Artificial Intelligence and U.S. Electricity Demand: Trends and Outlook to 2040

Summary

Artificial Intelligence (AI) is emerging as a major driver of electricity consumption in the United States. From energy-hungry model training clusters to ubiquitous AI-driven services, the growth of AI-related computing is beginning to reshape electricity demand patterns. This report provides an in-depth, evidence-based assessment of how AI is impacting U.S. power consumption today and how it is projected to influence demand through 2040. We examine historical trends in computing electricity use, forecast the rise of AI-related loads (in data centers, industry, and at the edge), and analyze implications for generation mix, grid infrastructure, and decarbonization goals. The findings draw on reputable sources – including U.S. government agencies, national labs, industry consultancies, and utilities – to give energy professionals and decision-makers a clear, data-driven outlook.

Historical Context: Computing and Electricity Use in the U.S.

Early data centers and computing infrastructure were a niche contributor to U.S. power demand, but their footprint expanded rapidly with the internet boom. U.S. data center electricity use roughly doubled from 2000 to 2005 amid a surge in server installations. By the mid-2000s, policymakers grew concerned that unchecked growth of digital infrastructure could significantly strain power systems. Indeed, U.S. data centers' energy consumption increased nearly 90% in the first half of the 2000s. An EPA report to Congress in 2007 warned that if trends continued, data centers would become a major grid burden [1][2][3].

However, around 2010 a dramatic efficiency turnaround occurred. Widespread virtualization, improved cooling, and IT hardware efficiency allowed computing output to rise without commensurate growth in energy use. From 2010 to 2014, total U.S. data center electricity consumption grew only about 4% (reaching ~70 TWh in 2014, ~2% of national power use). This plateau was a huge departure from prior years – efficiency gains essentially "bent the curve," preventing a potential explosion in energy use. One analysis found that continued efficiency improvements from 2010 onward saved about 620 billion kWh of electricity by 2020, compared to where consumption would have been if 2010-era technologies had remained static [2].

By the mid-2010s, cloud computing and hyper-scale data centers (run by firms like Amazon, Google, Microsoft) began to dominate the landscape, enabling greater efficiency at scale. Yet demand for digital services accelerated so much in the late 2010s that absolute energy use began rising again. Total U.S. data center power consumption climbed from 58 TWh in 2014 to roughly 176 TWh in 2023. In other words, data center electricity use tripled over the past decade, reaching about 4.4% of total U.S. electricity consumption in 2023. The surge is attributed to explosive growth in internet traffic, cloud services, and recently AI computing workloads that require power-intensive server farms. Notably, companies like Google and Microsoft each more than quadrupled their electricity usage from 2016 to 2023, citing expansions in AI and data centers as key drivers. In short, after a period of relative stability, computing-related electricity demand is again on a steep upward trajectory entering the 2020s [4][5][6].

AI-Driven Electricity Demand Projections (2025–2040)

Near-Term (2025–2030): The advent of modern AI – especially energy-intensive machine learning model training and inference – has triggered a wave of new data center construction and load growth. Both industry analysts and government forecasts now anticipate an extraordinary rise in electricity demand attributable to AI and digital infrastructure in the next decade. The U.S. Department of Energy reports that data center energy use could double or even triple from 2023 levels by 2028 to roughly 325–580 TWh. This implies data centers alone might draw 6.7% to 12% of all U.S. electricity by 2028, up from ~4% today. Consulting firm Bain & Company likewise finds that after nearly two decades of stagnant load growth nationally, the late-2022 breakthrough in generative AI and the ensuing data center boom “blindsided” utilities and is driving an unprecedented upswing in demand. Bain estimates data centers will account for ~44% of all growth in U.S. power consumption from 2023 to 2028. In absolute terms, U.S. utilities may need to supply 7% to 26% more generation by 2028 than in 2023 just to meet data center and AI-related loads, far beyond any five-year increase in recent memory [4][7].

Multiple projections coalesce around a similar range of outcomes by 2030. The Electric Power Research Institute (EPRI) warned in 2024 that data centers could consume about 9% of U.S. electricity by 2030, roughly double their current share. This is equivalent to an annual usage of ~400–450 TWh in 2030 (up from ~180 TWh in 2023). S&P Global Market Intelligence compiled utility forecasts showing data center demand rising from ~185 TWh in 2023 to about 440 TWh in 2035. For context, an increase of that magnitude (~250 TWh) over a decade is comparable to adding the electricity consumption of an entire large state like California to the grid. McKinsey & Company analysis suggests U.S. data center loads could grow ~23% annually through 2030, adding roughly 400 TWh of demand by 2030. In fact, McKinsey expects data center capacity expansion for AI will constitute between 30–40% of all net new electricity demand in the U.S. through 2030. These forecasts align in painting the latter 2020s as a period of accelerating growth, where AI and digitalization emerge as top contributors to load expansion alongside electrification of transport and industry [7][8][9].

Globally, the pattern is similar. The International Energy Agency (IEA) estimates electricity consumption from data centers (including AI and crypto applications) could double between 2022 and 2026. Data centers worldwide used about 460 TWh in 2022 and are on track to exceed 1,000 TWh by 2026. While electrification of vehicles and industrial processes will be larger drivers of global demand growth, AI data centers are significant new loads, especially in advanced economies and Chinaiea.org. In the U.S., which is home to many hyperscale AI computing facilities, near-term demand projections have been repeatedly revised upward. The Energy Information Administration’s 2025 load growth forecast was adjusted eightfold higher within the span of a year due to emerging data center developments. Clearly, the scale and speed of AI-driven electricity uptake in the next 5–10 years carries substantial uncertainty, but the direction is unequivocally toward rapid growth [6][10].

Long-Term (2030–2040): Looking further out to 2040, forecasts necessarily become more uncertain, but scenario analyses suggest that AI and computing could help drive U.S. power demand to levels not seen in decades. For instance, McKinsey posits that under a high-AI adoption scenario, total U.S. electricity consumption in 2040 might reach about 6,900 TWh –

roughly 75% higher than today's usage. (For comparison, U.S. retail electricity sales in recent years have been ~3,800–4,000 TWh.) This scenario assumes sustained growth in data center loads alongside electrification of transportation and other sectors, resulting in a near-doubling of power demand by 2040. Utility planners are starting to grapple with such possibilities: Pacific Gas & Electric, California's largest utility, stated that statewide power demand could roughly double by 2040 largely due to "artificial intelligence, electric cars and other efforts to electrify more of the economy". A high-demand case from the National Renewable Energy Lab (NREL) similarly found that U.S. electricity use could diverge far above historical trends – the electricity intensity of the economy (electricity per unit GDP) could rise by 20+% by 2040 (reversing a long decline) if emerging electrification and digitalization trends take off [6][9][11].

Specific to data centers and AI, fewer published studies extend beyond 2030, but many experts anticipate continued (if moderating) growth. One utility consortium projection sees U.S. data center energy consumption climbing to ~8–12% of total electricity by 2030, with further increases by 2040 depending on efficiency improvements and saturation of AI deployment. If AI-driven services and automation permeate most sectors by 2040, it is conceivable that hundreds of gigawatts of new computing load (on par with the scale of today's entire power fleet) may need to be supplied. On the other hand, there is a potential for leveling off: once major companies have built out AI infrastructure and models, growth could stabilize in the 2030s, especially if offset by breakthroughs in energy efficiency or constraints on expansion (economic or regulatory). Thus, long-range outlooks often use scenario ranges. For example, EPRI's analysis suggests U.S. data centers might be as "low" as ~4.6% or as high as 9% of power use by 2030, depending on whether efficiency gains keep pace. Extrapolating those scenarios to 2040 yields a wide band of outcomes – a reflection of the sensitivity to key assumptions around AI adoption rate and technology progress (discussed further in a later section) [4][8][12].

In summary, all evidence indicates that AI and associated data infrastructure will be a dominant source of U.S. electricity demand growth in the coming two decades. Even under conservative cases, dozens of gigawatts of new load are headed for the grid; under aggressive cases, the U.S. could experience demand growth on the order of 50–100% by 2040, reversing the stagnation of the 2010smckinsey.com. This trajectory will have far-reaching implications for how utilities plan generation, transmission, and distribution [9][11].

Sector-Specific Demand Growth: Hyperscalers, Industrial AI, and Enterprise

Not all AI-related electricity demand is created equal – growth is concentrated in certain sectors and facility types. The hyperscale cloud providers (like Amazon Web Services, Google Cloud, Microsoft Azure, Meta, etc.) account for the bulk of the increase, as they race to build large AI training clusters and serve global cloud AI needs. Hyperscale data centers – often campus-style facilities 100 MW or larger – are proliferating to support AI services. The IEA notes that annual capital investment by the top U.S. tech firms in data centers doubled in the last two years, with Google, Microsoft and Amazon spending more on data infrastructure in 2023 than the entire U.S. oil & gas industry's capital expenditures. This massive investment (over 0.5% of U.S. GDP in data centers) underscores how cloud hyperscalers are "all in" on expanding compute capacity for AI. These companies are deploying tens of thousands of AI accelerator chips (GPUs and specialized AI ASICs) in newly built server farms, resulting in power-dense facilities requiring robust power supplies and cooling. A single large AI data center can demand on the order of 100–150 MW of power – equivalent to the consumption of hundreds of thousands of homes. Hyperscalers' willingness to pay for reliable power is high, since electricity is only ~20% of their operating costs and AI services generate substantial revenue. Thus, we see hyperscale operators pushing the envelope on both scale and density of data centers, which in turn drives sector-leading electricity consumption growth [9][13].

In contrast, enterprise and small-scale data centers (those operated by individual businesses, colocation providers, or on-premises server rooms) are a smaller and more slowly growing segment. In the 2010s, many enterprises migrated workloads to the cloud or improved their facility efficiency, which helped limit energy growth in this sector. Enterprise data center energy use has largely plateaued or even declined in some cases as older inefficient server rooms are retired. However, certain industries with sensitive data or low-latency needs still maintain their own data centers. These enterprise facilities are now also adding AI capabilities – for example, financial institutions deploying AI for fraud detection may upgrade their on-premises compute clusters. While any single enterprise data center is relatively small (usually a few MW or less), collectively they number in the thousands of sites. The net effect is modest growth in enterprise data center energy use, likely overshadowed by the outsized gains on the hyperscale side. One consequence is that the data center industry's energy profile is shifting: a decade ago, a significant share of load was in distributed enterprise server closets with poor efficiency, whereas today a greater share is concentrated in large, professionally managed cloud facilities with much lower PUE (Power Usage Effectiveness). This consolidation yields efficiency benefits, but also means new loads come in large, concentrated blocks rather than diffuse increments [2].

A third category is industrial and edge AI infrastructure. Industrial companies are increasingly adopting AI for automation, predictive maintenance, and analytics in manufacturing. This includes on-site high-performance computing (HPC) clusters at factories, research labs, and oil & gas facilities. For example, oil exploration firms use GPU-rich clusters for seismic data analysis; automotive firms run AI models for autonomous vehicle R&D. These industrial data centers (often 1–5 MW scale) add to electricity use in the industrial sector, although they remain a smaller fraction compared to process-related electric loads. Nonetheless, as AI is integrated into manufacturing ("Industry 4.0"), even traditional factories may house AI servers or edge devices that incrementally raise their power draw.

Edge computing and distributed AI is an emerging area that could become significant by 2030. To reduce latency and bandwidth usage, companies are deploying smaller data centers and AI inference servers closer to end-users – for instance, in telecom central offices, at cell tower sites, or in regional micro-data centers. These "edge" installations typically range from a few kilowatts up to a few hundred kilowatts each. Individually, they are minor loads on local distribution feeders, but in aggregate they represent an expanding distributed IT load on the grid. As 5G networks, autonomous vehicles, and IoT devices proliferate, analysts foresee thousands of edge data centers coming online (for content caching, real-time analytics, etc.). The power consumption of edge computing is expected to grow in tandem with core cloud computing. One industry projection (pre-generative-AI boom) was that U.S. data center capacity would double from ~17 GW in 2022 to 35 GW by 2030 even before accounting for new AI workloads – a trend that included both large and small facilities. Edge deployments will contribute to this growth by bringing AI inference closer to users, though some of that load is simply shifted from central clouds. The key point for utilities is that data-handling electricity demand will not be confined to giant campuses; it will also appear as many distributed sites drawing power from diverse parts of the grid. This complicates planning for distribution networks and raises the importance of local infrastructure upgrades (more on this in a later section) [14].

In summary, the hyperscale cloud sector is the primary engine of AI-driven electricity demand growth, with industrial and edge computing playing supporting roles. Hyperscalers are creating massive new electric loads that dominate the national statistics. Enterprise and on-premise facilities are generally optimizing or offloading to cloud, keeping their growth moderate. Meanwhile, industrial and edge AI are emerging as new load categories – smaller in scale but numerically abundant – that will add to the overall picture of rising demand. Understanding these sectoral differences is important for grid planners: concentrated vs. distributed load growth, constant vs. variable usage patterns, and differing capabilities to manage or shed load all depend on the type of AI infrastructure in question.

Regional Grid Impacts and Data Center Siting Trends

The AI and data center boom is highly geographic in its impacts. Certain U.S. regions have become hotspots for data center development, leading to disproportionate load growth and new stresses on local grids. Other regions have seen little to no data center expansion. These trends are reshaping the regional load map and influencing where new power infrastructure is most urgently needed.

Virginia (Northern Virginia): The corridor outside Washington D.C. (Loudoun, Fairfax, Prince William counties) is famously known as “Data Center Alley,” hosting the world’s largest concentration of data centers. Northern Virginia’s advantages – access to fiber networks, proximity to federal clients, favorable tax incentives, and reasonably priced power – made it a magnet for hyperscalers. By 2023, data centers consumed an astonishing 25–26% of Virginia’s total electricity usageenvironmentamerica.org. Dominion Energy, the state’s dominant utility, reported that 24% of its electricity sales in 2023 were to data centers. This load is equivalent to several gigawatts on the grid, and it continues to rise rapidly. The Northern Virginia Electric Cooperative (NOVEC) projects its peak load will grow 12% per year for the next 15 years, driven almost exclusively by data center demand. Such growth, if realized, would require a massive expansion of substations, feeders, and possibly new transmission lines to deliver power into Loudoun County. It is no exaggeration that data centers have become the single largest strain on Virginia’s grid planning: local regulators have convened studies on how these energy “hogs” may affect reliability, rates, and state clean energy goalsenvironmentamerica.org. Dominion has had to upgrade infrastructure and even consider locational constraints on new data center connections as the cluster’s load outpaces some equipment ratings. Virginia’s experience highlights how clustering of AI infrastructure can dramatically alter regional load shape – Dominion’s summer peak is now set in part by data center demand, and minimum loads have risen as well due to their 24/7 operation [12][15].

Texas: Texas has also emerged as a major data center growth area, particularly in the Dallas-Fort Worth metro (a hub for colocation facilities) and more recently in Central Texas for hyperscalers. In addition, Texas hosts a number of large cryptocurrency mining sites which share similarities with data centers in their high power use. From 2019 to 2023, Texas saw one of the largest increases in commercial-sector electricity use (adding ~13 billion kWh, +10%) largely due to large-scale computing facilitiesdatacenterdynamics.com. The Electric Reliability Council of Texas (ERCOT) has reported dozens of “near-miss” grid events where big loads like data centers or crypto mines tripped offline unexpectedly, causing sudden swings in power balancereuters.com. For example, in west Texas a substation failure in 2022 led to ~400 data centers and crypto miners disconnecting, abruptly removing 1,700 MW of demand (about 5% of ERCOT’s load)reuters.com. While an oversupply of power might sound benign, such rapid load drops can force generators offline and risk destabilizing frequency. Grid operators now realize that large industrial loads must be treated akin to power plants in planning contingenciesreuters.com. ERCOT has adjusted its interconnection requirements and is closely monitoring the addition of new high-MW data facilities (Texas has active proposals for both data centers and more crypto mining). With its competitive electricity prices and ample land, Texas will likely continue attracting data center investments – meaning ERCOT’s planners need to factor in both the steady consumption and the possibility of abrupt outages or ramping by these loads (for instance, some crypto mines participate in demand response, dropping load when prices spike, which can cause significant variability) [5][16].

Other Regions: A handful of other states have notable data center clusters: Iowa (major cloud campuses powered partly by wind energy), Oregon (especially around Prineville and Hillsboro where Facebook and others have centers tapping into hydro/wind), Arizona (Phoenix area is growing as a data center hub), Illinois (Chicago region for colocation and cloud), New Jersey (financial data centers), Georgia (recent large cloud builds near Atlanta), and Nebraska/Wyoming (which have smaller populations, so even a few big data centers make up a large share of state consumption). According to EPRI data, states like Nebraska, Iowa, Wyoming, Nevada, and Arizona all had 7–12% of their electricity use going to data centers in 2023visualcapitalist.com. California, despite being the tech heartland, had only about 3.7% of its electricity in 2023 for data centers – partly because California's overall demand is huge and efficiency is high, and some tech firms site their largest server farms out-of-state where power is cheaper. Interestingly, North Dakota saw the fastest relative load growth (up 37% from 2019–2023) due to a couple of large computing facilities (including an AI/data center campus and crypto operations) tapping into the state's inexpensive powerdatacenterdynamics.com. This shows that even relatively remote areas are being targeted for data center siting if they offer a favorable mix of low-cost energy and available grid capacity [12][16].

The siting criteria for AI mega-datacenters generally include: access to cheap and reliable electricity, robust fiber optic connectivity, available land, tax incentives, and sometimes cool climates (for cooling efficiency). Many hyperscalers also prioritize access to renewable energy (for sustainability goals) – for instance, Google built large centers in Iowa and Oklahoma to utilize wind power, and Microsoft in Wyoming and Texas with wind/solar PPAs. The result is that data center growth is spatially uneven. States like Virginia and Oregon proactively courted data centers with tax breaks and now host huge concentrations, whereas other states have only a few. However, as power and land become constrained in the traditional hubs (e.g. Northern Virginia has faced community pushback and grid limitations), companies are scouting new regions. There is a trend of considering "energy community" sites, such as repurposing retired coal plant locations for data centers. These sites often have strong transmission infrastructure in place and available land and cooling water, making them attractive for re-development as data center campuses. DOE has highlighted this strategy – leveraging existing grid assets at decommissioned fossil plants to supply new high-tech loads [4].

Transmission and Grid Infrastructure Impacts: The regional clustering of AI data centers means localized grid stress. Utilities in hub areas are having to invest in new substations, high-voltage transmission loops, and distribution upgrades to serve these energy-intensive campuses. For example, in Virginia, Dominion Energy has built multiple new high-voltage transmission lines and substations in Loudoun County over the past decade specifically for data center load, and more are planned. In some cases, the lead time for transmission is becoming a bottleneck – it can take 5–7 years to site and build new lines, while data centers can be constructed in 2 years or less. This misalignment of timelines has led to situations where data centers sit waiting for sufficient grid capacity, or conversely, where they come online and strain existing infrastructure. Grid operators like PJM are now explicitly modeling data center growth in their load forecasts: PJM's latest 2024 long-term forecast nearly doubled the anticipated load growth in the Mid-Atlantic region versus previous years, chiefly due to Northern Virginia data center additions. PJM expects its overall summer peak to rise ~2% per year, a sharp increase after years of flat demand, attributing much of it to the digital sector [12][17][18].

Another emerging issue is grid reliability under high data center penetration. Traditionally, planners worried about a large power plant tripping offline; now they must equally consider a large load block tripping off (or on). The North American Electric Reliability Corporation (NERC) warned in late 2024 that as data centers concentrate, an unplanned simultaneous outage of multiple centers could induce instability. The "near-miss" incident in July 2022 in Northern Virginia – where a cluster of data centers (~1,500 MW total) disconnected due to a grid fault – forced PJM to rapidly cut generation to avoid over-frequency, narrowly averting a cascading outage. Following that, NERC formed a task force to study the risk of "en masse" data center disconnections. As one NERC official put it, "the grid is not designed to withstand the loss of 1,500 MW of data centers" without additional resources or damping. This is driving considerations for special protection schemes, faster load shedding systems, or on-site backup support to ensure a graceful handling of such events. In ERCOT's case, some large flexible loads (e.g. crypto mines) have agreed to provide under-frequency load shedding – essentially turning off quickly if frequency drops – which helps balance their impact [5].

In summary, region matters. States like Virginia, Texas, Oregon, Arizona, and others with major AI/data center growth are experiencing a fundamental shift in load shape and infrastructure needs, while other areas remain relatively unaffected so far. Utilities in the hot spots must accelerate capacity expansion and carefully manage reliability risks associated with huge new loads. Going forward, we may see more geographic diversification of data centers (to mitigate hub constraints), including movement to rural areas or areas with excess renewable energy potential. Nonetheless, the existing hubs are slated to grow further in the near term, so the regional grids will need creative solutions – from stronger transmission ties to possibly co-located generation – to support the AI revolution.

Load Profile Evolution and Utility Capacity Planning

The integration of large AI and computing loads is not only a question of how much demand, but also when that demand occurs. Load profile characteristics of data centers differ from many traditional loads, leading to new considerations for daily and seasonal grid operations.

Data centers typically draw power at a fairly steady, 24/7 rate. Unlike residential or commercial building loads that peak during certain hours (e.g. evening or midday), a hyperscale data center tends to consume near its peak power round-the-clock, with only minor fluctuations. This high load factor means that as data centers become a bigger portion of the grid's load, the overall system load profile can flatten. For instance, areas with many data centers may see higher nighttime and weekend loads than they historically had, because servers keep running regardless of human activity cycles. In Northern Virginia, Dominion Energy has noted that data centers are boosting the region's off-peak demand, effectively raising the "valley" load levels at night. In fact, Virginia's statewide electricity demand has risen even as some other sectors stagnated, directly attributable to the continuous data center [6].

From one perspective, flatter load profiles can be beneficial: they allow generation assets to run at steadier outputs and improve the utilization of transmission infrastructure. Base-load resources (like nuclear or combined-cycle gas) can operate more consistently if there is less variability between day and night demand. For example, a nuclear plant that might have had to curtail output at night in a low-demand scenario can now serve data center load at all hours. Some grid operators welcome this "baseloading" effect of data centers, as it provides a stable revenue stream and can reduce the ramping requirements on other plants. On the other hand, the lack of demand flexibility is a concern. During peak grid stress events – say a hot summer afternoon with high air conditioning load – data centers generally do not reduce their consumption (unless they enroll in demand response programs or switch to backup generators temporarily). Their contribution is essentially an added firm load on top of existing peaks. Thus, utilities must plan enough capacity not only for traditional peak drivers but also to cover the ever-present data center demand even in extreme conditions.

There is also the question of coincidence with renewable generation. California's CAISO, for instance, has a surplus of solar power at midday but faces ramping needs in the evening. A data center with a flat 24-hour profile will draw heavily during the evening ramp (when solar drops off), which could exacerbate ramping challenges unless paired with storage or demand management. Some operators are exploring making data center loads more flexible to align with renewable availability. Google has implemented "carbon-aware computing" where certain non-urgent workloads (like batch processing or AI training tasks) are shifted to times of day when renewable energy is abundant or prices are low. This kind of temporal load shifting could in theory transform some data center load into a flexible resource that helps balance the grid (e.g., crunch AI jobs when wind output is high at night, or pause them during grid peaks). To date, most hyperscalers only do this on a limited basis, as many tasks are latency-sensitive. But as AI training workloads grow, there is potential to schedule them intelligently to provide a form of demand response – effectively using the data center as a “virtual battery” by varying its consumption within certain limits.

Utilities, in their capacity planning, increasingly must account for these new load profiles. Many utilities are now including explicit data center growth segments in their load forecasts and integrated resource plans. For example, PJM Interconnection’s latest planning outlook indicates it will need to ensure tens of GW of new capacity by the 2030s largely due to data center-driven load growth in its Mid-Atlantic zone. Similarly, the peak demand season could be affected: data centers themselves don’t have a strong seasonal swing (cooling needs can make their summer load slightly higher, but many use efficient cooling or water systems to manage heat year-round). So the grid’s winter vs summer peak differential might narrow in regions with big data center influence. In ERCOT, winter peak was historically lower than summer, but with the addition of large constant loads (including crypto mines), some winter days have approached summer-like demand levels when heating and those industrial loads coincide[17][19].

In terms of capacity credit and reliability, system planners treat data center demand as a relatively inflexible load that must be met with firm capacity. Some data centers have backup generation (diesel gensets or gas turbines) that could be called upon in emergencies. Indeed, there are discussions about leveraging those backup gensets as a last-resort grid support resource. For instance, if the grid is critically short on capacity, a utility could ask large data centers to fire up their onsite generators to temporarily offload the grid. Programs like this effectively turn data centers into demand-side resources for peak shaving. However, running diesel generators has emissions and cost implications, so it would only be a strategy for rare events. A more sustainable approach is encouraging on-site energy storage at data centers. Many centers already have battery UPS systems for reliability; if those were upsized, they could provide a buffer – drawing power during low-price hours to charge, and discharging to supply the facility (or even the grid) during peak price hours. This concept of data centers as “grid interactive” resources is gaining traction. The U.S. DOE is promoting data center flexibility through initiatives for onsite storage and generation, aiming to make data centers a grid asset rather than a pure load [4].

From a planning and rate design standpoint, some utilities are creating special tariffs for high-load customers like data centers. These can include time-of-use pricing, curtailment agreements, or construction allowances for the needed infrastructure. The goal is to send price signals that at least encourage data centers to locate or operate in ways beneficial to the grid (for example, lower rates in regions with surplus capacity, higher in constrained nodes). Regulators in states with big data center growth are also examining how to allocate the costs – ensuring that general ratepayers are not unduly burdened by the new construction needed for a few large customers. In

Virginia, debates have arisen about whether data center operators should contribute more to grid upgrade costs since they are driving them. Innovative rate structures (e.g., demand charges that strongly reflect peak contribution) are one tool being considered to maintain affordability while supporting expansion [4].

In summary, AI and data center loads are pushing utilities to think differently about load profiles and capacity. The old paradigm of daily peaks and valleys is giving way to a more continuous high load situation in some regions. This requires building sufficient capacity (generation, storage) to meet a higher baseline demand at all times, and ideally, finding ways to harness these loads for grid support (through smart scheduling or emergency backup use). Utilities are now planning not just for "N-1" generation contingencies, but also for "N-1" load contingencies (a large load drop). Resilience strategies, such as sectionalizing data center campuses or fast load shed schemes, are becoming part of the reliability toolkit. As AI electricity demand grows, successful grid integration will hinge on closer coordination between data center operators and utilities – sharing of operational data, agreements on interruptibility or flexibility, and co-investment in solutions like on-site storage or renewables.

Sensitivity Analysis and Scenario Comparisons

Forecasting AI-related electricity demand is challenging, and various scenarios can diverge widely depending on key assumptions. Energy planners and analysts use sensitivity analyses to explore how different factors might alter the trajectory of power consumption from AI and data centers. Here we consider a few major uncertainties and compare scenarios:

Rate of AI Adoption: Perhaps the largest swing factor is how quickly and pervasively AI technologies are adopted across the economy. A High Adoption scenario assumes continued breakthrough advances (e.g. ever-larger neural networks powering new services, widespread enterprise AI deployment) leading to exponential growth in compute needs. In this scenario, data center construction remains at record pace through the 2020s. Projections like the DOE/LBNL high case of 12% of U.S. power for data centers by 2028, or Goldman Sachs' estimate of 8–9% by 2030, would be realized or even exceeded. By 2040, AI workloads could be ubiquitous in manufacturing, transportation, healthcare, and beyond, potentially pushing data centers towards a mid-teens percentage of national load. Conversely, a Low Adoption (or slower) scenario might occur if AI deployment encounters headwinds – perhaps regulatory restrictions on data usage, public backlash, or simply diminishing returns on scaling model sizes. If the "AI boom" leveled off by early 2030s, data center demand might still grow, but at a decelerating rate. The low end of forecasts – e.g. staying near ~5% of U.S. electricity in 2030 – would imply more modest growth thereafter. In such a case, overall electricity demand by 2040 might fall in line with historical GDP-linked growth, rather than the high-demand scenarios [4][8][12].

Efficiency Improvements: Another crucial variable is the energy efficiency of computing. Historically, improvements in server efficiency and cooling (PUE reductions) allowed data center workloads to increase without proportional energy growth. Will this continue? A Optimistic Efficiency scenario posits that new technologies (advanced chip designs, optical interconnects, better algorithms) significantly reduce the energy per AI computation. For example, specialized AI accelerator chips (TPUs, neuromorphic processors) might perform the same tasks as today's GPUs at a fraction of the wattage. Cooling might shift to liquid or even cryogenic methods that cut ancillary power use. If each generation of AI hardware is dramatically more efficient, some of the anticipated load growth could be offset. The IEA notes that continued efficiency gains at hardware and software levels are expected to "mitigate" some data center energy growth, even as total demand still risesiea.org. On the other hand, a Stagnant Efficiency scenario (or Jevons paradox scenario) could occur if improvements slow or if efficiency gains just enable more intensive workloads rather than reducing energy use. The McKinsey analysis points out that in computing, unlike other areas, breakthroughs that lower cost or increase efficiency often increase demand – because they enable running more complex models and unlocking new use cases. In other words, if we make AI computing 2× more efficient, we might simply run a model that is 2× larger or serve 2× more inferences, ending up using the same or more total power. Therefore, scenario analysis must consider whether efficiency translates into actual energy savings or gets reabsorbed as higher computational output [2][9][13].

Alternate Workload Placement: Another sensitivity is where AI workloads are processed – centrally in big data centers vs. on edge devices or user devices. A Centralized scenario assumes most AI inference and training stays in cloud data centers (current trend), maximizing the grid-connected load. A Distributed scenario might imagine that by 2040, a significant portion of AI processing is done on billions of edge devices (phones, cars, IoT). If edge devices handle more inference locally, data center servers can offload some work, potentially easing the growth in data center energy. However, edge devices themselves draw power (usually from distribution grids or batteries). Some analyses of ICT energy use look at total "ICT footprint" (including user devices, networks, and data centers). It's possible that improved efficiency in data transmission and local processing could reduce the need for every minor AI query to hit a power-hungry data center. On the flip side, the growth of cloud-based AI (like large language model APIs, cloud AI services) suggests centralization is still very strong. Scenarios can explore extremes: one where every vehicle and factory has on-prem AI servers (more distributed consumption, possibly lower transmission needs) vs. one where nearly all compute is done in a few cloud regions (higher concentration, requiring big transmission). The overall energy difference may not be huge in magnitude, but it affects which systems (central power plants vs. local distribution) carry the loads.

Demand Management and Flexibility: We should also consider a scenario where data centers become active grid participants providing demand response. For instance, a Flexible scenario might assume by 2035 many hyperscale data centers contract with utilities to shed or shift a portion of load during peak emergencies (using their backups or delaying non-urgent tasks). This could shave the top off peaks and effectively reduce the net load growth that must be met by new generation. A Rigid scenario would assume data centers remain completely uninterruptible and run at full tilt regardless of grid conditions, requiring one-for-one new generation for every kW of load. Most current forecasts (like EIA, EPRI) implicitly assume rigidity – they count the full demand in needing new supply. If, however, even 10-20% of data center load could be treated as flexible, it introduces a significant resource for grid balancing. Sensitivity analysis could assign different values to "interruptible load" from data centers and see how it impacts capacity expansion needs. Preliminary experiments (such as centers agreeing to throttle if frequency dips) are promising, but broad adoption is uncertain without clear market signals or regulations.

To illustrate scenario ranges: the 2024 LBNL/DOE data center report provided a low and high case for 2028 – 6.7% vs 12% of U.S. power – basically reflecting different assumptions about AI uptake and efficiency. By 2040, we can envision a Low scenario where data centers maybe stabilize at ~10% of load, and a High scenario where they approach 20% or more (if AI truly permeates everything and efficiency gains plateau). Total U.S. electricity demand in those scenarios could range from maybe 5,000 TWh up to 7,000+ TWh, a huge divergence impacting generation investment. Such wide uncertainty underscores why utilities are wary – as the Kansas City Fed noted, recent forecast revisions show the difficulty, with 2025 demand growth forecasts revised upward eightfold once AI trends were accounted for [4][6].

Another factor is economic and policy context. If climate policies or electricity price increases lead to aggressive energy efficiency measures in other sectors, that could free up headroom for data center loads without growing total demand as much. Conversely, if electricity remains cheap and plentiful, AI developers might be less incentivized to optimize for efficiency, leading to higher consumption. There's also a wild card in quantum computing on the horizon (post-2030) – if quantum computers become viable, they might take over certain AI workloads far more efficiently, or they might themselves require new cryogenic facilities with high energy use. Scenarios toward 2040 sometimes include possibilities like this.

In summary, scenario analysis for AI-driven electricity demand typically compares a base case (steady improvement, moderate adoption) with a high case (breakneck AI growth, slower efficiency gains) and sometimes a low case (slower growth or societal pushback limiting AI). The high cases are the ones raising flags: they point to unprecedented load additions that challenge our generation and T&D build-out schedules. The low or managed cases suggest that with technological and policy interventions, we could still reap AI's benefits without straining the grid to the breaking point. For prudent planning, many utilities are now conducting sensitivity studies specifically on data center growth – essentially stress-testing their systems if, say, an extra 500 MW or 1000 MW of unexpected load shows up in five years. As one utility executive put it, it's the "magnitude and uncertainty" of these new AI loads that keeps planners awake at night. Strategies to handle this uncertainty include shorter-term resource procurements, modular generation that can scale up if needed, and close dialogue with large tech customers about their expansion plans [7].

Decarbonization Challenges and Mitigation Strategies

The surge in AI-related electricity demand comes at a time when the U.S. power sector is undergoing a transition to cleaner energy. This juxtaposition creates both challenges and opportunities for achieving decarbonization goals such as net-zero emissions by 2050. On one hand, additional load from data centers could make it harder to decarbonize – more demand could mean more fossil generation in the near term if renewables can't be built fast enough. On the other hand, if managed properly, the AI boom might catalyze investments in clean energy technologies and even provide flexible demand to help integrate renewables. Here we analyze the interplay.

Emissions Impact of Increased Load: If the grid's generation mix remained the same, a large increase in electricity use would translate to higher carbon emissions. For example, training a single large AI model can consume megawatt-hours of electricity; if that power comes from coal or natural gas, the carbon footprint of AI becomes significant. A key worry is that data center growth could slow the retirement of fossil-fueled plants or even trigger new ones. Indeed, some utilities facing rapidly rising AI loads are considering building natural gas peaking plants to ensure reliability through the 2020s. A Bain analysis estimated that globally, rising data center power consumption could require over $2 trillion in new generation resources, and if not steered to clean sources, a lot of that could be fossil. This scenario would pose a challenge to climate targets – more CO_2 unless renewable deployment greatly accelerates. NERC's reliability outlook noted that nearly all U.S. regions face elevated risk of energy shortfalls in the next decade partly due to new data centers; one stopgap solution has been to keep some coal or gas units on reserve that would otherwise retire. In Virginia, for instance, the state's clean energy progress is under scrutiny because data center energy demand is "exploding" at the same time coal plants are retiring – raising the question of whether gas generation will fill the gap [5][7][15].

However, there is a strong counter-current: the major AI/data center operators have made ambitious renewable energy and carbon commitments, which is actually driving decarbonization investment. Companies like Google, Amazon, Microsoft, Meta have all pledged to power their data centers with 100% renewable energy (and some aim for 24/7 carbon-free energy). In practice, these companies have become some of the world's largest buyers of renewable power purchase agreements (PPAs). For example, Amazon announced it had achieved 100% renewable electricity for its operations as of 2022, primarily by investing in wind and solar farms around the world. If every incremental MWh that an AI data center uses is matched by a new MWh of wind or solar on the grid (via PPAs), then net emissions need not rise proportionally with load. Indeed, one opportunity is that the AI sector's demand can underpin financing for vast new renewable projects. The sustained high PPA prices and appetite from hyperscalers have been crucial in expanding U.S. renewables in recent years [9][20].

There are still challenges in the details: matching "annual" renewable energy is one thing, but achieving round-the-clock clean supply is harder. Data centers often consume power even when the wind isn't blowing or sun isn't shining, meaning the grid mix at those hours might include fossil fuel. Companies like Google are working on 24/7 clean energy strategies (e.g. integrating storage, load shifting, and firm zero-carbon resources) to truly decarbonize their consumption hour-by-hour. Until that is achieved broadly, increased AI load could cause short-term upticks in fossil generation even if over a year it's netted out by renewables.

One notable trend is interest in carbon-free firm power for data centers, such as nuclear energy. In 2023 and 2024, several tech firms made headlines for exploring advanced nuclear projects to power their data centers. For instance, Microsoft agreed to purchase output from a nuclear plant (the operator of Pennsylvania's Three Mile Island) to supply its data center needs. Amazon announced a partnership to develop small modular reactors (SMRs) in Washington state, aiming for ~320 MW of nuclear capacity online by the early 2030s to help power its AWS cloud region therebizjournals.com. The rationale is that SMRs or other next-gen nuclear can provide steady, carbon-free electricity that aligns with data centers' 24/7 load. While these projects are in early stages, they indicate that AI-heavy companies are willing to invest in innovative generation to secure clean power. McKinsey also pointed out the potential alignment of hyperscalers and nuclear developers: hyperscalers have the capital and long-term outlook that could support nuclear despite high upfront costs. Several challenges (cost, timing, public acceptance) remain for nuclear solutions, but they are clearly on the table [9][20].

Beyond generation, data centers are exploring other decarbonization strategies: such as using green hydrogen or biogas for backup generators instead of diesel, optimizing cooling to reduce electricity use (some are locating in cooler climates or using renewable-powered cooling systems), and even reusing waste heat. In a few cases, the waste heat from data centers is being piped to heat nearby buildings or greenhouses – effectively offsetting other energy use. This isn't widespread in the U.S. yet (more common in Nordic Europe), but it's being piloted as a way to improve overall energy efficiency and carbon footprint.

From the grid perspective, one challenge with integrating so much new load while decarbonizing is ensuring the transmission build-out for renewables keeps pace. Many of the regions with big data center growth (e.g. Virginia, Pacific Northwest, etc.) will need new transmission to bring in wind/solar from remote areas if they are to meet that load with clean energy. If transmission lags, local fossil generators might run more. To mitigate this, there's discussion of siting data centers nearer to renewable generation (for example, putting future data centers in the Midwest near wind farms, or in the Southwest near solar fields). Some companies are indeed locating facilities in regions like rural Nebraska or Wyoming with excellent renewable resources, effectively coupling the load with new clean power plants. DOE has even suggested using energy community sites (like old coal plant sites) for data centers, which often are near robust grid connections and could host new clean energy as well. This strategy could ensure that as old high-carbon generation is retired, new high-tech load at that site is fed by incoming renewables on those same wires, maintaining utilization [4].

Another opportunity is that AI and digitalization can themselves be tools for grid decarbonization. AI is increasingly used in grid management – from forecasting renewable output to optimizing dispatch and detecting faults. The energy sector can leverage AI to improve efficiency and integrate higher shares of renewables (for example, AI algorithms for demand response, smart EV charging, and optimizing battery storage). So in a broader sense, while AI's energy appetite is a challenge, AI is also a key enabler of a smarter, more flexible grid that can handle decarbonization better. The IEA has highlighted the "enormous potential" for digitalization (including AI) to either help or hinder clean energy transitions. It will hinge on how we direct these technologies [21].

In policy terms, ensuring that AI's growth aligns with climate goals may involve new standards or incentives. One concept is establishing energy efficiency standards for data centers or AI hardware – similar to appliance efficiency standards – to push the industry toward best practices. Another is mandating or incentivizing green power procurement for large new loads. Some jurisdictions are considering requiring that new data centers have a plan for renewable sourcing or are even net-zero (through offsets or clean onsite generation). States like Washington have clean energy laws that will force any incremental load (like Amazon's SMR project in WA) to be met with carbon-free sources by certain dates. Additionally, the industry itself through groups like the Climate Neutral Data Centre Pact (in Europe) or similar initiatives in the U.S. is setting voluntary targets for efficiency (e.g. a PUE below a certain threshold, water usage limits, etc.) to minimize the environmental impact [20].

To summarize, the decarbonization challenge posed by AI's electricity demand is twofold: scale up clean energy supply fast enough to meet the new load, and do so without compromising grid reliability or affordability. The opportunity is that the tech sector's demand can be a driver for innovation in clean energy – whether it's massive renewable purchases, investment in advanced nuclear, or pioneering demand flexibility solutions. Many utilities are optimistic that they "can meet this growth with clean energy" if proper actions are taken. The U.S. Energy Secretary noted that rising AI/data center demand underscores why DOE is deploying technologies like next-generation geothermal, advanced nuclear, and long-duration storage – these will be crucial to meet the demand surge while cutting emissions [4].

In practical terms, strategies to align AI demand with net-zero goals include:

- Rapid renewable energy expansion: Streamlining permitting and interconnection for solar, wind, and storage to ensure supply keeps up. Data center companies will likely continue aggressive PPA procurement for new projects.
- Firm low-carbon capacity: Pursing solutions such as nuclear (SMRs), geothermal, or gas with carbon capture to provide reliable clean power for when renewables fall short, especially in regions with big 24/7 loads.
- Grid upgrades and energy storage: Strengthening transmission to connect renewable-rich areas with load centers, and deploying storage (battery and beyond) to buffer the intermittency. Data centers themselves might host on-site batteries that help both the facility and the grid.
- Energy efficiency and R&D: Supporting R&D in more efficient AI algorithms and hardware (so that each increment of compute uses less kWh). Also continuing improvements in cooling (e.g. AI-driven cooling management) to lower PUE across the industry.
- Demand-side management: Encouraging data center participation in demand response, and developing tariffs or programs that reward flexibility and off-peak consumption. This might include pricing that incentivizes scheduling non-critical AI tasks to mid-day (solar peak) or overnight (wind peak).
- Policy and collaboration: Developing policies that ensure new large data centers are built as “good grid citizens” – e.g. located in areas that aid grid stability, using waste heat productively, and collaborating with utilities on integrated resource planning.

The coming decades will be a delicate balancing act: embracing the economic and societal benefits of AI and digitalization, while aggressively mitigating the environmental footprint of the supporting infrastructure. The stakes are high – unmanaged, AI’s electricity hunger could put a dent in decarbonization progress, but managed wisely, it can be the impetus for modernizing and greening the electric grid at an accelerated pace.

Conclusion

AI is poised to be a transformative force not just in the digital realm, but in the physical domain of energy systems. In the United States, the rise of AI and its computational demands are already lifting electricity consumption after years of stagnation. By 2040, AI-related infrastructure – from colossal cloud data centers to distributed edge nodes – could represent one of the largest slices of national power usage. This report has examined how such a trend is likely to unfold, drawing on the best available forecasts and current indicators. The historical context shows that computing can drastically alter energy trajectories, as seen by the plateau and then spike in data center usage. Projections uniformly anticipate robust growth in AI-driven demand, though with a range of uncertainty tied to technology and adoption scenarios. We explored how growth breaks down by sector, with hyperscale operators leading the charge, and how it varies regionally, with certain states under heavy strain while others see new opportunities. We discussed the evolving load profile introduced by constant AI loads and the planning adaptations utilities must make in response. Scenario analysis underscores that the future is not set in stone – policy choices, efficiency innovations, and industry practices will influence whether we land on the low or high side of the range.

Crucially, we assessed the intersection with decarbonization: increased AI electricity demand is coming at the same time as the push for net-zero emissions. This convergence can be a collision course or a virtuous cycle, depending on our approach. The findings suggest that proactive strategies – massive clean energy deployment, flexible load management, and embracing new technologies – can enable the grid to support AI's growth sustainably. AI itself can be harnessed to enhance grid efficiency and renewable integration, creating a feedback loop where digital technology helps power the clean energy transition even as it relies on it.

For energy professionals and decision-makers, the key takeaways are clear. Electric load forecasting and resource planning must incorporate AI/data center growth as a core element, not a peripheral factor. Traditional demand models may need an overhaul as we enter an era where "bits" drive electrons more than population or economic GDP alone. Utilities in regions of high tech expansion should engage early and often with data center developers to align capacity expansion plans. Investment in grid infrastructure – both "steel in the ground" and digital intelligence – will be critical to maintain reliability amidst these new loads. At the same time, meeting sustainability targets will require leveraging the considerable influence of AI companies in procuring clean energy and perhaps setting standards for energy-conscious AI design.

In summary, the relationship between AI demand and electricity consumption is one of accelerating interdependence: AI will shape future electricity demand trends, and the state of the electricity system (its cost, carbon intensity, and availability) will shape the trajectory of AI deployment. Navigating this dual revolution – in computing and in energy – will be a defining challenge and opportunity for the coming decades. With prudent planning, innovation, and collaboration between the tech industry and energy sector, the U.S. can support the computing power needs of the AI era while advancing toward a resilient, decarbonized electric future.

Bibliography

1. International Energy Agency / LBNL, "U.S. Data Centres Energy Use Report," *IEA 4E*, accessed May 2025, https://www.iea-4e.org/wp-content/uploads/publications/2016/06/05j_-_LBNL_-_US_Data_Centres_Energy_US.pdf.
2. Data Center Knowledge, "Here's How Much Energy All U.S. Data Centers Consume," *Data Center Knowledge*, accessed May 2025, https://www.datacenterknowledge.com/energy-power-supply/here-s-how-much-energy-all-us-data-centers-consume.
3. U.S. Environmental Protection Agency, "News Release Archive," *EPA*, accessed May 2025, https://www.epa.gov/archive/epapages/newsroom_archive/newsreleases/4be8c9799fbceb028525732c0053e1d5.html.
4. U.S. Department of Energy, "DOE Releases New Report Evaluating Increase in Electricity Demand from Data Centers," *U.S. Department of Energy*, accessed May 2025, https://www.energy.gov/articles/doe-releases-new-report-evaluating-increase-electricity-demand-data-centers.
5. Reuters, "Big Tech's Data Center Boom Poses New Risk to U.S. Grid Operators," *Reuters*, accessed May 2025, https://www.reuters.com/technology/big-techs-data-center-boom-poses-new-risk-us-grid-operators-2025-03-19/.
6. Federal Reserve Bank of Kansas City, "Powering Up: The Surging Demand for Electricity," *Kansas City Fed*, accessed May 2025, https://www.kansascityfed.org/research/economic-bulletin/powering-up-the-surging-demand-for-electricity/.
7. Utility Dive, "Data Center Load Growth Could Hike U.S. Electricity Bills: Bain," *Utility Dive*, accessed May 2025, https://www.utilitydive.com/news/data-center-load-growth-us-electricity-bills-bain/730691/.
8. Environment America, "Electricity Demand from U.S. Data Centers Could More Than Double by 2035," *Environment America*, accessed May 2025, https://environmentamerica.org/center/articles/electricity-demand-from-u-s-data-centers-could-more-than-double-by-2035-thats-like-adding-another-california/.
9. McKinsey & Company, "How Data Centers and the Energy Sector Can Sate AI's Hunger for Power," *McKinsey & Company*, accessed May 2025, https://www.mckinsey.com/industries/private-capital/our-insights/how-data-centers-and-the-energy-sector-can-sate-ais-hunger-for-power.

10. International Energy Agency, “Electricity 2024: Executive Summary,” *IEA*, accessed May 2025, https://www.iea.org/reports/electricity-2024/executive-summary.

11. Bloomberg Law, “PG&E Sees Underutilized Grid as Ready for Jump in Power Demand,” *Bloomberg Law*, accessed May 2025, https://news.bloomberglaw.com/environment-and-energy/pg-e-sees-underutilized-grid-as-ready-for-jump-in-power-demand.

12. Visual Capitalist, “Mapped: Data Center Electricity Consumption by State,” *Visual Capitalist*, accessed May 2025, https://www.visualcapitalist.com/mapped-data-center-electricity-consumption-by-state/.

13. International Energy Agency, “What the Data Centre and AI Boom Could Mean for the Energy Sector,” *IEA*, accessed May 2025, https://www.iea.org/commentaries/what-the-data-centre-and-ai-boom-could-mean-for-the-energy-sector.

14. Goldman Sachs, “AI Poised to Drive 160% Increase in Power Demand,” *Goldman Sachs*, accessed May 2025, https://www.goldmansachs.com/insights/articles/AI-poised-to-drive-160-increase-in-power-demand.

15. Environment America, “Data Centers Could Derail Virginia’s Clean Energy Progress,” *Environment America*, accessed May 2025, https://environmentamerica.org/virginia/articles/data-centers-could-derail-virginias-clean-energy-progress/.

16. Data Center Dynamics, “Data Center Hotspots Driving Growth in U.S. Electricity Usage: EIA,” *Data Center Dynamics*, accessed May 2025, https://www.datacenterdynamics.com/en/news/data-center-hotspots-driving-growth-in-us-electricity-usage-eia/.

17. Ascend Analytics, “PJM’s Data Center Surge Calls on Battery Storage,” *Ascend Analytics*, accessed May 2025, https://www.ascendanalytics.com/blog/pjms-data-center-surge-calls-on-battery-storage.

18. Utility Dive, “PJM Load Forecast Shows Dominion, Virginia Face Explosive Data Center Growth,” *Utility Dive*, accessed May 2025, https://www.utilitydive.com/news/pjm-load-forecast-data-center-dominion-virginia/735056/.

19. Aurora Energy Research, “Northern Virginia Data Center Demand Could Incentivize up to 15 GW of Natural Gas by 2030,” *Aurora Energy Research*, accessed May 2025, https://auroraer.com/media/new-aurora-report-finds-northern-virginia-data-center-demand-could-incentivize-up-to-15-gw-of-additional-natural-gas-generators-by-2030/.

20. Puget Sound Business Journal, “Amazon Backs New Nuclear Power Plant to Power Data Centers,” *Puget Sound Business Journal*, accessed May 2025, https://www.bizjournals.com/seattle/news/2024/10/17/amazon-nuclear-power-plant.html.

21. International Energy Agency, “Data Centres and Data Transmission Networks,” *IEA*, accessed May 2025, https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks.