AI infrastructure: the race for data centres rages on

Image source: iStock/piranka

The global competition for AI-related technology is intensifying in 2025.

Adrien Pichoud
Chief Economist

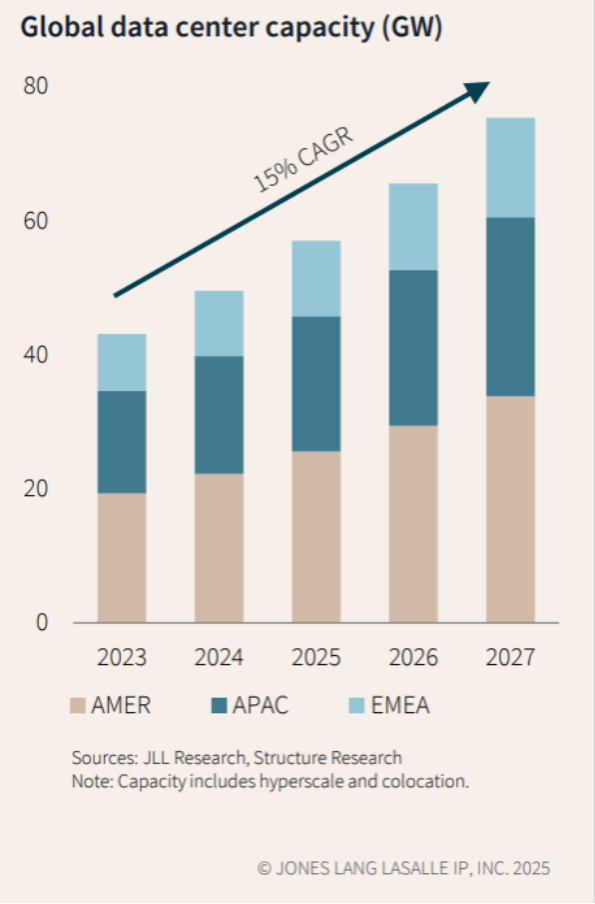
Jakub Dubaniewicz
Senior Equity Analyst

Emma Gillioz
Intern

SYZ PRIVATE BANKING

AI is rapidly transforming industries, from finance and healthcare to defence and autonomous systems, making it a focal point of strategic investment and regulation. The US has long sought to build and maintain technology leadership, a stance unlikely to change under Trump. The 'tech war,' particularly in AI, is poised to shape the coming years—if not decades.

Data centres have become a critical asset in this tech race. Countries and companies are investing billions in building and expanding data centre capacity to match the demand for computing power and data storage resulting from the surge in AI adoption. In the United States, tech giants are leading the charge, expanding infrastructure to support AI workloads. In China, the government and state-backed enterprises are investing in domestic data centre expansion to support AI development while countering trade restrictions and semiconductor shortages imposed by the US. Europe, meanwhile, is balancing investment growth with strict data sovereignty laws, sustainability mandates, and efforts to reduce dependency on foreign technology providers.



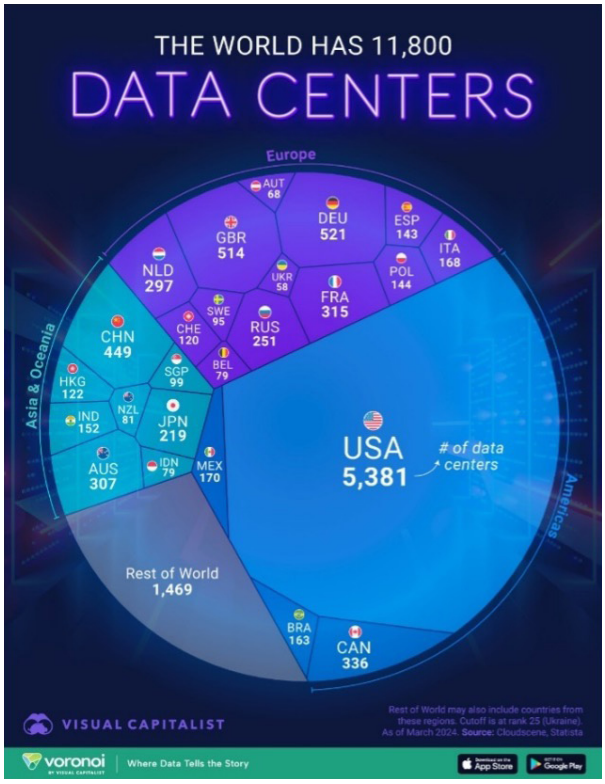
From the Stargate Project in the US to the DeepSeek release in China or the AI Action Summit in Europe, technological breakthroughs and gigantic investments in the AI field are announced almost daily across all major economies. The United States intend to maintain their leadership with a $500bn investment program centred around OpenAI and Oracle. China has become a serious contender and the release of DeepSeek is seen as a technological breakthrough challenging the established US-led framework.

Europe is at risk of being left on the sideline but appears to have finally received the message: a joint declaration from European Central Bank President Christine Lagarde and European Union President Ursula von der Leyen pledges to "do whatever is necessary to bring Europe back on track." The race for building AI infrastructures, data centres and the energy sources required to power them has already started.

The current data centres market

The United States is leading the race

With nearly 5,400 data centres, the US leads the race, prioritising rapid expansion through private investment and large-scale deployment. The most ambitious project to date, Stargate, is a joint venture between Oracle, OpenAI, and SoftBank. Announced by President Donald Trump in January, Stargate aims to build 20 data centres nationwide, with no government funding, for an estimated sum of $500 billion over four years. Along with Stargate – Meta, Amazon, Alphabet and Microsoft – are pouring tens of billions into its AI capabilities, while companies including Nvidia, Dell, and xAI are expanding their US operations.



China is coveting the top spot

China is making rapid progress in AI, despite currently having fewer than 500 data centres. The government has invested over 43.5-billion-yuan ($6.12 billion) in computing infrastructure and recently launched a 60-billion-yuan ($8.2 billion) AI investment fund. To lower costs and improve energy efficiency, it introduced the "Eastern Data, Western Computing" initiative, setting up eight major data hubs in key regions. By 2030, studies estimate relocating data centres could reduce emissions by 16-20% and bring $53 billion in direct economic benefits.

Europe must catch up

With Germany hosting around 520 data centres, Europe is not entirely out of the race. The combined infrastructure of all European data centres amounts to less than half of the US and European companies owned less than 5% of the region's data centre capacity as of 2023. However, recent announcements show that significant AI investments are also to be deployed in Europe: Microsoft and Amazon have committed €4 billion in France and €15.7 billion in Spain, respectively. Meanwhile, France announced on February 6 a plan to build Europe's largest AI campus, featuring a vast data centre funded by the UAE, with investments ranging between €30 billion and €50 billion. And the European Commission just launched InvestAI, a plan to mobilise €200 billion for investment in AI.

However, the European Commission is also balancing AI innovation with regulation. It has adopted a legal framework on AI in 2024, the AI Act, and has introduced sustainability rules for data centres. In parallel, it is investing €54 million in Open Euro LLM, an open-source AI project uniting 20 European companies, universities, and supercomputing centres. Still, AI's economic impact will also depend on its adoption and scaling—a key challenge for Europe, where labour productivity has been slowing. According to McKinsey, AI could boost productivity by up to 3% annually through 2030, reinforcing the need for accelerated digital transformation across the continent.

Challenges ahead

Tech is a battlefield for China's challenge to US global dominance

The United States has been tightening key technology export restrictions to limit China's AI progress, blocking access to advanced AI chips, high-bandwidth memory (HBM), and semiconductor manufacturing equipment. Despite these efforts, China is making progress in AI model training, chip production, and data centre infrastructure.

In September 2024, China Telecom's AI Institute revealed that its TeleChat2-115B model was trained using tens of thousands of domestically produced chips. Huawei, meanwhile, is shifting supply chains, sourcing memory chips from Chinese companies, such as ChangXin Memory Technologies (CXMT) and Yangtze Memory Technologies Corporation (YMTC), though some models still incorporate chips from South Korea's SK Hynix.

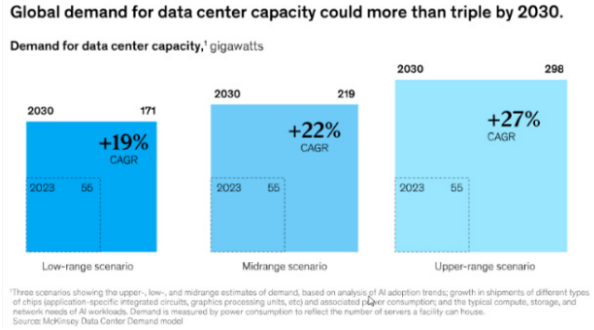
DeepSeek's R1 model has delivered results comparable to OpenAI's latest releases. Its optimised architecture drastically reduces computing power requirements and energy use. The efficiency gains that DeepSeek has achieved are open source, meaning that they will eventually become widely adopted by all industry players. Initially, skeptics suggested that this poses a risk of data centre overinvestment. However, the prevailing view is that more efficient models will drive wider usage, and therefore cloud giants like AWS, Azure or GCP have all stepped up their AI capex plans for 2025.

China's big tech firms are doubling down on AI infrastructure even if they remain far behind their large US competitors. ByteDance, the parent company of TikTok, has pledged $12 billion in AI development for 2025, including $5.5 billion for domestic AI chip production and $6.8 billion to expand model training. Its Doubao 1.5 Pro AI model already reached 78.6 million monthly active users this January, positioning it as a strong competitor, though still lagging behind ChatGPT's 180.5 million monthly active users. Meanwhile, the startup Moonshot AI's model, Kimi k1.5, claims to match or outperform OpenAI's o1 model in math, coding, and multimodal tasks.

Growing demand for data centres

DeepSeek has demonstrated that significant efficiency gains are possible in AI, and as technology evolves, more workloads may transfer to smaller data centres, which will be increasingly energy efficient.

McKinsey reported predictions that global data centre demand will triple by 2030, with annual growth rates between 19% and 27%. To keep up, cloud service providers would need to build twice as much capacity as has been constructed since 2000.

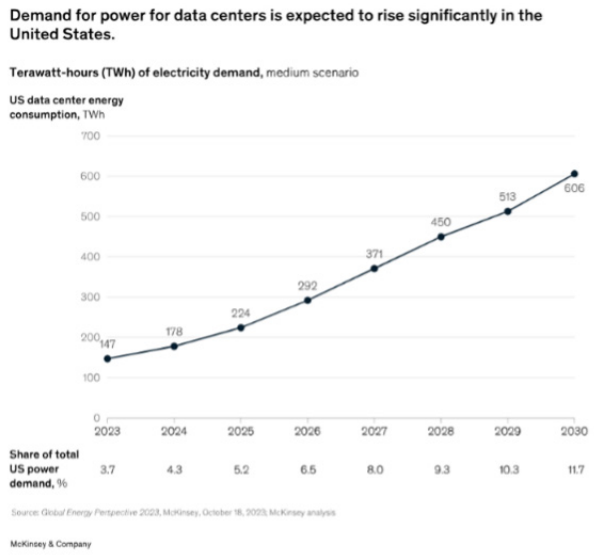


Source : McKinsey

Energy consumption

The rapid expansion of AI-driven data centres is putting pressure on energy and water resources. As AI workloads become more complex, companies are redesigning infrastructure to lower cooling costs, improve efficiency, and reduce power consumption. Colder regions are ideal for data centres due to lower cooling needs, but proximity to users remains crucial to avoid delays in data transmission.

By 2028, the U.S. Department of Energy estimates that data centres could consume 12% of the country's electricity, up from 4.4% in 2023, with half of new data centres potentially facing power shortages by 2027. To secure their energy supply, tech companies are turning to alternative sources. Google, Microsoft, and a data centre real-estate company Switch have invested in nuclear power, with plans to develop reactors. At the same time, startups like Oklo, backed by Sam Altman, and Radiant, which develops microreactors, are betting on next-generation nuclear technology to provide small-scale, efficient power solutions for AI infrastructure.



Welcome to Syzerland®

For further information

Banque Syz SA
Quai des Bergues 1
CH-1201 Geneva
T. +41 58 799 10 00
syzgroup.com

Adrien Pichoud
Chief Economist
adrien.pichoud@syzgroup.com

Emma Gillioz
Intern
emma.gillioz@syzgroup.com

Jakub Dubaniewicz
Senior Equity Analyst
jakub.dubaniewicz@syzgroup.com