

Artificial Intelligence, Data Centers, and Energy Needs

To provide a better understanding of how we think about various credit sectors, we are producing a series of pieces that outline significant issues impacting sectors including domestic banks, pharmaceuticals, technology, and aerospace and defense. Part one of this series focused on domestic banks, and part two reviewed headwinds in the pharmaceutical space. In part three, Winnie Cheng, CFA, Vice President and Principal Credit Analyst, will discuss the demand for and cost of artificial intelligence (AI) technology.

What are the different types of AI?

There are several types of AI where the scope of capability is very different. Artificial narrow intelligence (ANI) is the most common and prevalent form of artificial intelligence today. ANI is highly specialized and restricted to performing a specific task based on predefined programming. Examples include voice assistants such as Siri, Alexa, and Google Assistant, image recognition that can identify objects, faces or emotions in images or video, such as facial recognition technology used in smartphones or security cameras. ANI is used in recommendation systems, where AI analyzes past behavior or browser history and recommends products, movies, or shows. Examples include shopping advertisements and recommendations for shows or music.

Generative AI refers to systems that create original content such as text, images, video, audio, and software based on user input. It relies on machine learning models, particularly deep learning models, to simulate the learning and decision-making processes of the human brain.

Agentic AI is an artificial intelligence system that can accomplish a specific goal with limited supervision, using machine learning models that mimic human decision-making to solve problems in real time to automate complex tasks, improve efficiency, and enable organizations to focus on strategic initiatives rather than

routine operations. Unlike traditional AI models that operate within predefined constraints and require human intervention, agentic AI exhibits autonomy, goal-driven behavior, and adaptability.

An agentic AI system can use generated content to complete complex tasks autonomously by calling external tools. Examples include automated customer service, predictive maintenance in manufacturing where the AI continuously monitors machinery through sensors to detect anomalies and triggers maintenance before breakdowns occur to minimize unplanned downtimes and reduce maintenance costs. Agentic AI can be used to automate sales outreach communications by building custom messages, track responses, and follow up sales leads based on engagement levels. This integration with customer relationship management systems helps keep sales pipelines organized and updated in real time. To date, most of the agentic AI uses are to improve operating efficiency and productivity to lower costs.

Differences between general compute and AI compute model

The distinction between general compute and AI compute lies in their application and capabilities. General compute is used for a wide range of tasks, including data processing, simulations, and applications that do not require the level of intelligence or adaptability that AI systems do. AI compute uses large language models (LLMs), a type of artificial intelligence algorithm, to train on vast amounts of data to process and understand human languages, enabling them to perform a wide range of tasks. AI compute is specifically designed for tasks that require intelligence, such as learning, decision-making, and problem-solving. The amount of computing power used in AI training is significantly higher than general compute and the amount of high-performance compute usage has been increasing exponentially. As AI systems become more



complex and capable, they will require more compute power to train, improve, and employ.

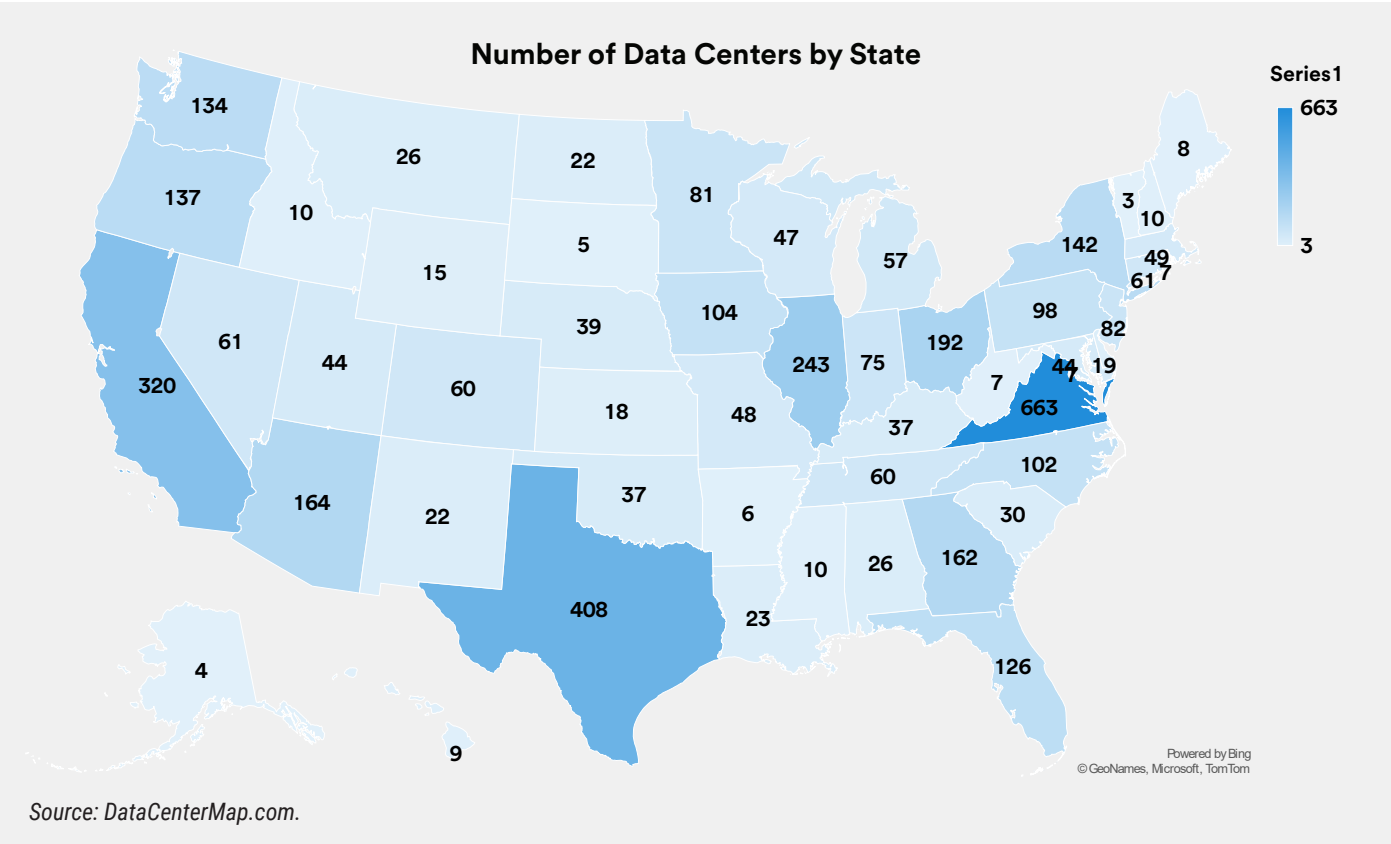
AI computing requires AI data centers equipped with cutting-edge technologies (GPUs (graphic processing units), TPUs (tensor processing units), and other AI accelerators), advanced storage architectures, resilient networking with broad bandwidth (maximum amount of data that can be transferred over a network connection in a given amount of time), low latency (delays in data transfer), and robust cooling systems to manage the heat generated from high energy demands of AI workloads.

Where are data centers located?

Older data centers were located closer to metropolitan centers where the businesses are, but the newer AI data centers are in rural or outer suburban areas with cheaper land, power availability, lower energy prices, connectivity to major digital infrastructure hubs for low latency connections for enterprises requiring fast data transfers, and regulatory incentives. As companies rush to build more datacenters, maps become outdated soon after publication.

The three hyperscalers Amazon Web Services, Microsoft Azure, and Google, plus Meta and Oracle are building multiple immense AI data centers. These companies have large backlogs due to inadequate capacity and continue to book multi-billion dollar deals to host cutting edge AI development work. Increasing demand for AI LLM training and inference has caused newcomers with minimal experience running massive data centers to build AI data centers with plans to rent their infrastructure to AI companies. Some technology companies are entering into joint ventures with investment companies to fund AI data center development to provide the technology companies operational and financial flexibility.

Hundreds of billion dollars have been pledged to build AI data centers around the world with significant focus in the U.S. Technology company debt issuance exceeded \$180 billion year-to-date as of October 30, 2025. Some of the debt will be used to fund rapidly scaling data center operations. The U.S. data center market is projected to add around 6.5 gigawatts of power capacity by the end of 2025, with a total of 27.5 gigawatts expected by 2027.





Energy Needs

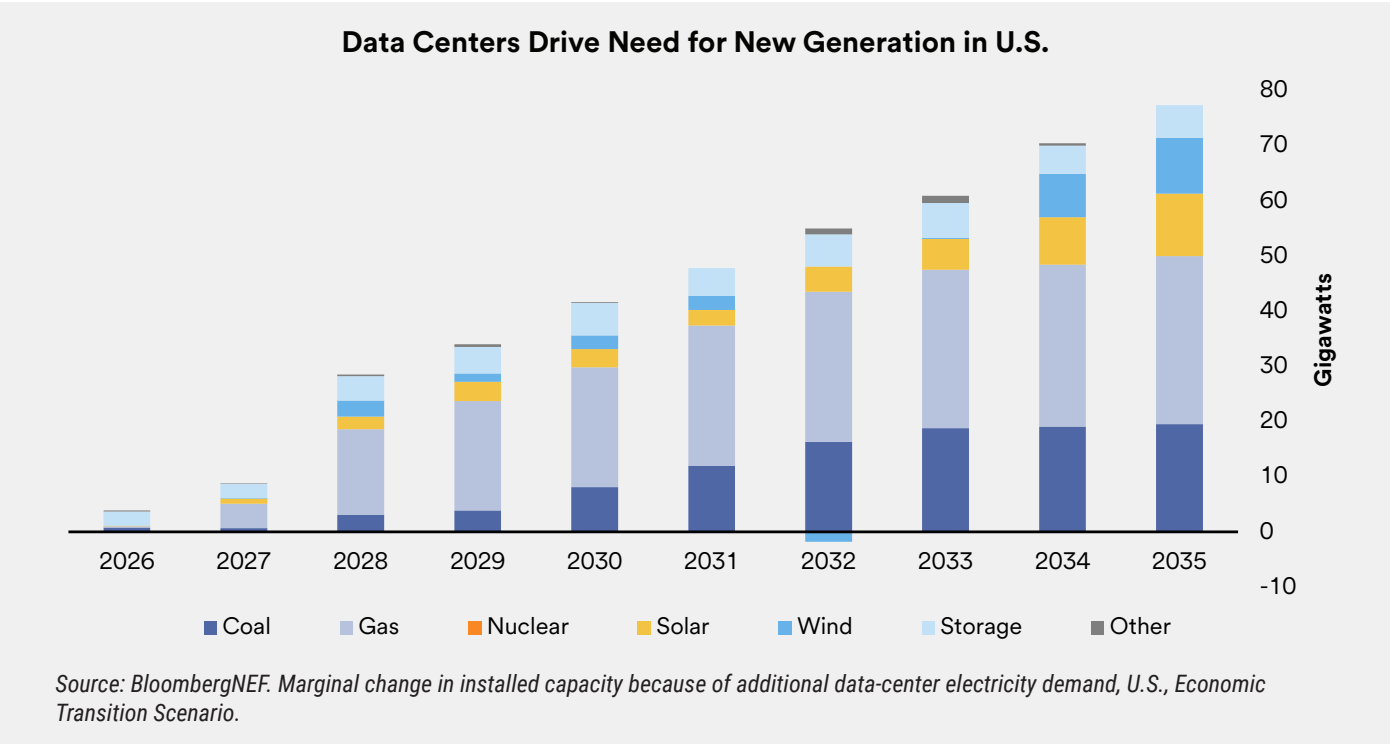
As data centers expand, their energy and water demands are increasing, raising concerns about sustainability and local resource management. Data centers have fluctuating demand throughout the day and need consistent energy availability to support high demand periods. The boom in data center construction has led to a backlog of data centers waiting for connection to the power grid, in some cases the wait could take five years according to the Lawrence Berkeley National Laboratory. The aging U.S. electrical grid needs significant infrastructure upgrades to support the much higher demand or create alternative backup systems to stabilize the grid during periods of intense usage.

Some data centers are going off-grid (or “behind the meter”, the energy industry’s term for generating one’s own power) by building their own power plants, mostly natural gas generators, to avoid waiting years to connect to the grid and quickly meet the enormous electricity demands of AI. Depending on the size of the data centers, the gas generation plants may not generate enough electricity to meet usage goals but is enough to start while it waits to connect to the grid. Once connected to the grid, the generators provide backup should high demand exceed the grid’s capacity.

Despite interest in renewable energy, all forms of renewable energy have intermittent generation and availability issues due to overcast weather or nighttime (solar), lack or change in wind patterns (wind), and draught (hydro). The government’s cancellation of renewable energy projects adds to the limited availability of this energy source. Use of renewable energy would require back-up alternatives such as gas, coal, nuclear, or battery power.

Nuclear power generation has been in the news about deals between technology and energy companies to restart shuttered nuclear power plants and build new small modular nuclear reactors (SMR) that will have a significantly smaller footprint than traditional nuclear plants. The SMRs are next-generation nuclear reactors with simpler designs for faster deployment and lower construction costs. Unlike other forms of electricity generation, nuclear plants will take much longer to build or refurbish. Traditional nuclear plants took around a decade to build.

Technology companies are also designing new semiconductor chips to require lower energy usage by improving efficiency and reducing cooling needs. There are increasing investments in battery storage systems to leverage all kinds of power generation – solar, wind, industrial gas turbines, and nuclear to provide back-up power to data centers.





Our Outlook

In the near term, profitability for the technology companies will be lower from high upfront investments in product development and data center construction before they recognize revenues from sales of products and services and usage of the data centers. Larger technology companies are actively investing in or purchasing small and start-up hardware and software companies working in the AI space to expand capabilities and ecosystems. The idea of a rising tide lifting all boats applies here, as a greater ecosystem of suppliers of AI products and services will increase and support corporate and consumer demand.

There is concern among investors that the industry may be in a bubble situation with large amounts of money invested in start-up companies with unproven products and huge capital investments in data centers that may exceed demand. Questions about data center profitability also hang in the air. We note that the largest technology companies have large and growing backlogs of unfulfilled contracts that are in the hundreds of billions of dollars due to capacity constraints. It takes approximately two years to build an AI data center, depending on the size and complexity of the infrastructure. During the construction phase, high capital expenditures will result in lower profitability measures, a decline in free cash flow, and higher leverage from increased debt. Revenues will jump as capacity comes online, and companies can start charging customers. The better rated technology companies all have significant and growing operating cash flows and solid balance sheets that allow for debt issuance to support the higher capital intensity. Still, profit margins may be lower than in the past as some data center operators tried to win large contracts with low bids.

Most U.S. corporations have not converted from on-premises data centers to the public clouds or to a hybrid cloud model. As companies decide to adopt AI capabilities to improve operating efficiencies and revenue generation, they will need to invest in new AI enabled software and modernize their hardware or move to a public or private cloud. However, concerns of tariffs, inflation, and an economic slowdown may delay corporate adoption of cloud and AI services.

In the long run, there is no doubt that AI will continue to affect how businesses operate, as more companies adopt AI products and services to improve operating efficiency. This will create additional opportunities for technology companies. We are only at the beginning of the adoption and monetization of AI. Much like at the start of any seismic technology change, there will be many companies that will benefit from providing new AI related products and services, while old technology providers will fall behind over time unless they can pivot and provide other value-added products and services. It is positive that inference usage (actual usage of LLMs, not just the training of LLMs) is increasing. More inference usage means higher adoption of AI. More adoption also means more data to continue LLM training. Improved LLMs mean even greater adoption. All of which increases data center usage. At some point within the next two to three years, we should see data center operators (hyperscalers, Oracle, etc.) start raising prices for using their data centers, which will improve profitability.

If you have any questions about this report, please reach out to your relationship manager.

Stay tuned for future commentary on the aerospace and defense sectors.

The views expressed within this material constitute the perspective and judgment of U.S. Bancorp Asset Management, Inc. at the time of distribution and are subject to change. Any forecast, projection, or prediction of the market, the economy, economic trends, and equity or fixed-income markets are based upon current opinion as of the date of issue and are also subject to change. Opinions and data presented are not necessarily indicative of future events or expected performance. Information contained herein is based on data obtained from recognized statistical services, issuer reports or communications, or other sources, believed to be reliable. No representation is made as to its accuracy or completeness.

PFM Asset Management serves clients in the public sector and is a division of U.S. Bancorp Asset Management, Inc., which is the legal entity providing investment advisory services. U.S. Bancorp Asset Management, Inc. is a registered investment adviser, a direct subsidiary of U.S. Bank N.A. and an indirect subsidiary of U.S. Bancorp. U.S. Bank N.A. is not responsible for and does not guarantee the products, services, or performance of U.S. Bancorp Asset Management, Inc.

NOT FDIC INSURED : NO BANK GUARANTEE : MAY LOSE VALUE

pfmam.com A Division of U.S. Bancorp Asset Management

pfm asset
management