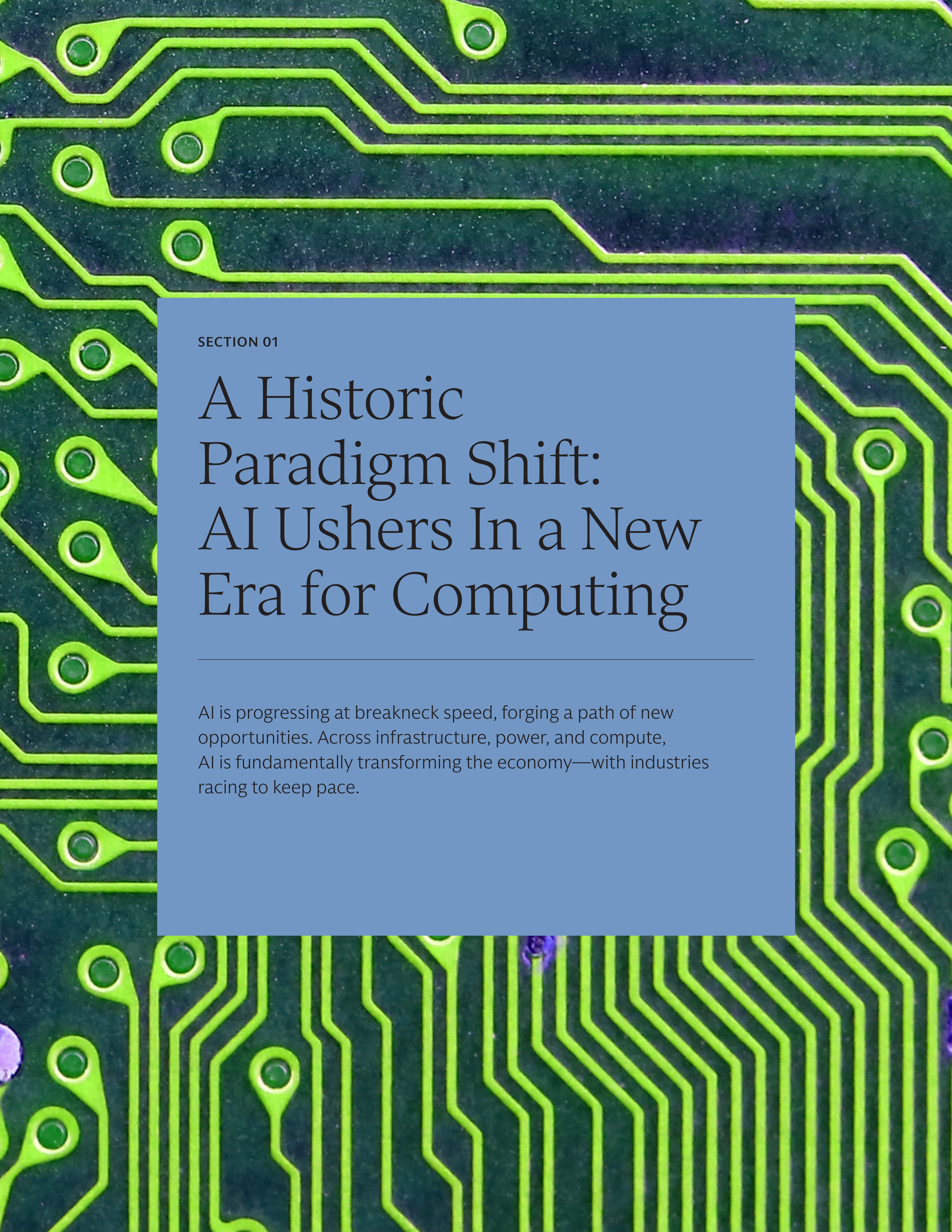
data center power demand growth by 2030[3]

$5T

funding required for digital infrastructure & power[10]

The future of AI is still being written but is already creating new business models and industries. To unlock AI's full potential, innovative financing solutions across public and private market sources are needed to meet unprecedented capital demands.





SECTION 01

A Historic Paradigm Shift: AI Ushers In a New Era for Computing

AI is progressing at breakneck speed, forging a path of new opportunities. Across infrastructure, power, and compute, AI is fundamentally transforming the economy—with industries racing to keep pace.

Before the mainstream adoption of generative AI, cloud computing was the most recent paradigm shift in digital infrastructure—with companies transitioning their storage and workflows from on-premises servers to inexpensive servers housed in external data centers at massive scale. Companies like Amazon, Google, Meta, and Microsoft quickly dominated this category by “hyperscaling” resources and data to efficiently process user workloads—lowering costs and resource utilization. Thousands of cloud data centers were built to meet surging demand—operating on traditional computer processing units (CPUs), large-scale storage arrays, and standard networking infrastructure to handle client workloads.

Yet even as cloud migration accelerated, data centers’ power demand remained flat due to efficiency gains that kept consumption in check. Those efficiency gains have largely been captured, and the infrastructure needed to run AI workloads is exponentially more complex and resource intensive. In 2027, AI server racks will require 50x more power than cloud equivalents five years ago,[11] and GPUs are significantly more energy intensive—requiring intricate liquid cooling systems to offset excessive heat generated by dense GPU clusters.

Source: Nvidia 2025 GTC Keynote Presentation

Hyperscalers are at the forefront of AI and are projected to invest $1T in the technology by 2027,[12] as “scaling”—the assumption that AI improvements require ever-more data and compute resources—drives CapEx. Outside the traditional cloud hyperscalers, other companies are establishing themselves as “neoclouds”—spending billions of dollars on GPUs and data centers. These investments are concentrated in training more advanced models, which inherently requires larger, power-hungry data centers capable of handling these AI workloads. Despite these significant investments, cloud computing spend has not slowed. AI isn’t cannibalizing cloud budgets—it’s expanding them.

AI training—teaching models to recognize patterns and generate responses—is particularly energy intensive, necessitating data center campuses with high-powered GPUs, dedicated power supply, and advanced cooling systems. Importantly, location is less crucial for training models because their development does not require interaction with end users. Instead, training infrastructure is being built in regions with abundant land, access to power, and favorable regulatory conditions—including less-populated US states such as Iowa and Nebraska, along with the Nordics and parts of Southeast Asia, where climate conditions support natural cooling.

Under the Trump administration’s Stargate initiative—a joint venture between OpenAI, Oracle, and SoftBank, that plans to invest up to $500B developing digital and energy infrastructure[13]—the first Stargate data center will be built in Abilene, Texas. This area is energy-abundant and population-scarce, offering inexpensive power and existing data centers that can be repurposed for AI.

"When we think about the scale of what's happening with AI, the cloud era offers one of the closest parallels. Cloud required an enormous buildout of computational and physical infrastructure. What did that deliver? Nearly a trillion dollars of new revenue and the emergence of new companies, platforms, and applications."

Jung Min | Global Co-Head of Technology, Media, and Telecom, Investment Banking

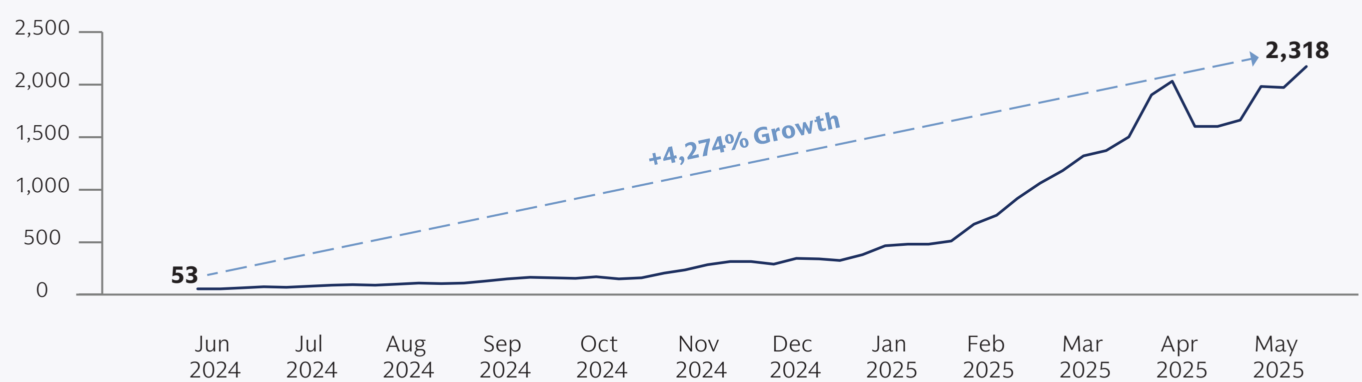
The CapEx playbook for data centers may soon evolve.

Inferencing—where trained and released models interact with real-world users (e.g., ChatGPT)—is not as compute hungry. However, speed and latency are paramount, requiring proximity to end users and adding complexity and risk to decisions around data center builds. Breakthroughs in AI have been difficult to anticipate—as evidenced by the launch of DeepSeek—posing a risk that data centers built for training in remote locations will become less relevant if the transition to inferencing is faster than expected.

Regardless, data center demand is rapidly outstripping supply. Data center vacancy rates sit at a record low 3%, and near 0% in the most sought-after markets—with new power at scale often not coming online until 2028 or beyond.[14] Accordingly, the total data center development footprint now exceeds 50 million square feet—double the volume of five years ago.[15] Even ignoring power constraints, the capital needs for producing the necessary supply are unprecedented. Companies spent $800B+ on internet infrastructure during the dot-com era[16]—hyperscalers are expected to surpass that figure by 2027.[17]

With AI progressing at breakneck speed, financing AI data centers is increasingly complex. For cloud data centers, most costs were related to the "shell" (i.e., power, land, physical infrastructure)—not the servers and CPUs inside—and financing mirrored traditional real estate in using the shell and MEP (mechanical, electrical, plumbing) equipment as collateral. With more sophisticated and expensive hardware, the compute equipment inside AI data centers costs 3–4x[18] more than the physical data center itself. The pace of innovation also creates uncertainty around the enduring relevance of this hardware—a challenge for lenders and investors underwriting data centers. The next generation of AI innovators face challenges in financing their massive compute requirements—simultaneously powering their own model development and providing cost-effective compute to their customers. Unlike established hyperscalers with investment-grade ratings and robust balance sheets, these emerging AI labs operate with startup financial profiles. Innovative financing structures and creative credit enhancements are critical to reducing their cost of capital and maintaining competitiveness.

Demand for Compute Is Soaring (Total Tokens Consumed per Week)



Source: OpenRouter as of May 26, 2025

New Challenges Require New Players and Financing Solutions

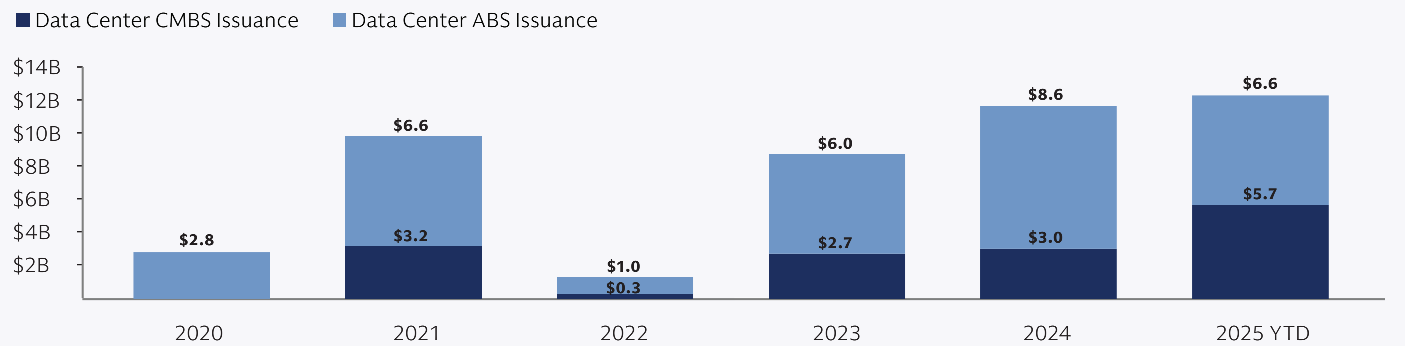
When building new data centers, hyperscalers typically own the process themselves—securing the power and land—or outsource the heavy lifting to third-party developers who provide plug-and-play solutions. To keep pace with record demand, these developers are tapping creative financing solutions—including joint ventures, borrowing base facilities, construction loans, and project finance.

After securing the powered land and tenant leases, banks are more willing to lend—with private credit funds providing additional funding via mezzanine debt and/or preferred equity to help developers stretch their equity further.

Joint ventures are also helping raise development capital—in 2024, the American REIT Equinix formed a joint venture with Singapore's sovereign wealth fund and the Canadian Pension Plan Investment Board to raise $15B for building and expanding the footprint of US hyperscale data centers.[19] Developers are also creatively recycling capital by securing development funding and long-term financing up front. Crusoe, the developer for Stargate's first data center, formed a fully funded forward takeout venture with Blue Owl Capital and Primary Digital Infrastructure to fund construction and permanent financing.[20] Once leased and cash-flowing, owners will often refinance a data center into a single-asset/single-borrower (SASB) CMBS/ABS or private placement—bonds secured entirely by the completed data center.

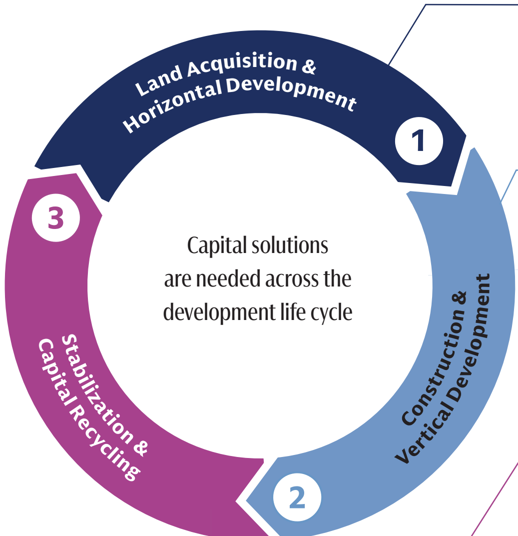
By packaging a fully leased ("stabilized") data center into SASB CMBS/ABS or private placement bonds, owners replace short-term construction loans with more permanent forms of financing that cut interest expenses by 50–100 bps. Additionally, with institutional investors (i.e., insurance funds, pension funds, etc.) considering stabilized data centers less risky, they accept a lower cap rate—boosting the property's appraisal value and unlocking 25% of cash that can be recycled into funding the next data center. The first data center SASB priced in 2021, but the product now represents 13% of the SASB market,[21] and multiple $2B+ transactions have occurred in 2025.[22] Hyperscalers, meanwhile, are leaning on joint ventures from their own balance sheets: seeding a project with an anchor lease or power contract with outside investors supplying most of the equity.

Data Center CMBS and ABS Issuance Continues to Accelerate



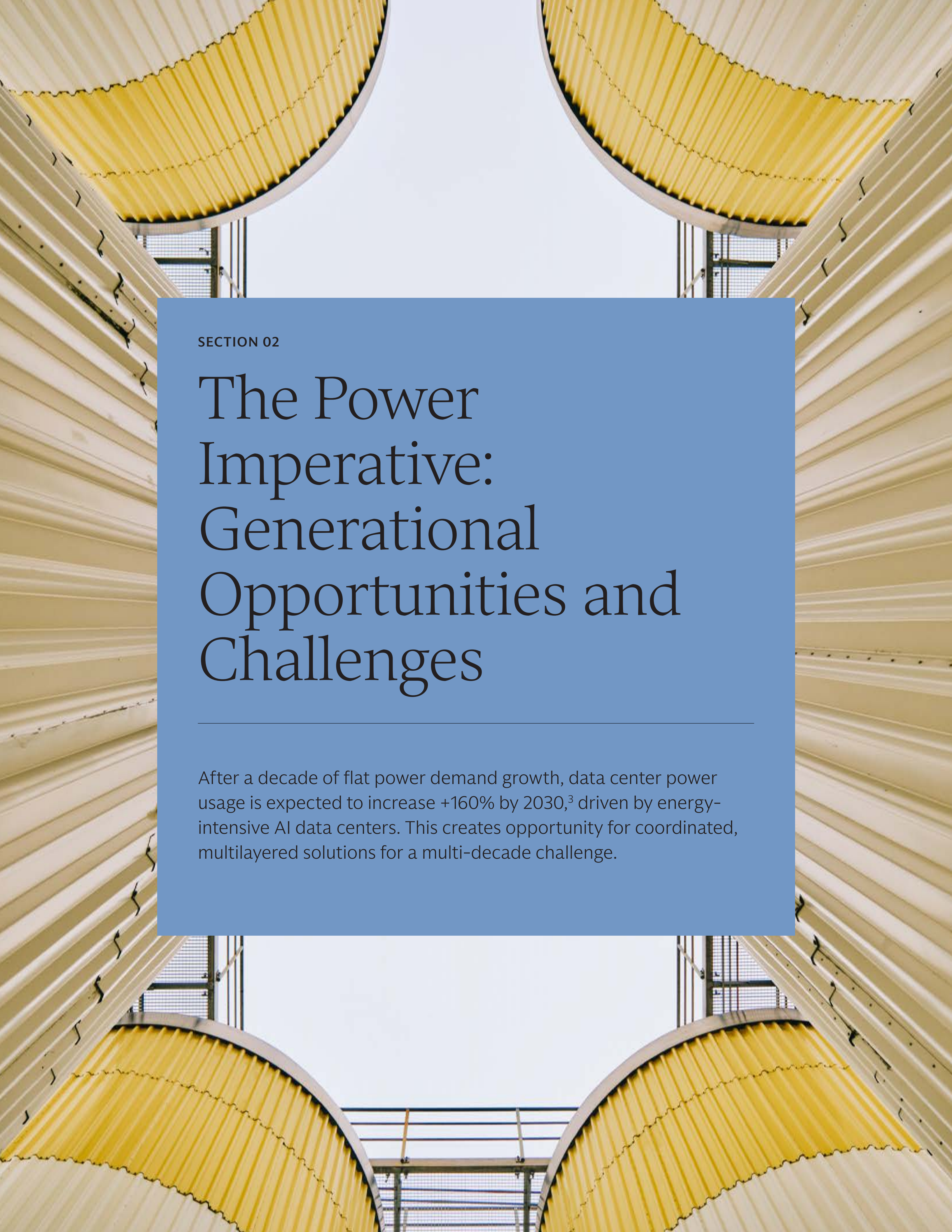
Source: Green Street as of June 24, 2025, Finsight as of June 24, 2025

A diverse set of investors—from private equity and private credit firms to real estate funds and endowments—want exposure to data center development, but are financing each stage of the life cycle in pieces according to risk tolerance, which increases complexity and execution risk. Alternatively, operating under one contract throughout the development life cycle reduces friction costs while providing capital cost certainty—structuring solutions to provide construction financing and permanent capital on day one. This visibility enables developers to accelerate timelines and commit to larger projects with confidence.



Capital	Key Players	Key Debt Products
Mostly Equity	**Equity:** Private equity, endowments, family offices **Debt:** Private credit	Corporate revolvers, borrowing base facilities, land and pre-development loans
Debt + Equity	**Equity:** Value-add infra funds, value-add RE funds, private equity, sovereign wealth, pension funds **Debt:** Large banks, private credit, insurance funds	Construction loans, project finance, DevCo facilities, borrowing base facilities
Debt + Equity	**Equity:** Core funds, pension funds, insurance funds, net lease funds **Debt:** Insurance funds, pension funds, asset managers	Corporate revolvers, borrowing base facilities, CMBS, ABS, private placement, term loans

But the most critical obstacle for unleashing AI's potential is not capital—it's power. Global power demand for data centers is expected to rise +50% by 2027—60% of that growth will need to be met by *new* capacity—and +160% by 2030.[3] The ability for hyperscalers and data center operators to continue recycling capital and fund new development is key.



SECTION 02

The Power Imperative: Generational Opportunities and Challenges

After a decade of flat power demand growth, data center power usage is expected to increase +160% by 2030,[3] driven by energy-intensive AI data centers. This creates opportunity for coordinated, multilayered solutions for a multi-decade challenge.

The technologies underpinning AI are advancing at breakneck speed—a stark contrast with the US power grid, composed of infrastructure assets that average 40 years old.[23] This structural mismatch has become a critical bottleneck that both industries are racing to resolve.

Challenges facing power and utility companies cannot be solved with a "move fast and break things" ethos. Expanding power capacity requires navigating regulatory hurdles, federal and state permitting processes, supply chain challenges, and a power grid unprepared for rapid demand growth. The ambitions of hyperscalers, AI companies, and data center operators are sobered by this governor on growth.

For over a decade, the United States enjoyed access to an abundance of low-cost natural gas, declining cost of wind and solar installation, historically low interest rates, and overall flat power demand, which kept electricity prices low. The combination of low wholesale power prices and the influx of intermittent wind and solar generation—creating the now infamous "duck curve"—undermined the economics of 24/7 baseload coal and nuclear plants and accelerated their retirement across the grid.

Today, the grid lacks sufficient baseload capacity to meet projected peak demand growth, a challenge that will require faster permitting for new projects, more flexible demand, and smarter planning. A silver lining? Rising price signals and longer lead times for traditional generation components are spurring R&D into new technologies that will help address the imbalance and make better use of existing infrastructure. Necessity is the mother of invention: we are seeing progress across carbon capture, distributed generation adoption, longer-duration storage solutions, and demand flexibility.

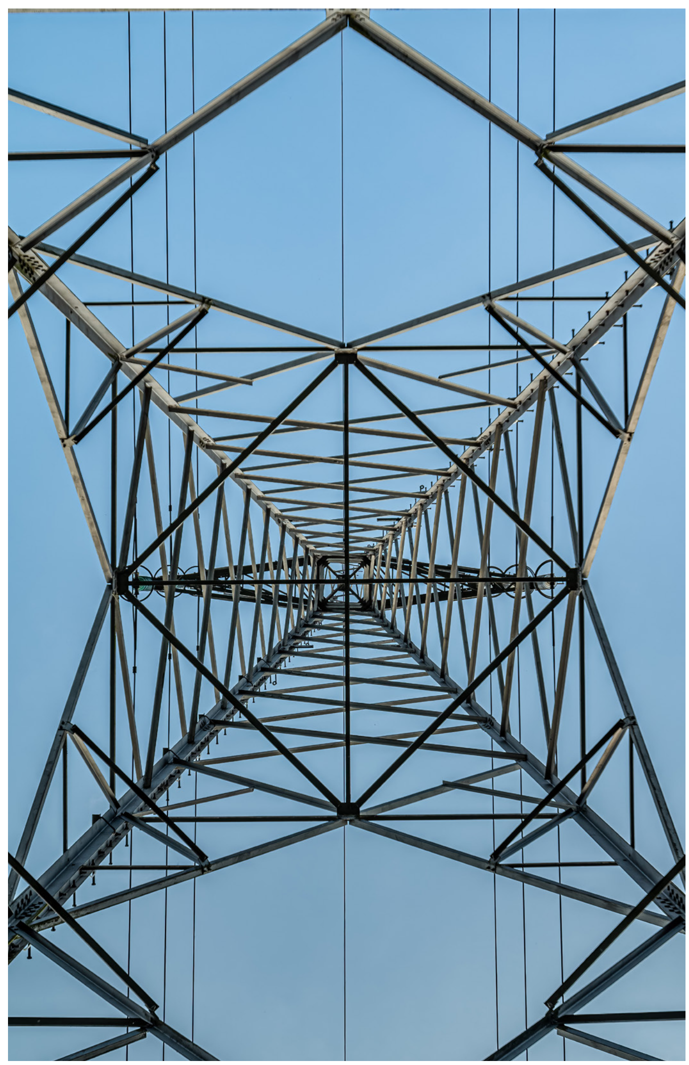
A New Era for . . . Utilities?

Regulated US utilities are facing an existential dilemma: how to reconcile their public service mandate with aging infrastructure in need of investment—and power demand that is rapidly surpassing traditional growth models. Utility sector CapEx has more than doubled over the last decade—not just to support load growth, but to harden the system against more extreme weather, ensure extreme reliability, and replace aging infrastructure nearing the end of its useful life. Capital spending across the sector is now at unprecedented levels, reflecting the scale and urgency of the energy transition.

Many utility companies are also wary from previously spending heavily on new assets that are then underutilized. DeepSeek and potential breakthroughs in chip efficiency have stirred questions on whether this historic power demand pans out. AI leaders, however, have underscored that efficiency breakthroughs won't offset the growing need for compute as enterprise applications, cloud services, and agentic AI reach mass adoption.

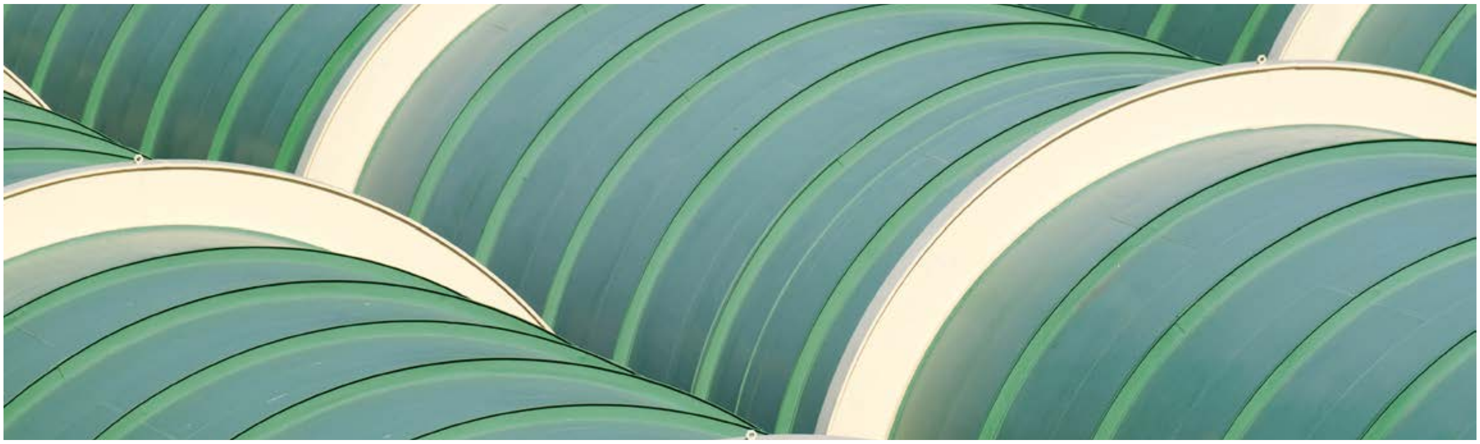
Regardless, power companies are increasingly seeking new rate structures for large-load customers like hyperscalers and data centers that mitigate stranded asset risk—exploring mechanisms such as take-or-pay contracts, up-front capital contributions, and long-term capacity commitments.

Partnerships are emerging between power and technology companies—like Entergy's deal with Meta—to co-develop generation and transmission assets that ensure reliable, long-term power for data centers. These ambitious projects, however, are not without consequences, as regulators assess the burden on local ratepayers—the average US electricity bill rose 23% from 2019–2024. These partnerships are being structured to protect existing ratepayers, ensuring that the cost of serving large, energy-intensive customers is borne by those customers.[24]



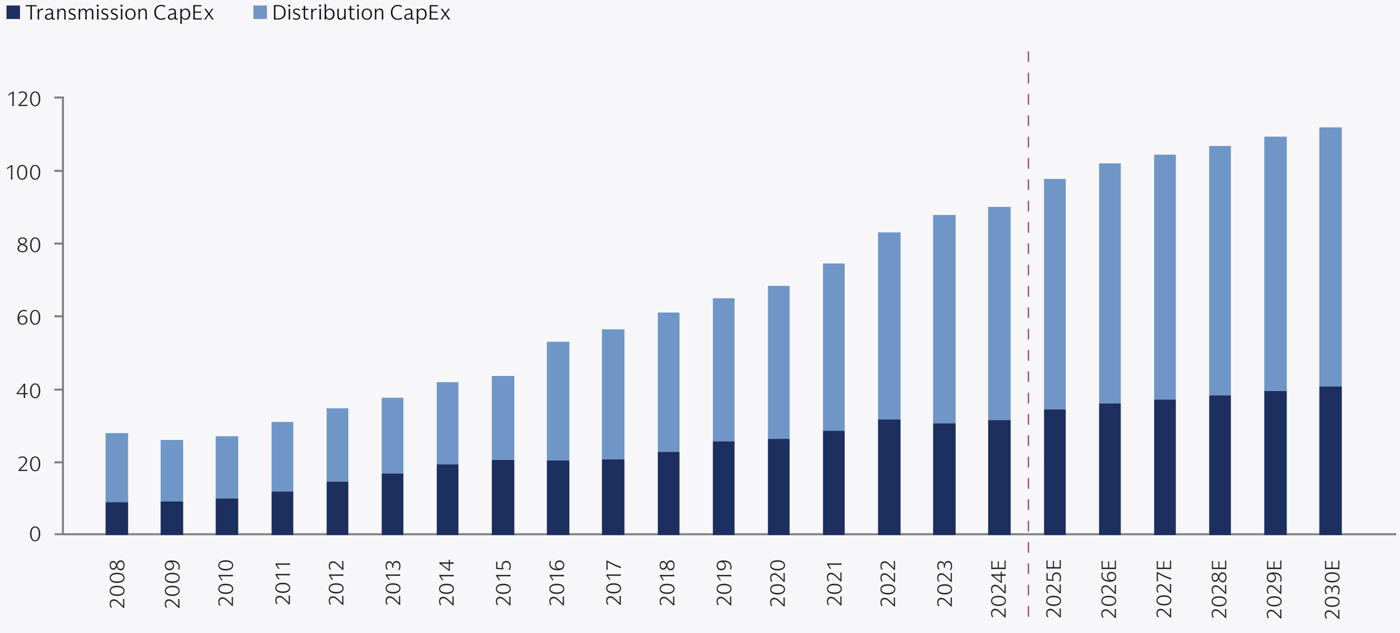
"Nuclear power is particularly exciting, with several developers advancing small modular reactors (SMRs) and fusion technologies following substantial investments in making these innovations commercially viable. That said, long lead times and high costs mean utilities, hyperscalers, and OEMs will require bespoke capital solutions and possible government support to fund and price new technology development risks."

Tyler Miller | Global Co-Head of Power, Utilities, and Infrastructure, Investment Banking



The generation of power is only half the equation, with transmission an equally important bottleneck for getting new plants online. With an abundant supply in the United States, most data centers are powered by natural gas—but permitting, transmission, and critical supply chain challenges (e.g., gas turbines) have resulted in 5–7 year timelines for getting new natural gas plants online and connected to the grid. This resurgence in natural gas, fueled by AI momentum, has already spurred dealmaking. In January 2025, Constellation Energy, the largest producer of clean, emission-free energy in the United States, announced its acquisition of Calpine Corporation, America's largest generator of electricity from natural gas, for $29.1B.[25] This strategic transaction will create the nation's largest clean energy provider at a time when power demand is surging.

More than $700B of Grid Investment Expected Through 2030 in the US



Source: Company data, EEI, Goldman Sachs Research

Federal policy will be critical for relieving these permitting delays. The Trump administration has voiced commitment to US power grid improvements, but full realization will require support from both sides of the aisle. The bipartisan Energy Permitting Reform Act, which was introduced in the last Congress, would grant the federal government authority to permit interregional transmission projects. As stress on the power grid intensifies, we expect support for this bill and similar proposals to garner greater attention from policymakers.

"The era of flat power demand is over, with data centers representing the first major tailwind of growth in the power sector for many, many years. Today, we are on the cusp of megaprojects and infrastructure spending on a scale not seen in decades."

Rebecca Kruger | Partner, Natural Resources, Investment Banking

A Multi-Decade Problem with Multilayered Solutions

These permitting delays and supply chain challenges are forcing hyperscalers and data center operators to simultaneously address short-term (five years) and long-term (10–15 years) power constraints. The unique challenges of today's landscape require a combination of strategic partnerships, creative financing, and new methods for sourcing power. Zooming out, long-term solutions like nuclear power still require significant investments and CapEx today that won't pay off for years—introducing another layer of complexity.

Hyperscalers Are Expected to Source Data Center Power via Multiple Sources

	CAPACITY	ADVANTAGES	CHALLENGES
Solar		carbon footprint	land requirements, intermittent power
Onshore Wind		carbon footprint	land requirements, intermittent power
Nuclear Large Scale		small land footprint, reliable, carbon footprint	supply chains, waste, labor, enriched uranium supply, lead time, legacy public fears
Nuclear SMR*		small land footprint, reliable, carbon footprint	supply chains, waste, labor, enriched uranium supply, lead time, legacy public fears
Battery Storage		enables greater clean energy reliability	capacity limits
Natural Gas CCGT**		reliable, small land footprint	carbon footprint
Natural Gas Peaker		reliable, small land footprint	carbon footprint, less carbon efficient than CCGT
Grid (Natural Gas)		reliable, land footprint	carbon footprint, interconnection wait times

Source: GS Research as of April 2025—"GS Datacenter supply / demand model update: Loosening earlier than expected; still tight by historical standards" and "Generational Growth: AI, data centers and the coming global power surge"

*small modular reactor **combined-cycle gas turbines

Goldman Sachs Research estimates that ~60% of data center demand growth needs to be met with new capacity, likely powered by 30% natural gas CCGT, 30% natural gas peakers, 27.5% solar, and 12.5% wind.[26] Facing extended timelines for getting new natural gas plants online, renewables remain the fastest and most efficient method for securing incremental power supply. However, wind and solar power only function when the weather cooperates, and battery technology is currently unable to store power long enough for reliable baseload power. While renewables are an important part of today's landscape, they do not provide the 24/7 capacity data centers need. That said, these dynamics could change with advances in battery technology like iron-air and/or sodium-ion battery systems that would extend discharge times, improve efficiency, and lower costs.

60%

of data center power through 2030 needs to be met with new capacity.[26]

Going Nuclear

Hyperscalers are pursuing combinations of power sources to meet short-term demand, and investing in long-term solutions like nuclear energy—but with caution. Tech companies have been willing to invest, but have carefully avoided development risk or asset ownership—preferring to accelerate progress via solutions like forward-start power purchase agreements (PPAs). Alphabet recently signed an agreement with Elementl Power to pre-position three sites for advanced nuclear energy.[27] Others are reviving shuttered reactors where the infrastructure is already in place—Microsoft struck a 20-year PPA with Constellation Energy to restart the 835 MW Unit 1 reactor at Three Mile Island,[28] and NextEra Energy is in the process of recommissioning Iowa's Duane Arnold plant.[29]

The advantages of nuclear are clear. It provides reliable, 24/7 zero-carbon baseload generation—making it a critical asset for decarbonization and grid stability, and an ideal match for data centers that demand uninterrupted, high-capacity power to support mission-critical and AI-driven operations.

Excitement around a potential nuclear renaissance is growing, but the history of cost overruns and delays challenge traditional financing models. Vogtle Unit 3, which came online in 2023, was estimated to be over $17B,[30] more than double the original budget, and roughly seven years behind schedule. The silver lining: it proved the efficacy of Westinghouse's AP1000 nuclear technology, which materially benefited Vogtle Unit 4 a year later through cost and execution improvements.

There is optimism more units will follow—but policy support is needed to absorb early-stage cost and timeline risks until the industry reaches standardized builds at scale. In May 2025, President Trump signed four executive orders aimed at accelerating the deployment of nuclear energy in the United States and improving supply chain issues—aiming for 400 GW of nuclear capacity by 2050, vs. the 100 GW currently operating.

Small modular reactors (SMRs) are emerging as an option for reliable, carbon-free power—and hyperscalers are actively exploring investing in them, and/or taking the power via long-term PPA, as a potential solution to meet the massive, 24/7 energy demands of AI and data center growth.

This power imperative is a multi-decade problem with multilayered solutions.

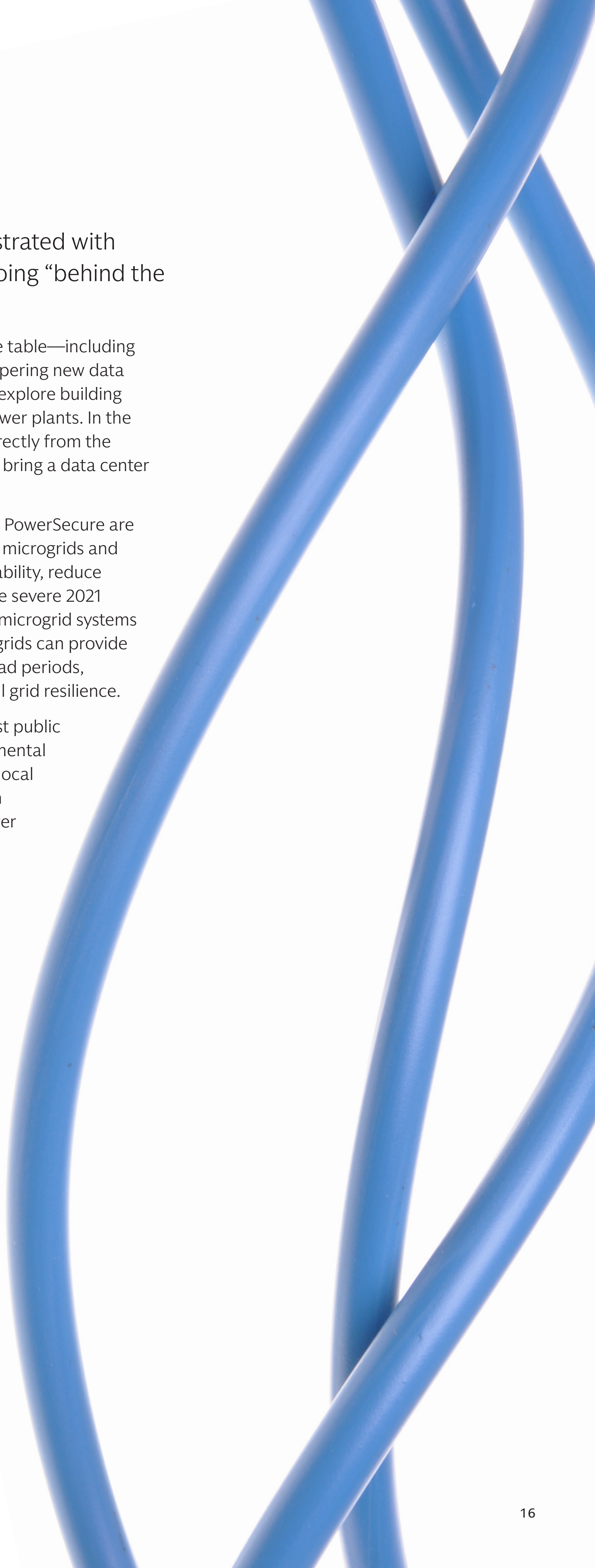
Going Behind the Meter

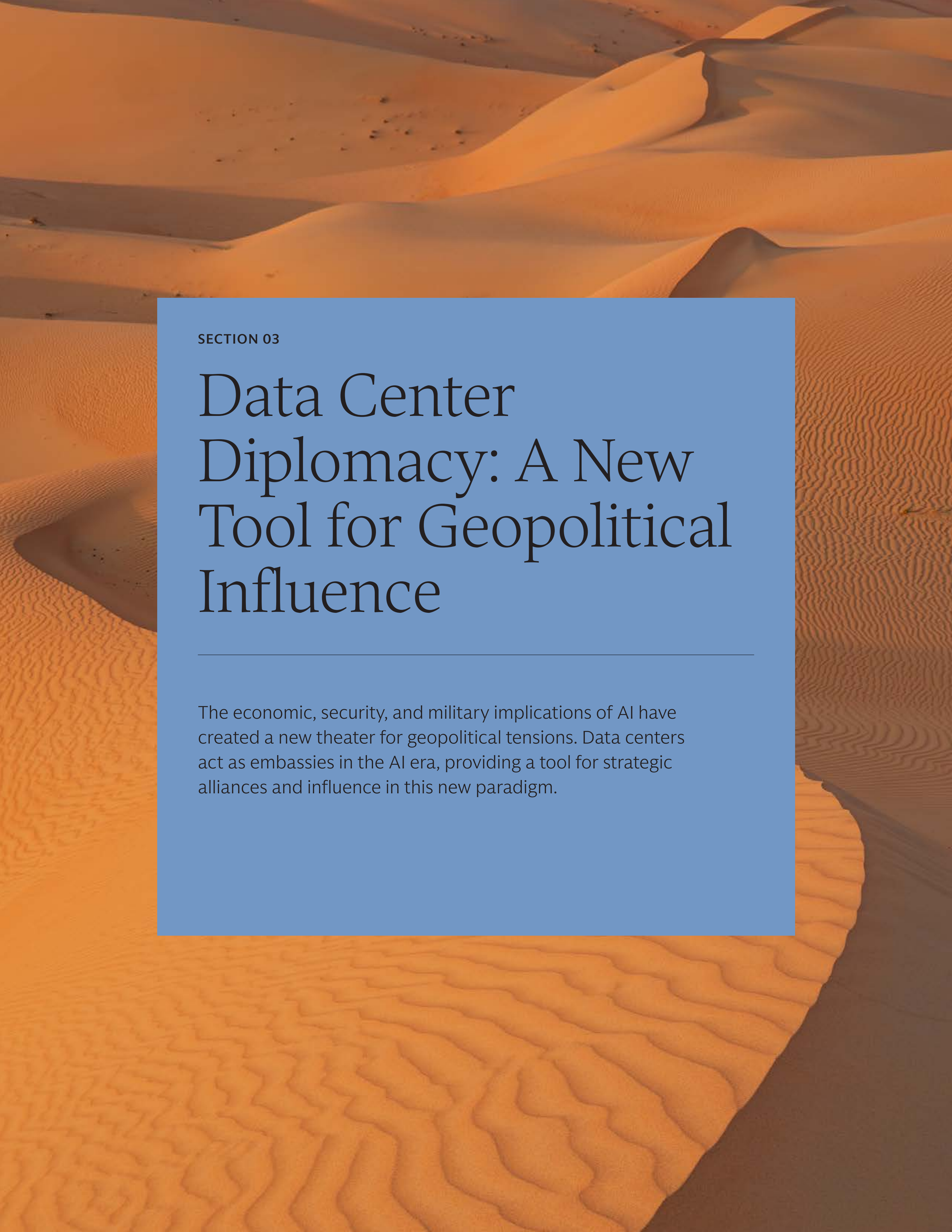
Many tech companies and data center developers frustrated with long lead times are taking an alternative approach—going "behind the meter" and acting as their own power supplier.

For data centers in urgent need of baseload power, all options are on the table—including those beyond the US power grid. Extended interconnection queues hampering new data centers' ability to secure reliable power have pushed many operators to explore building microgrids directly on-site—or locating data centers close to existing power plants. In the latter scenario, hyperscalers purchase and receive their power supply directly from the generation source, not the utility grid—significantly reducing the time to bring a data center online or scale up its power capacity.

Additionally, companies like Solaris Energy Infrastructure, VoltaGrid, and PowerSecure are offering turnkey distributed energy solutions—from solar and storage to microgrids and on-site generation—to help utilities and large energy users enhance reliability, reduce emissions, and manage grid constraints amidst rising demand. During the severe 2021 winter storm in Texas, when the state's power grid failed, PowerSecure's microgrid systems provided more than 2.26 GWh of reliable energy. Grid-connected microgrids can provide peak shaving by reducing demand on the broader system during high-load periods, lowering and stabilizing prices for local ratepayers, and enhancing overall grid resilience.

However, behind-the-meter solutions are not without controversy amidst public debates around cost burdens for local ratepayers and potential environmental concerns. At xAI's Project Colossus facility in Memphis, for example, the local community has publicly complained about the pollution emanating from the site. Other plans to colocate data center campuses with nuclear power plants have been blocked by the Federal Energy Regulatory Commission over concerns that the development would raise consumer prices.





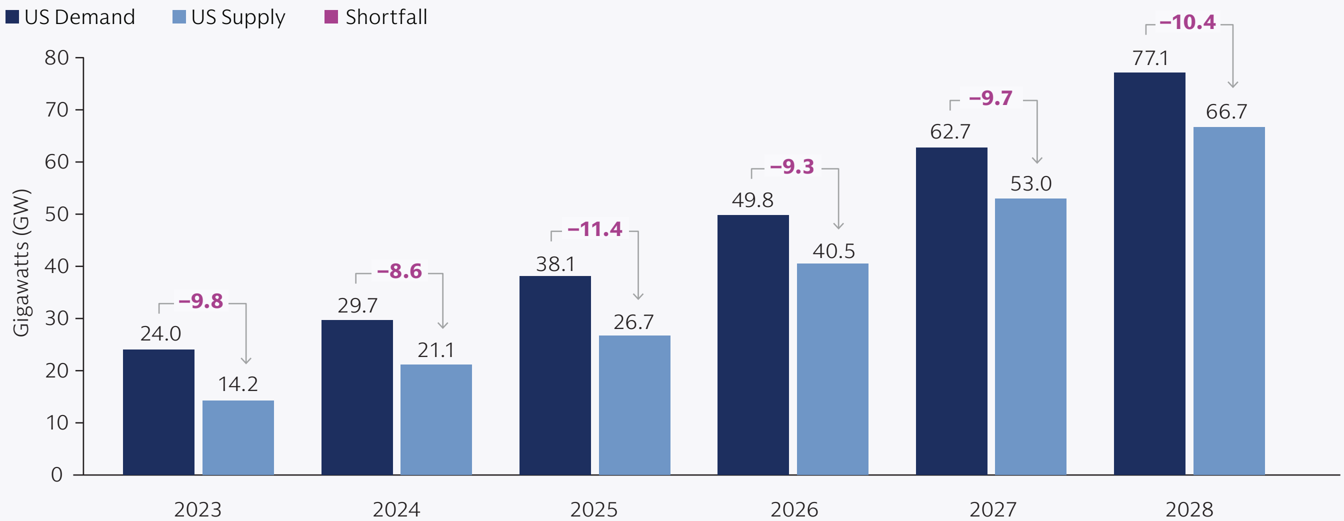
SECTION 03

Data Center Diplomacy: A New Tool for Geopolitical Influence

The economic, security, and military implications of AI have created a new theater for geopolitical tensions. Data centers act as embassies in the AI era, providing a tool for strategic alliances and influence in this new paradigm.

While most of the hyperscalers' data center development activity has been centered in the United States, other regions are rapidly scaling their capacity. Data is often referred to as the "new oil" in the digital economy, but there is a key distinction. Unlike oil reserves, which are naturally determined by geography, data centers can be strategically built in locations chosen by businesses and governments. This flexibility allows nations to leverage data center infrastructure as a critical geopolitical and economic tool in the AI era.

Shortfall of Supply vs. Demand for Data Center Capacity



Source: Wall Street broker estimates—JLL, CRBE, JP Morgan and C&W

The complexity of technology demands international cooperation. The United States already relies on a network of global commercial partners like China, Taiwan, and the Netherlands due to specific technologies in which those countries specialize. Outside of the underlying technologies, real estate for digital infrastructure provides another avenue for strategic partnerships. If the United States faces domestic bottlenecks in data center expansion due to regulatory, financial, or logistical constraints, it will need to establish overflow options with global partners for training models—where location and latency matter less. Through carefully planned AI infrastructure investments, nations can strengthen alliances, enhance economic competitiveness, and assert influence in the evolving digital economy.

Countries working with companies to host AI data centers gain economic, political, and technological advantages, but data centers also present national security sensitivities. They often house high-end, export-controlled semiconductors utilized by governments, businesses, and everyday users to transmit some of their most sensitive information.

Founders, hyperscalers, and investors are already thinking globally. Amazon, Google, and Microsoft have all focused investments in Middle East data centers. Latin America is also emerging as a potential hub, and with expectations that the data center market will double in the next five years,[31] billions of investment dollars are expected in Brazil alone—where ~90% of power is renewable. While the global expansion of data centers offers significant growth opportunity, there is also significant risk. Some data centers are being built in regions that already struggle with power access, raising concerns about straining local infrastructure and furthering inequities.

These alliances are also sensitive to policy changes. For example, as digital policymaking comes to the forefront for government leaders globally, proposed tax reforms and data privacy could create headwinds for US technology companies, potentially limiting operational flexibility. Global regulations and sanctions limiting the number of American-made chips that can be distributed to certain countries further complicate efforts to expand data center hubs, adding layers of geopolitical and supply chain risk.

SECTION 04

Meeting the Moment with Capital Solutions

AI is reshaping dealmaking, underwriting, and capital flows—requiring new financing solutions and an expanded universe of capital market participants to meet unprecedented demand.

In 2024, global hyperscalers' CapEx equaled ~$800M[32] per day as the race for artificial general intelligence (AGI) rapidly accelerated. In parallel, the US utilities sector collectively spent $200B in annual CapEx[33]—a marked jump from figures in recent years. As the scale of infrastructure and associated funding needs continue to grow, efficiently sourcing and deploying capital will be a key determinant of success.

Strategic partnerships, alongside creative combinations of public and private capital, are unlocking opportunities in this new landscape. Joint ventures in particular can leverage the unique expertise and capital resources of diverse players to drive efficiency and scale—with collaborations between public pension funds, financial sponsors, sovereign wealth funds, and data center operators gaining momentum. For example, in 2024, American REIT Equinix, the Canada Pension Plan Investment Board, and Singapore's sovereign wealth fund GIC formed a joint venture aiming to raise $15B for building and expanding the footprint of US hyperscale data centers.[34]

Financial sponsors are sitting on more than $4T[35] of dry powder, with a desire to return capital amidst a slowdown in exits and pressure on distributions to paid-in capital (DPI). AI presents a unique opportunity to strategically invest in an infrastructure expansion that offers a self-reinforcing cycle of growth and returns throughout a data center's life cycle.

$12T
Expected capital demand for energy transition by 2030[36]

$3T
Expected capital demand for power and utilities by 2030[36]

$2T
Expected capital demand for digital infrastructure by 2030[36]

The AI ecosystem is currently being financed in parts, with different capital solutions for each segment—from the land and power to the underlying chips. But as the scale of data centers compounds, clients will need more comprehensive solutions that provide greater certainty in the availability and pricing of capital.

Resolving AI Bottlenecks with Long-Term Capital

Global data center supply has increased from 30 GW in 2019 to 57 GW in 2024[37]—equaling $324B of investments, assuming costs of $12M/MW. By 2030, another 65 GW of capacity is expected to come online, and the scale of each data center is massive—Meta alone is currently working on a 2 GW facility.[38] To keep pace with record demand, hyperscalers and developers need effective strategies for generating liquidity from stabilized data centers and expanding the universe of capital participants beyond traditional markets.

Develop equity capital around stabilized data centers: Traditional lower-cost-of-equity capital solutions for investors seeking exposure typically offer higher yields than government bonds, and are considered relatively stable due to their fixed revenue streams. Emerging data center and real estate core funds are also attracted to this opportunity set. Another public market option, the $68B Singapore REIT (S-REIT) market, has become a popular destination for listing REITs holding global data center assets due to favorable tax and regulatory treatments. Alternatively, the public markets offer another viable path, as shown recently in the $1.3B IPO of Australian data center operator DigiCo.

Expand investor base by optimizing risk allocation: Creative and bespoke financing structures can help access large, untapped pools of capital. For example, joint ventures can align the large-scale funding needs of digital infrastructure with insurance companies that need to pair long-term liabilities with long-term investments. Longer contract terms, removing early termination rights, and ensuring triple net lease structures could make stabilized and/or developing data centers a suitable investment for insurance firms, retirement plans, and other long-term sources of capital.

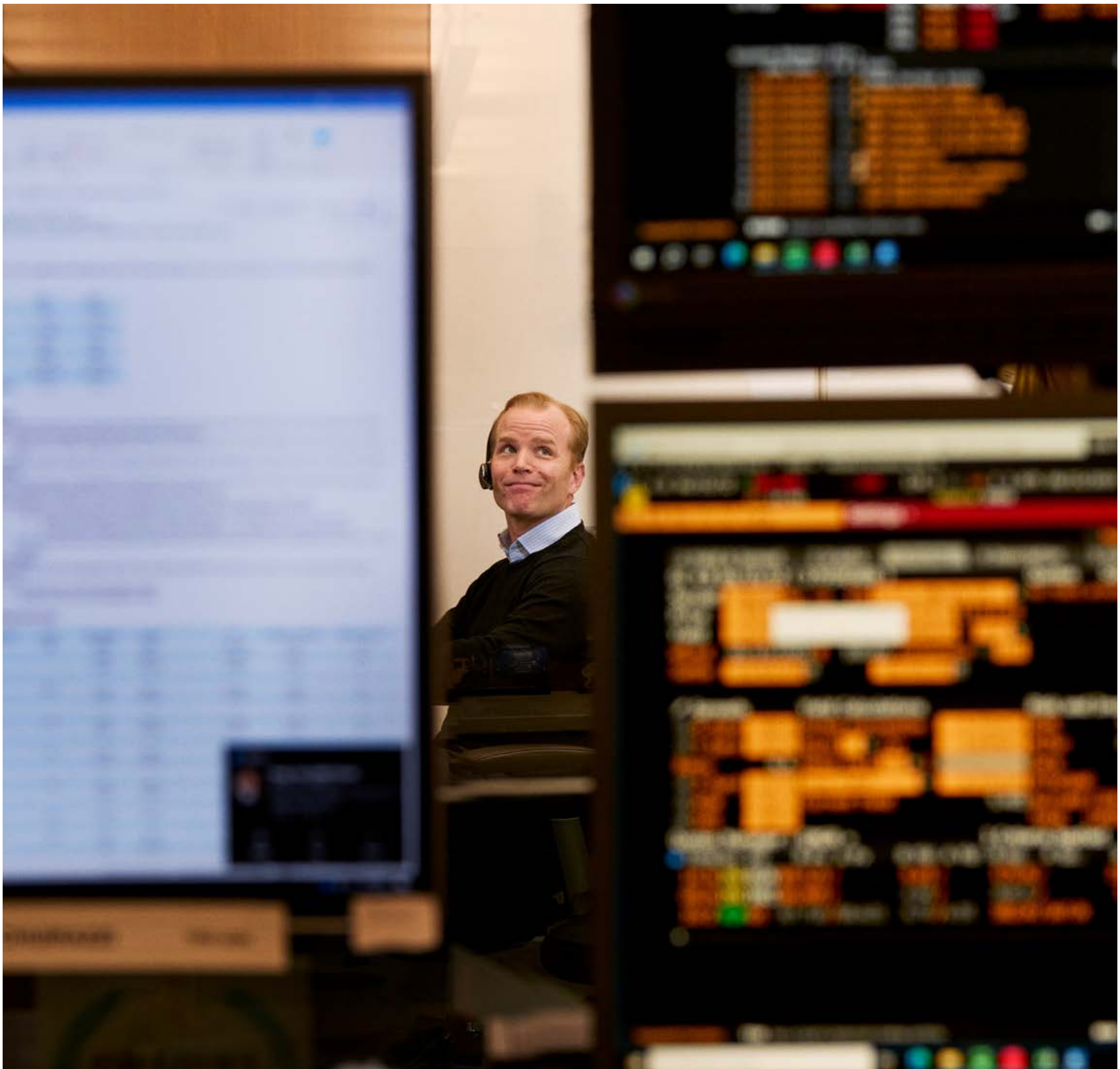
“From public cloud hyperscalers to emerging players in AI, hyperscalers are evolving from infrastructure consumers to enablers of capital formation. Their continued support can unlock under-invested forms of capital in digital infrastructure such as insurance and 401(k)s—lowering their own cost of growth while broadening access to the AI economy.”

Jason Tofsky | Partner, Technology, Media, and Telecom, Investment Banking

The New CapEx Playbook: Strategies and Shifts in the AI Era

Historic energy and capital demands require combinations of both private and public markets that leverage the strengths of each. Public market financing offers greater efficiency, deeper liquidity, and tighter pricing—and private markets excel in bespoke solutions. But the current pace of activity will also bring a more efficient, standardized approach to private capital deployment over time, reducing friction costs and the illiquidity premium as a result.

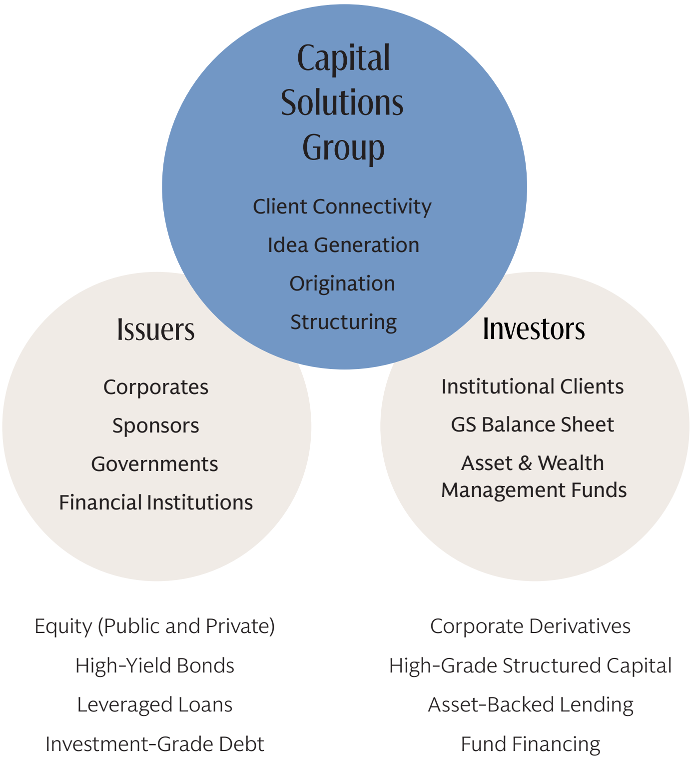
As the emergence and growth of private credit and other private asset classes materially shifts financing markets in real time, staying ahead requires agility, perspective, and access.



Capital Solutions Group: Partnering with Corporates and Sponsors to Fund the Future

Goldman Sachs is positioned at the fulcrum of these tectonic industry and market shifts.

Our Capital Solutions Group, formed in early 2025, is underpinned by the world's #1 advisory franchise, a leading origination business, and a world-class investing platform that deploys capital across liquid and alternative asset classes. This integrated unit allows for both more holistic oversight of investment ideas and related capital—and the strategic deployment of that capital across key channels.



Spotlight: High-Grade Structured Capital Solutions

Significant, low-cost private capital sources from insurers and infrastructure funds can offer non-debt, non-dilutive capital to IG corporates for a range of strategic and financing objectives applicable to a wide range of industries, segments, and assets with stable and recurring cash flows (e.g., data centers).

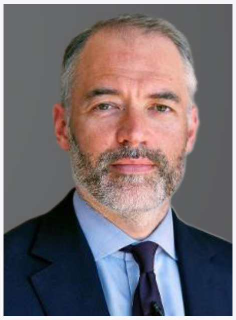
This capital is cheaper than traditional forms of equity, but at a premium to unsecured debt. Structures can be tailored to unique situations and market conditions while allowing for shared resources between partners and lowering individual parties' risk.

Our preeminent global Investment Banking business is built on a culture of enduring partnerships and a commitment to exceptional execution—helping clients seize new opportunities to unlock growth and transformation.

Global Investment Banking Leadership

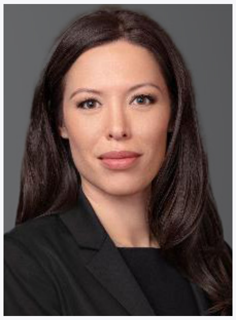
MATT MCCLURE

Global Co-Head of Investment Banking



ANTHONY GUTMAN

Global Co-Head of Investment Banking; Co-CEO of Goldman Sachs International



KIM POSNETT

Global Co-Head of Investment Banking

PETE LYON

Global Co-Head of Capital Solutions Group

MAHESH SAIREDDY

Global Co-Head of Capital Solutions Group

Contributors

JOHN GREENWOOD
BRIAN HAUFRECT
KYLE JESSEN
GEORGE LEE
REBECCA KRUGER
TYLER MILLER
JUNG MIN
CHRISTINA MINNIS
JASON TOFSKY
MIRIAM WHEELER
EMILY BAKER
JOHN ELEOTERIO
DAN KEYSERLING
MATT LUCAS
KARA MANGONE
EDUARDO MANZUR
RYAN NEWMAN
GEOFFREY OKAMOTO
SALVADOR PAREJA
ACHAL SAKARIA
MATTHEW SIMONETTI
ALEX TINGLE
BERTIE WHITEHEAD
TIM WHITEHEAD
STEPHANIE WISNER
FRANK LONG
MILA MILATOVIC
DAMIEN PETZOLD
LARKYN SINCLAIR
JONATHAN TAN
MAHESH BHIDE
JAMIE CATHERWOOD
LUCAS GREENBAUM
PETER HAGERTY
CAROLINE BROOKS

Endnotes

[1]Telecommunications Industry Association. Inflation adjusted.

[2]Based on average construction costs of $12M/MW. Compute equipment typically costs 4x the cost of physical "shell" and MEP. Cushman & Wakefield, Data Center Development Cost Guide 2025

[3]GS Research

[4]Finaeon

[5]Library of Congress

[6]National Bureau of Economic Research

[7]Boston University

[8]Telecommunications Industry Association. Inflation adjusted.

[9]The New York Times. June 20, 2001

[10]Estimates based on GIR reports: "Carbonomics: The GS net zero carbon scenarios—a reality check"; "Generational growth: AI, data centers and the coming US power demand surge"; Power & Utilities figure globalized based on a US estimate of ~$1.4T. Estimates from Apollo Investor Presentation (all estimates for the next 10 years): Energy Transition: ~$30–50T+; Power & Utilities: ~$30T+; Digital Infrastructure: ~$15–20T+.

[11]https://www.goldmansachs.com/insights/articles/rising-power-density-disrupts-ai-infrastructure

[12]New Street Research as of August 9, 2024

[13]https://openai.com/index/announcing-the-stargate-project/

[14]Public press releases, UBS, datacenterHawk, WallStreet Research, CBRE Investment Management

[15]Newmark, 2025 U.S. Data Center Market Outlook

[16]Telecommunications Industry Association. Inflation adjusted.

[17]New Street Research as of August 9, 2024

[18]Corporate reports, 650 Group, Mercury Research, Yale, FactSet, and NSR estimates and analysis

[19]Press release. October 1, 2024

[20]Press release. October 15, 2024

[21]Green Street as of May 31, 2025

[22]Green Street

[23]Reuters

[24]https://www2.deloitte.com/us/en/insights/industry/power-and-utilities/funding-growth-in-us-power-sector.html

[25]Press release. January 10, 2025

[26]GS Research

[27]Press release. May 7, 2025

[28]Press release. September 20, 2024

[29]American Nuclear Society

[30]https://apnews.com/article/georgia-nuclear-power-plant-vogtle-rates-costs-75c7a413cda3935dd551be9115e88a64

[31]Latin America Data Center Market Landscape 2025–2030

[32]FactSet. Global hyperscalers includes Meta, Amazon, Alphabet, Microsoft, Oracle, Tencent, Baidu, Alibaba.

[33]S&P Global Ratings

[34]Press release. October 1, 2024

[35]Preqin as of September 30, 2024

[36]Estimates based on GIR reports: "Carbonomics: The GS net zero carbon scenarios—a reality check"; "Generational growth: AI, data centers and the coming US power demand surge"; Power & Utilities figure globalized based on a US estimate of ~$1.4T.

[37]https://www.goldmansachs.com/insights/articles/ai-to-drive-165-increase-in-data-center-power-demand-by-2030

[38]https://www.reuters.com/technology/meta-invest-up-65-bln-capital-expenditure-this-year-2025-01-24/

Disclaimer

Goldman Sachs Investment Banking