Power Consumption Trends from Increased AI and Data Center Utilization

Jordan Aljbour | Technology Innovation – Strategic Research

August 2023

Introduction

Artificial intelligence (AI) and data centers, the engines of the worldwide digital economy, are driving innovative solutions and novel business models across the transforming energy industry. However, their increased utility poses a set of significant challenges that EPRI is diligently assessing—most importantly the current energy consumption trends, contributing factors, and management strategies.

Research Question

"How do emerging trends in AI and data center usage impact US electricity infrastructure, and what insights can we gain from these trends?"

This subject explores the nexus of digital transformation, energy use, and environmental sustainability. It seeks to examine the implications of escalating AI and data center usage on US energy systems.

Power Consumption Insights

AI usage broadly, particularly large-scale enterprise models, are commonly tasked with Process Automation & Optimization (PAO), Predictive Analytics (PA), and Natural Language Processing (NLP), demanding intensive resources for training and operation. For instance, in 2020, ChatGPT-3, a prevalent generative NLP AI required ~10.57 GWh for both initial training and annual search request operation. A 2019 study indicated a single NLP AI model's carbon footprint equaling five average American cars' full lifecycle, emitting ~280 tons of CO2e. Despite these figures, potential for more efficient and sustainable AI implementation exists with broad adoption and technological advancements.

Data centers, supporting digital services like cloud computing and AI, are a significant energy consumer, though they constitute only a fraction of the market. From small-scale to large-scale commercial data centers, which include enterprise centers (*private*), colocation centers (*shared tenancy*), and hyperscalers (*global-scale facilities*). In 2020, US data centers accounted for 1.8% of total electricity consumption, akin to Wisconsin's annual usage. Globally, data centers represented 1% of worldwide electricity use, expected to rise with increasing digital reliance.

Factors Driving Increased Power Consumption

There are several key drivers behind increasing power consumption of AI and data centers:

1. **Explosion of Data:** The 'big data' surge necessitates substantial storage and processing in data centers, escalating their energy demands. Global IP traffic is set to triple by 2022, reaching 396 exabytes monthly or ~1.98e6 GWh; resulting in providing AI models with a valuable resource—diverse and vast datasets.
2. **Model Complexity:** With AI models becoming more intricate and data-intensive, they demand increased computational power, resulting in higher energy consumption. Resources needed for large AI training runs amplified by 300,000 times from 2012-2018.
3. **24/7 Operation:** Data centers' continuous operation for an "always-on" digital economy ensures relentless energy consumption. In 2020, US data centers used about 69 billion kWh, equivalent to the annual consumption of 6.4 million homes.

Data center distribution in the US

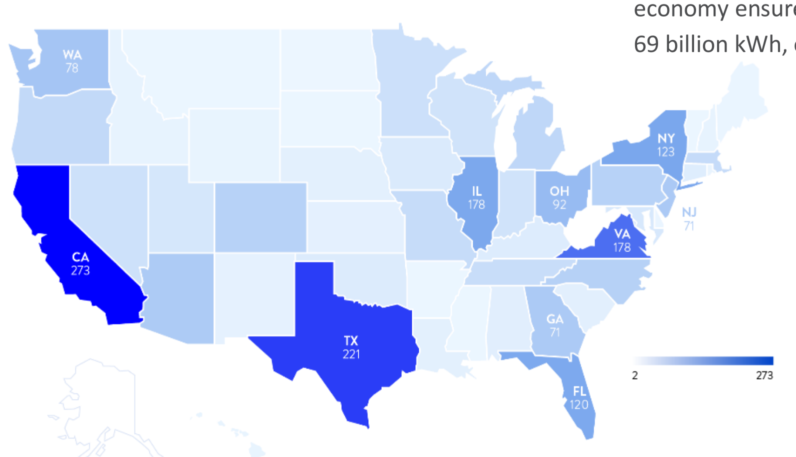

***Figure 1**. US Data Center distribution aggregated together by general classifications of commercial, co-location, and hyperscaler. As of 2022, approximately 2,701 data centers were operational in the U.S.—with the geographic concentration following the sequential distribution of: California, Texas, Virginia, and New York. Illustrating the socio-economic determinants of population density and grid infrastructure on data center location.*

Strategies to Manage Efficiency, Usage, and Environmental Impact:

1. **Efficient Algorithms:** Developing energy-efficient AI algorithms, such as pruning, quantization, and knowledge distillation, can offer high performance with lower power use, reducing AI models' computational demands.
2. **Hardware Enhancements/Optimization:** Adoption of energy-efficient hardware like TPUs and FPGAs can notably cut power usage. Plus, advanced cooling methods could reduce energy use by 40%, and server utilization optimization can counter energy waste.
3. **Greener Data Centers:** The use of renewable energy sources can help decrease ecological footprints. Tech companies, like Google, are increasingly investing in renewables for their data centers, targeting 100% carbon-free energy use by 2030.

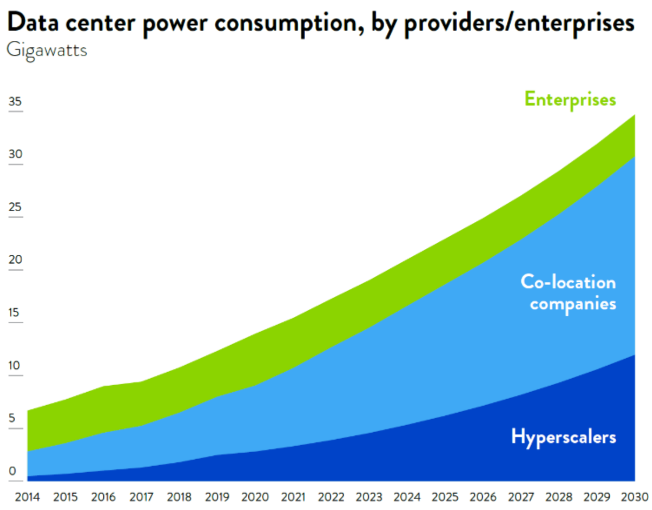

Figure 2. According to some industry sources, U.S. data center marketsize demand is forecasted to grow by an average of 7.1% annually until 2030. Suggesting that the current value of 148,920 GWh (2022) is estimated to increase to ~233,500 GWh by 2030. These estimates are posited from real-estate growth forecasting, wherein corresponding energy consumption ratios were applied to each data center type to gain a more accurate mosiac of how forecasted changes in US data center marketsize could impact power consumption.

Future Scenarios – Improved Efficiency, "Business as Usual", and Depreciated Efficiency:

1. **Optimistic Scenario:** Successful research leading to highly energy-efficient algorithms, hardware, and widespread green data centers could cut AI and data center energy use by 50% in a decade.
2. **Neutral Scenario:** With moderate adoption of energy-efficient technologies and data center design improvements, the energy consumption could remain stable over the next decade, despite expected growth.
3. **Pessimistic Scenario:** Without significant energy-efficient tech advances and adoption of green data center practices, escalating demand for digital services and data explosion could double AI and data centers' energy use in a decade.

KEY TAKEAWAYS

AI and data center systems, despite their transformative impact, pose substantial environmental challenges due to high energy consumption. This growing energy demand underscores a critical trade-off between digital innovation and environmental sustainability. However, advances in software and hardware efficiency present promising strategies to mitigate this energy footprint, paving the way towards a more sustainable future.

The future leans towards greener AI and data centers, but this direction hinges on relentless research and collaboration between tech companies and policymakers to establish best practices and regulations. In doing so, we can continue to leverage the transformative benefits of AI and databases while remaining cognizant of our environmental responsibilities.

NEXT STEPS

EPRI's ongoing research investigates the historical and projected trends for AI, ML, and data center load consumption, analyzing their impacts on regional demand growth and infrastructure planning. The study also examines load consumption during the AI & ML learning and user phase to provide comprehensive insights.

For more information, contact:

EPRI Customer Assistance Center
800.313.3774 • askepri@epri.com

3002027934 August 2023

EPRI

3420 Hillview Avenue, Palo Alto, California 94304-1338 USA • 650.855.2121 • www.epri.com