

The Gigawatt Bubble

Is the massive buildout of power infrastructure for AI sustainable?

A global race is on to build ever bigger and more powerful data centers to fuel the boom in demand for computing power generated by AI. Industry forecasts are already projecting that global electricity demand from data centers will more than double by 2030 to 4-6% of global power consumption, from 1.5% in 2024—equivalent to the energy use of Germany, the world's third largest economy.

GCC countries feature among the countries ramping up their ambitions around this roll out, taking advantage of their unique economic, technological and geographical advantages. Data center capacity in the GCC is expected to increase from **1 GW to 4 GW** or more over the next five years. If this growth materializes, data center demand will represent 3-5% of total GCC electricity consumption in 2030, an especially significant share given the region's comparatively later start in capacity building.

As a result, key players in the data center industry are preoccupied with ensuring that plans for new infrastructure are in place to handle the AI power boom. But what if the real risk is not being left behind, but instead the opposite: massive overbuild and stranded assets? Behind the current attention-grabbing forecasts lies tremendous uncertainty.

Multiple factors drive uncertainty in AI-driven energy demand growth

History shows that when capital investment far outpaces demand, as was the case with over-investment in fiber during the dotcom boom, a bust often follows. And the evidence is that capital expenditure for AI infrastructure is surging, including in the GCC. OpenAI's **UAE Stargate initiative** is expected to add 5 GW of data center capacity, of which 1 GW is slated for go-live in 2026. Recently announced initiatives from Saudi Arabia include the **US\$100 billion Transcendence AI Initiative** backed by the Public Investment Fund (PIF) and a **US\$5.3 billion commitment from Amazon Web Services (AWS)** to develop new data centers.

While the use of AI—including through large language models and, increasingly, AI agents—is indeed growing rapidly amongst businesses and consumers, the resulting implications for data center energy demand are not a foregone conclusion. While demand could double or triple by 2030, multiple factors could sharply bend down the growth curve.



The story of overheating AI power infrastructure could be a cautionary tale in the making.

First is compute and algorithmic efficiency. AI technology itself could deflate the AI power boom. Algorithmic and hardware breakthroughs—including AI chip efficiency and better data center design—are continuously reducing the computational power needed for a given level of AI performance. An acceleration of these trends could dramatically slow power demand growth. Performance gains from larger and newer AI models could also decelerate. While context-focused models trained on proprietary data will continue to push the frontier, **upgrades to existing models are showing some symptoms of the classic law of diminishing returns** as early breakthroughs give way to more incremental gains.

Second is a drought of data for training large language models. AI models will eat through essentially **all public text data by 2028-2032 at current scaling rates**. Training data scarcity could therefore become a serious limiting factor. Privacy and copyright barriers could also undercut the AI industry's implicit assumption of infinite data. The **EU's GDPR** and other privacy laws illustrate how regulators and content platforms are erecting barriers to data access.

Third are regulatory barriers to growth, including regulatory roadblocks that could affect the construction of data centers. In response to grid strain, policymakers are already devising new rules—including **Texas' Senate Bill 6**, the "Kill Switch Bill"—to ensure big new loads don't jeopardize reliability. Continuing caution from regulatory authorities could slow the pace and/or increase the cost of data center buildouts.

Building power sector resilience in a period of uncertainty

If the AI build-out overshoots actual needs, the fallout could be profound. For every 1 GW of newly built data center capacity that sits idle, **up to \$12 billion of data center investments and \$2 billion of power infrastructure investments** could be at risk. Scaling this up to 50-60 GW of stranded capacity by 2030—our estimate of the difference between a medium demand scenario and a high supply scenario—the risk to the power sector, exceeds \$100 billion. Under current regimes, this risk is often funded through the public exchequer or passed on to consumers, and consumers are already paying the price in de-regulated markets. Between 2020 and 2025, for example, **wholesale electricity costs more than doubled in Baltimore, USA, with some consumers reporting energy bill increases of 80%**.

The story of overheating AI power infrastructure could be a cautionary tale in the making. AI's growth will undoubtedly shape the future of power systems, but predictions of exponential growth may be only one possibility. Stakeholders across the GCC power sector—including sovereign wealth funds, investors, utilities, and technology leaders—need to consider whether they are building only for the exponential scenario, or whether they should plan to be resilient across multiple scenarios. That way, if this is a bubble that bursts, public finances and the cost to serve consumers will be protected.

Shihab Elborai
Partner
shihab.elborai@strategyand.pwc.com

Virender Vannam
Manager
virender.vannam@strategyand.pwc.com

Ramzi Hage
Partner
ramzi.hage@strategyand.pwc.com

Aditya Harneja
Principle
aditya.harneja@strategyand.pwc.com

www.strategyand.pwc.com/me