



Insight

AI Data Centers: Why Are They So Energy Hungry?

SHUTING POMERLEAU, ANGELA LUNA | JULY 15, 2025

Executive Summary

- AI data centers have two unique energy-related features compared to traditional ones: They require enormous amounts of electricity, with generative AI consuming 10-30 times more energy than task-specific AI, as well as advanced liquid-based cooling solutions as they generate substantially more heat.
- Electricity demand from AI data centers is expected to grow rapidly from 4.4 percent of total U.S. electricity in 2023 to between 6.7-12 percent by 2028, with faster growth expected in the future as AI data centers continue to grow as more industries discover new AI use cases and adopt the infrastructure to deploy the technology.
- This insight provides an overview of the capabilities of AI data centers compared to traditional data centers, analyzes why AI data centers are so energy intensive, and explores the projected future electricity demand from the sector.

Introduction

Several tech giants such as Microsoft and Google have announced billions of dollars of investment projects in artificial intelligence (AI) data centers - which are critical in supporting generative AI. Compared to earlier forms of AI, generative AI is capable of producing new content, relying on massive datasets and requiring performance computing to run various tasks at the same time, including everything from training deep learning models to making real-time inferences.

AI data centers have two unique energy-related features compared to traditional ones: They have an enormous appetite for electricity, with generative AI consuming 10-30 times more energy than task-specific AI, and require advanced liquid-based cooling solutions as they

generate substantially more heat than traditional operations.

Electricity demand from AI data centers is expected to grow rapidly, from 4.4 percent of total U.S. electricity in 2023 to between 6.7-12 percent by 2028, with even-faster growth expected in the future as AI data centers continue to expand and as more industries discover new AI-use cases and adopt the infrastructure to deploy the technology.

This insight provides an overview of the capabilities of AI data centers compared to traditional data centers, analyzes why AI data centers are so energy intensive, and explores the projected future electricity demand from the sector.

What Are AI Data Centers?

A data center is a [facility](#) designed to house the computing infrastructure that powers digital services, from storing data to running applications. Typically, there are three components of data center [infrastructure](#): compute, storage, and network. Computing infrastructure includes mainframes and servers with varying levels of internal memory and processing power. Storage infrastructure is used for data and digital file storage. Network infrastructure refers to networking devices, such as cables and routers that connect the data center components to end-user locations.

Traditional data centers

Traditional data centers are engineered to support a broad spectrum of computing needs, including web hosting, enterprise applications, databases, and general cloud services. These centers also support AI workloads, mainly earlier forms of AI, known as [task-specific AI](#), which are systems designed to perform specific tasks or limited sets of tasks and include recommendation engines such as Netflix or virtual assistants such as Siri and Alexa.

AI data centers

AI data centers are critical in supporting generative AI. Since the rise of generative AI, the demand for data centers has increased and evolved, introducing infrastructure needs that differ significantly from those needed to support traditional computing and earlier forms of AI.

In contrast, generative AI models - capable of producing new content such as text, image and video - rely on massive datasets and require high performance computing to [run](#) various tasks at the same time, including everything from training deep learning models to making real-time inferences.

Unique Energy-related Features of AI Data Centers

AI data center infrastructure differs from traditional data centers in several ways: It requires substantially more powerful hardware per task, generates significantly more heat – necessitating advanced cooling systems – and supports continuous, real-time workloads for AI inference. Additionally, AI data centers demand higher storage capacity, faster internal networking, and more sophisticated systems to manage massive training and deployment tasks.

There are two main energy-related features that distinguish AI data centers from traditional ones: an enormous appetite for electricity and their utilization of advanced cooling solutions.

AI data centers require a substantially larger amount of power compared to traditional data centers. This is because AI [is trained on](#) a massive amount of data to identify the underlying patterns to create new content. Besides the high-performance computing capabilities, AI data centers also require advanced storage architecture and resilient networking. Generative AI is estimated to [consume](#) between 10–30 times more energy than task-specific AI. Specifically, using generative AI to create an image [requires](#) the amount of energy equivalent to fully charging a smartphone. The Electric Power Research Institute [estimates](#) that a traditional Google search uses approximately 0.3 watt hours (Wh), compared to 2.9 Wh for a ChatGPT query.

AI data centers also require advanced liquid-based cooling solutions as AI workloads generate significantly more heat than traditional operations. [Liquid cooling](#) is more effective than conventional air-cooling methods (e.g., using fans) to cool AI data centers because water has three times more heat capacity than air. Liquid cooling includes both [evaporative cooling](#) (using water to cool off hot air) and [pumping coolant](#) directly into a server to absorb heat emitted by processors.

Future Power Demand of U.S. AI Data Centers

The Energy Information Administration forecasts that the total U.S. electricity consumption [will grow](#) in 2025 and 2026 after the relatively flat demand from the mid-2000s to early 2020s. Electricity consumption is expected to grow at an average rate of 1.7 percent annually from 2020 to 2026, partly driven by the electricity demand from data centers.

A 2024 Department of Energy (DOE)'s report [estimated](#) that data centers consumed about 4.4 percent of total U.S. electricity in 2023 and are expected to consume between 6.7–12 percent of total U.S. electricity by 2028. DOE also projected that the total electricity usage by U.S. data centers will increase from 176 trillion-watt hours (TWh) in 2023 to 325–580 TWh by 2028 driven by data center expansion and AI applications.

The U.S. AI data center electricity forecast aligns with the global forecasts. [Goldman Sachs](#) forecasts that global electricity demand from data centers will increase 50 percent by 2027 and by as much as 165 percent by 2030 compared to the level in 2023.

The Big AI Data Center Projects Boom in the United States

The United States remains a global leader in data center infrastructure. As of early 2025, it [had the largest number](#) of active facilities in the world at over 5,400.

As various AI-powered tools such as digital agents and content generators transform workflows and daily activities in real time and [offer](#) productivity gains, investing in specialized AI data center infrastructure has become a strategic necessity for maintaining global competitiveness. In 2024 alone, major tech companies [spent](#) more than \$200 billion on AI and cloud infrastructure, and combined AI infrastructure spending from Microsoft, Amazon, Alphabet, and Meta is expected to [surpass](#) \$300 billion in 2025.

Several tech giants have announced large investment projects in AI data centers. For example, OpenAI, SoftBank, and Oracle this year announced their [Stargate Project](#), a four-year, \$500-billion initiative that will build a network of advanced AI data centers in Texas, powered by Nvidia's chips. [Microsoft](#) announced it would invest \$80 billion in building datacenter facilities in 2025. [Alphabet](#), Google's parent company, also announced it would invest \$75 billion in 2025 to expand its global data center footprint, while Meta [announced](#) plans to invest \$60-\$65 billion in AI infrastructure this year.

While most of today's data centers are [hybrid](#), meaning they have both traditional and AI capabilities, [projections](#) show that by the end of 2025, an estimated 33 percent of global data center capacity will be dedicated to AI applications, with that number reaching 70 percent by 2030.

Looking Forward

As AI data centers continue to grow, and as more industries discover new AI use cases and adopt the infrastructure to deploy the technology, electricity demand from AI data centers will increase rapidly. How the U.S. electricity market can keep up with the soaring energy demand from AI data centers is key to the successful expansion of the technology.