IFRI
PAPERS

FEBRUARY
2025

AI, Data Centers and Energy Demand

Reassessing and Exploring the Trends

Geopolitics of Technology Center

Center for Energy and Climate

Laure DE ROUCY-ROCHEGONDE
Adrien BUFFARD

The French Institute of International Relations (Ifri) is a research center and a forum for debate on major international political and economic issues. Headed by Thierry de Montbrial since its founding in 1979, Ifri is a non-governmental, non-profit foundation according to the decree of November 16, 2022. As an independent think tank, Ifri sets its own research agenda, publishing its findings regularly for a global audience.

Taking an interdisciplinary approach, Ifri brings together political and economic decision-makers, researchers and internationally renowned experts to animate its debate and research activities.

ISBN: 979-10-373-1000-2

How to quote this publication:
Laure de Roucy-Rochegonde and Adrien Buffard, "AI, Data Centers and Energy Demand: Reassessing and Exploring the Trends", *Ifri Papers*, Ifri, February 2025.

Ifri

27 rue de la Procession 75740 Paris Cedex 15 – FRANCE

Tel: +33 (0)1 40 61 60 00 – Fax: +33 (0)1 40 61 60 60

Email: accueil@ifri.org

Website: Ifri.org

Authors

Laure de Roucy-Rochegonde has been the Director of Ifri's Center for the Geopolitics of Technology since February 2024. She was previously a researcher at Ifri's Center for Security Studies starting in March 2020, where she worked on the military applications of technology, artificial intelligence, normative conflict and arms control. With a PhD in Political Science, she is also an associate researcher at the Center for International Research (CERI, Sciences Po/CNRS). In parallel to her research activities, she teaches the ethics of war and arms control at Sciences Po Paris and at the University of Paris 2 Panthéon-Assas. She also holds a Master's degree in public policy and a Bachelor's degree from Sciences Po Paris, during which she spent a year in the War Studies department at King's College in London.

Adrien Buffard is a Master's student in International Energy Transitions at Sciences Po Paris. He previously interned at Ifri's Center for Energy and Climate. Beforehand, he worked on offshore wind planning and calls for tenders at the French Directorate-General for Energy and Climate, on the Baltic countries' energy dependence towards Russia at the French Embassy in Lithuania, and as an impact financial analyst at Bloomr. He holds a Bachelor of Arts from Sciences Po and a Bachelor of Literature from the Université Sorbonne Nouvelle.

Executive summary

The information and communication technologies sector today accounts for 9% of global electricity consumption, data centers for 1-1.3%, and artificial intelligence (AI) for less than 0.2%. The energy consumption of data centers was stable at around 200 terawatt-hour (TWh) per year until 2019, thanks to continued efficiency gains that totally offset skyrocketing computational and storage needs. However, efficiency gains then started to plateau and power usage effectiveness completely stalled, while the electricity needs of power data centers began rising exponentially to reach 460 TWh in 2022 (cryptocurrencies included), despite more efficient chips, algorithmic improvements and bigger data centers that have led to economies of scale. The growing energy demands of cloud services first, and now AI workloads (10% of today's data centers electricity demand), have exacerbated this trend. In the future, hyperscale data centers will gain shares amongst all kinds of data centers and AI will probably account for around 20% of data centers electricity demand by 2030.

If data center electricity consumption is booming at the global level (expected to be between 1,000 and 2,000 TWh in 2030), it remains a modest driver of the global growth in final electricity demand, far behind air conditioning and the electrification of industries. Since AI deployment offers a wide range of opportunities for the industry and energy sectors, ranging from efficiency gains to fusion monitoring and smart grids, it is pushing up the electricity demand in some regions where it is widely concentrated. In the US, where more than half of the world's data centers are located, they could make up to 13% of the total electricity consumption in 2030 (compared with 4% in 2024), representing 560 TWh of consumption then. In Europe, AI needs should account for 4 to 5% of total electricity demand by then (up from 2-3% in 2024).

Some uncertainties over AI's future impacts remain as the technology must cope with numerous challenges that will likely hamper its full deployment and usage in the years ahead. Grid queues are rising and data centers are reinforcing the pressure on already strained grids in the US and Europe. To cope with the surging electricity demand, additional dispatchable capacity generation is needed and data centers may be a long tailwind for gas production, especially when coupled with Trump's "drill, baby, drill' agenda. Likewise, the chip market concentration amidst a deepening trade war may slow down AI deployment. Other concerns and challenges encompass the lack of flexibility, water supply, and workers' skills.

Big Tech is playing a key role in the rollout of data centers. However, the fast deployment of AI is jeopardizing firms' climate targets as electricity growth and the need for better infrastructures are outpacing their green agenda. Still, with record power purchase agreements (PPAs) in renewables, nuclear power, and small modular reactor (SMR) start-ups –in addition to a large fleet of bidirectional charging electric vehicles–, Big Tech companies are increasingly becoming important energy actors and could further play a key role in providing demand response services, especially if data centers proved to be flexible. Meanwhile, Big Oil looks carefully at the AI irruption, reconsidering its renewables investments and willing to take advantage of the technology to decarbonize faster and further its drilling operations.

In the future, data centers will most likely be even more concentrated in some places. The US will remain their favorite host, whilst European data centers will prefer either countries with cheap and abundant power (Scandinavia and France) or where industries, finance and tax incentives are located (Germany, the UK, and Ireland). With the geopolitical competition on AI capabilities and megaprojects such as $500bn Stargate or France's €109bn announced investments, immense data centers will be built with capacities reaching 5 GW, providing incentives for Big Tech to look closer to traditional nuclear power plants. Leaders are raising awareness of the environmental impacts of AI, as the AI Action Summit in early February 2025 in Paris highlighted, insisting on the need for sustainable AI.

Résumé

Le secteur des technologies de l'information et de la communication représente aujourd'hui 9 % de la consommation mondiale d'électricité, les centres de données (*data centers*) 1 à 1,3 % et l'intelligence artificielle (IA) moins de 0,2 %. La consommation d'énergie des centres de données est restée stable autour de 200 TWh par an jusqu'en 2019, grâce aux gains continus en efficacité qui ont compensé l'envolée des besoins d'instances de calcul et de capacités de stockage. La demande électrique des *data centers* est alors devenue exponentielle et a atteint 460 TWh en 2022 (cryptomonnaies incluses), malgré l'utilisation de puces et la conception d'algorithmes plus efficaces ainsi que l'augmentation de la taille moyenne des *data centers* qui ont permis de réaliser des économies d'échelle. La demande croissante d'énergie du *cloud* d'abord, et maintenant de l'IA (10 % de la demande d'électricité des *data centers* aujourd'hui), a exacerbé cette tendance.

Si la consommation d'électricité des centres de données est en plein essor au niveau mondial (elle devrait se situer entre 1 000 et 2 000 TWh en 2030), elle reste un moteur modeste de la croissance mondiale de la demande finale d'électricité, loin derrière la climatisation et l'électrification des industries. Le déploiement de l'IA offre un large éventail d'opportunités pour les secteurs de l'industrie et de l'énergie, allant de gains d'efficacité à une meilleure modélisation de la fusion nucléaire, en passant par les réseaux intelligents. La demande d'électricité explose ainsi dans certaines régions où les centres de données sont largement concentrés. Aux États-Unis, où se trouvent plus de la moitié des centres de données du monde, ceux-ci pourraient représenter jusqu'à 13 % de la consommation totale d'électricité en 2030 (contre 4 % en 2024), soit 560 TWh de consommation à cette date. En Europe, les besoins en IA devraient représenter 4 à 5 % de la demande totale d'électricité d'ici là (contre 2 à 3 % en 2024).

Certaines incertitudes demeurent autour des effets futurs de l'IA, étant donné qu'elle doit faire face à de nombreux défis qui pourraient entraver son déploiement et son utilisation dans les années à venir. Les files d'attente pour la connexion aux réseaux des centres de données s'élargissent et renforcent la pression sur des réseaux déjà sous contraintes aux États-Unis et en Europe. Pour subvenir à l'explosion de la demande électrique, des capacités de production pilotables supplémentaires sont nécessaires. Les centres de données pourraient ainsi constituer un moteur de la hausse de production de gaz, en particulier aux États-Unis où ils profiteront de l'agenda trumpiste. De même, la concentration du marché des puces dans le contexte d'une guerre commerciale qui s'intensifie pourrait ralentir le déploiement de l'IA. Les

autres défis liés au déploiement de l'IA comprennent le manque de flexibilité, l'approvisionnement en eau et la formation des travailleurs.

Si les GAFAM jouent un rôle clé dans le déploiement des centres de données, le déploiement rapide de l'IA met en péril leurs objectifs climatiques, étant donné que la croissance de la demande électrique et des besoins en infrastructures dépassent la vitesse d'atténuation de leur empreinte carbone. Toutefois, avec un nombre record de PPAs dans les énergies renouvelables, l'énergie nucléaire, les petits réacteurs modulaires (SMR) – ainsi qu'une flotte importante de véhicules électriques à chargement bidirectionnel – les GAFAM deviennent des acteurs énergétiques de plus en plus importants et pourraient jouer un rôle clé dans la gestion de la demande électrique, en particulier si les centres de données s'avèrent flexibles. De leur côté, les géants des hydrocarbures examinent attentivement l'irruption de l'IA qui les pousse à reconsidérer leurs investissements dans les énergies renouvelables et à envisager l'emploi de cette technologie pour décarboner leurs activités plus rapidement.

À l'avenir, les *data centers* seront probablement encore plus concentrés dans certains endroits. Les États-Unis resteront le premier pays accueillant des *data centers*, tandis qu'en Europe ce sont les pays où l'énergie est abondante et bon marché (Scandinavie et France) ou ceux où se trouvent les industries, la finance ou des aménagements fiscaux (Allemagne, Royaume-Uni et Irlande) qui les accueilleront. L'annonce des projets prométhéens de *Stargate* aux États-Unis (500 milliards de dollars) et les 109 Md€ promis par la France augurent la construction de gigantesques centres de données dont les capacités atteindront les 5 GW, ce qui devrait inciter les géants de la *tech* à considérer d'autant plus l'utilisation de centrales nucléaires traditionnelles pour leurs *data centers*. Le sommet pour l'action sur l'IA, tenu début février 2025 à Paris, a rappelé l'importance de concevoir une IA « durable » et compatible avec la transition énergétique.

Table of contents

Introduction

The day after his inauguration, Donald Trump announced the launch of what the founder of OpenAI, Sam Altman, described as "the most important project of our time." Dubbed Stargate, the initiative aims to invest $500 billion over the next four years in the development of artificial intelligence (AI). Backed by three tech giants—cloud computing specialist Oracle, Japanese investment firm SoftBank, and the now-ubiquitous parent company of ChatGPT, OpenAI—Stargate's goal is to "build the physical and virtual infrastructure to power the next generation of AI" while creating "over 100,000 jobs" in the United States.

At the heart of this massive project lies the construction of data centers. AI technologies rely entirely on the ability to capture, store, and process data, making data centers a crucial piece of the puzzle. Thence, according to Oracle's chairman, Larry Ellison, Stargate will primarily focus on establishing these centers across the U.S. This marks a turning point for OpenAI, which has so far relied on infrastructure provided by its partner Microsoft but will now be able to leverage its own data centers. In a report published in September 2024, Sam Altman's company urged the U.S. government to support the creation of giant five-gigawatt data centers, with an estimated cost of $100 billion overall.

Trump appears to have heeded this call. The new president has announced plans to declare a "state of emergency" to remove regulatory barriers to data center construction—despite concerns over their environmental impact. He has also suggested that Stargate should generate its own electricity, but he did not provide further details on how this would be achieved. Some estimates suggest that data center energy consumption could triple by 2028—an astonishing figure.

However, just days after this high-profile announcement in the U.S., several American tech companies saw their stock prices plummet on Wall Street. Nvidia—the world's most valuable publicly traded company—lost nearly $600 billion in market capitalization on Monday, January 27, 2025. The reason? The breakthrough success of DeepSeek, a Chinese competitor to ChatGPT, which has quickly become the top-downloaded AI app across multiple platforms worldwide. DeepSeek is also touted as being far more energy-efficient than its competitors. Reportedly, it required only $5.6 million to train—compared to the $100 million spent on ChatGPT-4—and used computers equipped with just 2,000 Nvidia chips, whereas ChatGPT required 16,000 more advanced ones.

This drastic reduction in computing needs not only slashes costs but may also significantly lower the energy demand of data centers. If DeepSeek were to replace models like OpenAI's, overall energy consumption would drop considerably. As a result, DeepSeek's rise also triggered a sell-off in energy stocks, alongside the tech sector downturn. U.S. energy giant Constellation Energy, which signed a deal in 2024 to supply Microsoft with electricity, saw its stock price plunge by more than 20% on the Nasdaq.

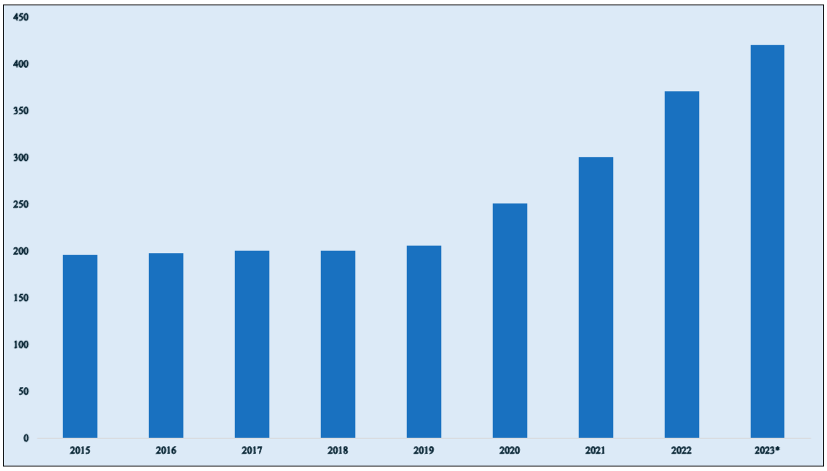
With the AI Action Summit, held in Paris on February 10-11, 2025, emphasizing the urgent need for "environmentally sustainable AI," it is now crucial to assess the true impact of AI and data centers on the energy sector—and whether the grid can handle this soaring demand. This report aims to address these questions.

The Rise of Data Centers

The Recent Soaring Electricity Demand as Efficiency Gains Level Off

The information and communication technology sector in 2022 accounted for 9% of global electricity consumption and data centers for 1-1.3%.[1] Until 2018, the electricity consumption of data centers was constant, with annual global electricity consumption plateauing at around 200 TWh (Figure 1) despite the surging number and average size of centers. From 2010 to 2018, power demand only rose by 6% in the time it took for computing activity to jump 550% and global installed data storage capacity to multiply by 26 (Figure 2). This great achievement was mainly due to efficiency gains. The efficiency of data storage improved ninefold over the 2010-2018 period, augmenting at a compound annual growth rate (CAGR) of 20%. The power usage effectiveness ratio drastically reduced from 2.50 in 2010 to 1.58 in 2018, thanks to great improvements in cooling systems and in auxiliary components, like security systems, lighting and power supplies. They respectively used to account for 40% and 20% of data center electricity consumption,[2] whereas they now represent around only 25% and 12%. This left open the possibility for the increased importance of data centers' intense energy usage (i.e., IT equipment accounts for 60-65% of the consumption today).

Figure 1: Global electricity demand from data centers (TWh) (excluding cryptocurrencies), 2015-2023



Source: IEA, Cisco, Goldman Sachs, GECF

1. "World Energy Outlook 2024", International Energy Agency, 2024, p. 186.
2. "Analyzing Artificial Intelligence and Data Center Energy Consumption, 2024 White Paper", Electric Power Research Institute, 2024.

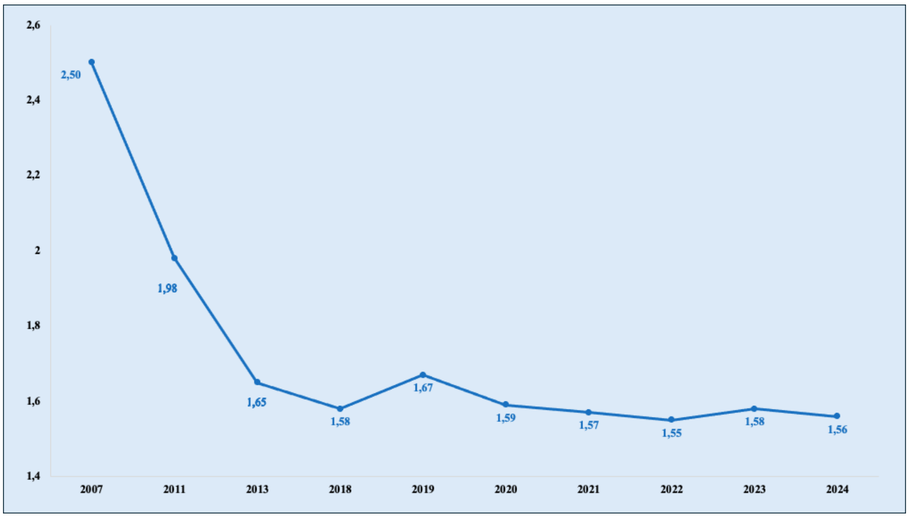
Figure 2: Trends in DC energy-use drivers, 2010-2018

Service demands have risen…	Relative change from 2010 to 2018 (mutlplier)
Global installed storage capacity (exabytes)	26
Global data centre internet protocole traffic (zettbaytes/year)	11
Data centre workloads and compute instances (millions)	6.5
Global installed base of servers (millions)	1.3
… while energy efficiency has increased	**Relative change from 2010 to 2018 (mutlplier)**
Average PUE	0.75
Typical server energy intensity (Wh per computation)	0.24
Average number of servers per wordload	0.19
Average storage drive energy use (kWh/terabyte)	0.11

Source: Masanet and al. (2020), Berkeley Lab

Yet, from 2018 onwards, data center-related electricity demand has soared. In 2022, data centers consumed around 350 TWh, excluding energy used for cryptocurrency mining, which was estimated to be around 110 TWh.[3] The same year, these data centers had the capacity to consume a combined total of 508 TWh of electricity if they were run constantly.[4]

Figure 3: Average data center's power usage effectiveness worldwide, 2007-2024



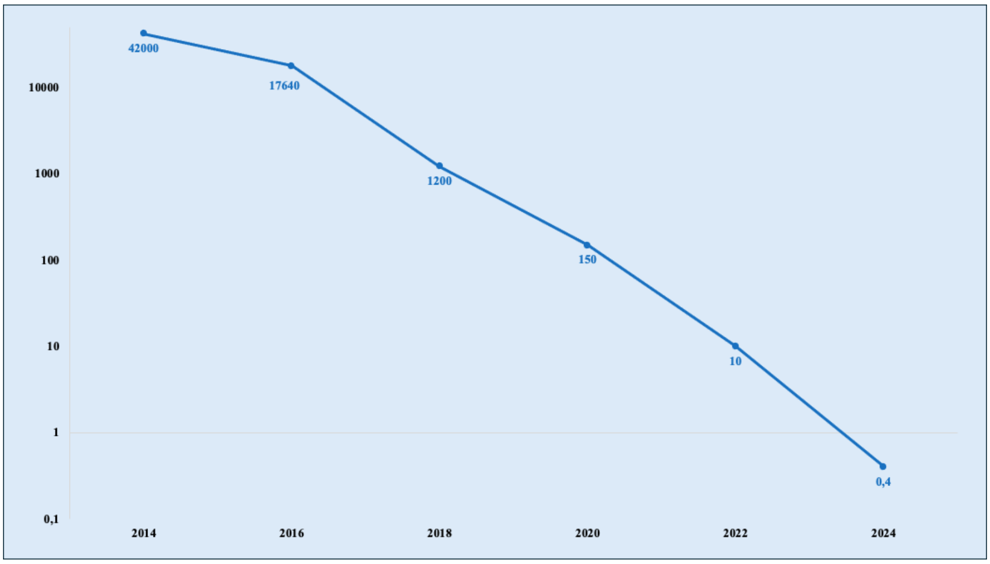
Source: Uptime Institute

If substantial technical efficiency improvements were achieved over the last decade, leading to a moderate data center overall consumption, the rate of technical efficiency gains has flattened in some aspects and is expected to continue doing so.[5] This led to sudden important growth. Indeed, the average

3. "Data Centers and Data Transmission Networks", International Energy Agency, February 2, 2025, available at: https://www.iea.org (based on Cambridge Bitcoin Electricity Consumption Index).
4. J. Saul *et al.*, "AI is Already Wreaking Havoc on Global Power Systems", *Bloomberg*, June 21, 2024.
5. E. Masanet *et al*, "Recalibrating Global Data Center Estimates", *Sciences Magazine*, February 28, 2020, pp. 984-986; D. Bizo, "Global PUEs - Are they Going Anywhere?", Uptime Institute, December 4, 2023, available at: https://journal.uptimeinstitute.com; "World Energy Outlook 2024", *op. cit.*, p. 186.

power usage effectiveness (PUE)[6] ratio of data centers has completely stalled in recent years, standing at 1.58 in 2023, the same level as in 2018 (Figure 3), even though in top modern data centers, with advanced technologies, despite the higher need for cooling due to more sophisticated GPUs, PUE can be driven down to 1.2. In some advanced countries, such as France, the average PUE is even 1.7[7] when 1.2 should be the standard cap. This means that previous improvements in cooling systems and auxiliary components have plateaued in recent years.

Figure 4: Nvidia's AI-chips efficiency (in J/Token), 2014-2024



Source: NVIDIA

Note: The graph uses a logarithmic scale. 2020, 2022 and 2024 respectively represent the A100, the H100 and the B100 Nvidia chips.

By contrast to PUE, chips (the main element of data centers both for storage as well as training and running AI) have continuously improved and have played a key role in bringing down energy consumption. Thanks notably to the emergence of "chiplets" that have allowed manufacturers to break up large monolithic chips into smaller, more efficient modular components and thus improve performance and scalability, the energy efficiency of AI-related chips has drastically improved, at an average compound pace of 4 times per year, meaning a 100,000-fold improvement in just 10 years (Figure 4). However, more efficient chips may also increase demand due to the rebound effect, some even thinking AI could become a new case study of the Jevons paradox.[8] Thus, energy efficiency improvements appear unlikely to fully offset the growing service demands due in part to AI.[9]

6. The power effectiveness ratio (PEU) is calculated by dividing the total energy consumed by a data center (including cooling and lighting) by the energy used solely for computing. A theoretical optimum value of 1 means that all energy is used for computing alone.

7. "Avis d'experts – Les data centers ou centres de données", Ademe, November 2024.

8. Microsoft CEO Satya Nadella wrote "Jevons paradox strikes again!" on X (ex-Twitter) on January 27, 2025.

9. "World Energy Outlook 2024", *op. cit.*, pp. 186-189.

Data centers' average size is the last key element to consider regarding the global evolution of their average efficiency. The concentration of bigger data centers contributed strongly to reducing the electricity consumption required for IT operations, thanks to economies of scale. Small-scale ones (0.5 to 2 MW), including *edge data centers* (located close to the end-user), represent about half of all data centers but only 10% of their load.[10] Large-scale one can be owned and run on-premises for enterprises (20-30% of the total load) or be hyperscalers (60-70% of the total load), immense facilities able to rapidly scale up their operations to meet the vast computing needs of cloud giants like Amazon AWS, Google Cloud, and Microsoft Azure.

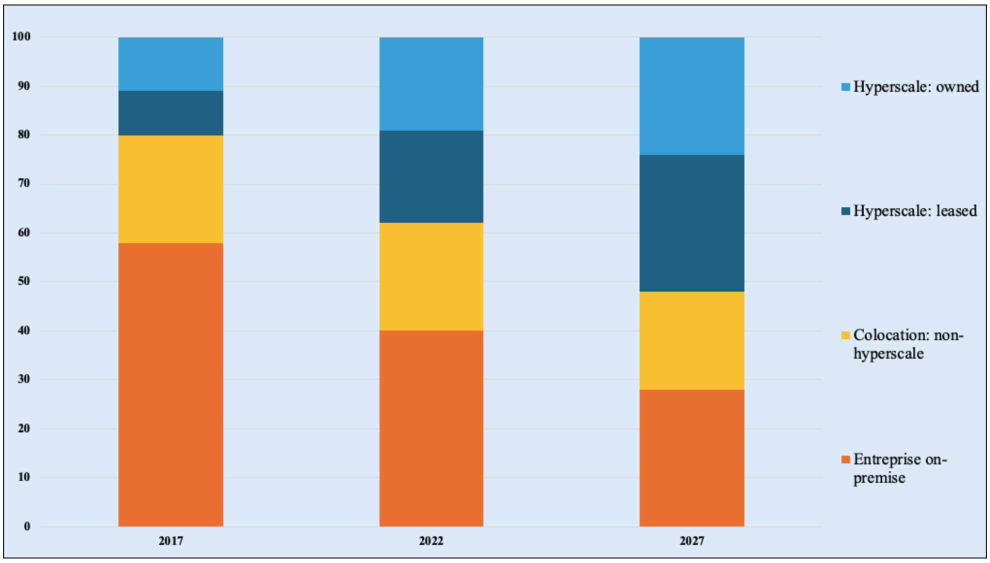
Data Centers' Growing Size and AI's Growing Role

Data centers are getting bigger, both in size and in capacity. As companies last decade reshaped their IT investments towards cloud services and were reticent about spending on their own data centers, the leading cloud providers have rapidly built huge global networks of hyperscale data centers. Hyperscale operator growth was further fueled by the rapid development of more consumer-oriented digital services such as social networking, e-commerce and online gaming. The average size of data center facilities worldwide is now 412,000 square feet (about 38,300 square meters), an almost fivefold increase from what it was in 2010.[11] In 2022, globally, 37% of data centers were hyperscalers and 23% were non-hyperscaler co-locations. Indeed, the share of on-premises data centers is shrinking (Figure 5), despite the projected growth of the global edge data center market of 22.1% at a compound annual growth rate (CAGR) through to 2030.[12] This trend is expected to continue. Around 70% of data center growth is expected to be fulfilled directly or indirectly (via cloud services, for instance) by hyperscalers by 2030.

It should be noted that the push in demand for data center services has firstly been led by higher needs and the soaring use of cloud services way beyond AI's requirements that pushed it later on, although AI also uses cloud services. AI currently accounts for only 10-15% of data center's energy consumption. Today it represents less than 0.2% of global electricity demand, although it is poised to surge. Estimates suggest that data center power demand will grow by 160% by 2030, and that by 2028 AI is expected to represent about 19% of the total power demand by data centers.[13]

10. "Analyzing Artificial Intelligence and Data Center Energy Consumption, 2024 White Paper", *op. cit.*
11. DC Bite, cited in J. Saul *et al.*, "AI is Already Wreaking Havoc on GLobal Power Systems," *op. cit.*
12. "Analyzing Artificial Intelligence and Data Center Energy Consumption, 2024 White Paper", *op. cit.*
13. "AI is Poised to Drive 160% Increase in Data Center Power Demand," Goldman Sachs, May 14, 2024, available at: https://www.goldmansachs.com.

Figure 5: Data center capacity trends, 2017-2027



Source: Synergy Research Group

AI and Data Centers: A Twofold Revolution

Massive Energy Consumption

Globally, the demand for data centers is growing rapidly. The International Energy Agency estimates that data center electricity consumption (including for cryptocurrencies) will overshoot 800 TWh in 2026, compared with 460 TWh in 2022.[14] This forecasts an astonishing 75% increase in four years. In the high-growth scenarios, it may even surpass the 1,000 TWh at the end of next year.[15] According to S&P, consumption may even reach 1,300 TWh in 2026. If we look forward to 2030, anticipations range from 1,100 TWh[16] to an impressive 2,000 TWh[17] which would respectively account for 3-4% and 6-8% of global electricity final consumption.

The huge variations over the exact amount of electricity that data centers will require come from the uncertainties on the use and deployment of AI, and possibly also, vested interests to portray AI as an extraordinary game changer. Indeed, AI leads to significant differences in the electricity consumed, depending on the training (from scratch or from pre-trained models), the number of parameters, and most importantly on the type of AI tasks made (Figure 6*)*. A ChatGPT text demand will typically only require 0.0005 Wh of energy whilst generating an image uses 0.5 Wh: i.e., 1,000 times more, but about 2,000 times *less* than the energy consumption of a washing machine cycle. Thus, even the energy to run a simple ChatGPT request is 10 times higher than a Google search, and it is estimated that using generative AI such as ChatGPT in each Google search would lead to annual electricity consumption of 29.2 TWh, the equivalent annual output of four to five nuclear reactors.[18] If this figure sounds grand, it is still less than the hundreds of extra TWh that forecasts indicate will be needed in the coming years. The growth will come from other uses too, mostly live-speech recognition, image captioning and generation. Regarding AI training, it does consume a lot of energy, but at the aggregate level, it represents a small portion of the energy consumed when it comes to widely deployed models like ChatGPT, whose 4th version training is

14. "Electricity 2024 – Analysis and Forecast to 2026", International Energy Agency, January 2024, p.31.
15. *Ibid*.
16. "AI is Poised to Drive 160% Increase in Data Center Power Demand," *op. cit*.
17. Schneider Electric Global Analysis.
18. J. Anderson *et al*., "Power of AI: Wild Predictions of Power Demand from AI Put Industry on Edge," S&P Global, October 16, 2023, available at: https://www.spglobal.com.

estimated to have required 50 GWh (the output of a French nuclear reactor running at full capacity for two days).[19]

Yet overall, data processing is a small driver of electricity demand growth at the global level. By 2030, data centers will account for around 5% of total electricity demand growth, which is roughly on par with demand growth for desalination, and less than a third of the demand growth for both EVs and air conditioning in the building sector, and less than a sixth of the electrification of industry.[20]

Figure 6: AI energy demand per task, 2024

Type of task	Average energy required (based on 20 popular open-source models)	Multiplier (text classification as a reference)
Image Generation	0.477 Wh	954
Image Captionning	0.109 Wh	218
Speech Recognition	0.025 Wh	50
Question Answering	0.001 Wh	2
Text Classification	0.0005 Wh	1

Source: Sasha Luccioni and al. (2024)

Data centers are clustered, and they tend to be even more so with time. In recent years, most new data centers have been installed in the US. While it accounted for slightly more than 1/3 of the 8,000 data centers in 2020, the US hosted more than half of them (5,381 out of 10,655 data centers globally), as of March 2024.[21] Furthermore, the concentration is also visible within the country: fifteen states account for 80% of the national data center load. However, while the US largely dominates the sector (51%), Europe also plays an important role by welcoming around 15% of the world's data centers: Germany (521), the UK (514), and France (315) are the top 3, with Ireland far behind in the rankings (around 15th-20th, with less than 100 data centers). The other main actors are China, that accounts for 4% (449 data centers as of March 2024 and so is 4th in the world), and Canada for 3% (336 data centers and the 5th country).

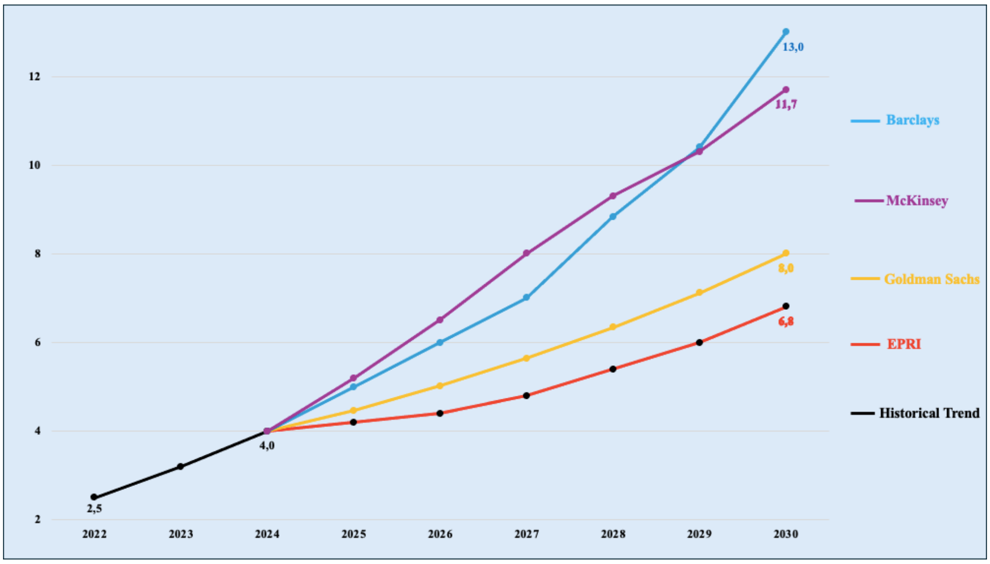
Thus, the impact on electricity demand varies depending on the location. In the US, data centers today account for 4% of total electricity consumption,[22] and the sector has already surpassed 10% of electricity consumption in at least five US states. In Virginia, it consumed ¼ of the total electricity in 2023. While data centers in the European Union used 55 TWh of electricity, which is 2% of their total electricity consumption,[23] in some

19. T. B. Minde, "Generative AI Does Not Run on Thin Air", Research Institute of Sweden, October 10, 2023, available at: https://www.ri.se.
20. "World Energy Outlook 2024", *op. cit.*, p. 187.
21. "Analyzing Artificial Intelligence and Data Center Energy Consumption, 2024 White Paper", *op. cit.*
22. *Ibid.*
23. P. Bertoldi and G. Kamiya, "Energy Consumption in Data Centres and Broadband Communication Networks in the EU", Publications Office of the European Union, Luxembourg, 2024.

specific countries like Ireland, data centers have already surpassed the impressive 20% mark of domestic electricity demand. In France, they account for 42% of the total digital load, which is around 10% of French electricity consumption.[24]

Forecasts are different too. The United States is expected to be the fastest-growing market for data centers, expanding from 25 GW of capacity and 175 TWh of annual consumption in 2023 to more than 80 GW of capacity and 560 TWh in 2030.[25] [26] By 2030, data centers are thus expected to account for 7 to 13% of US electricity consumption (Figure 7). They will even be the first driver of electricity demand growth. Between 2024 and 2030, electricity demand for data centers in the United States is expected to increase by about 400 TWh (almost France's annual electricity consumption) at a CAGR of about 23%.[27] Thus, out of the 2.4 percentage growth in electricity consumption in the US in the 2022-2030 period, 0.9 percentage points will be tied to data centers.[28]

Figure 7: US data center electricity consumption forecasts, 2022-2030



Source: Ifri, based on EPRI, Goldman Sachs, Barclays Research, McKinsey

In Europe, the impact on the grid will be less significant. Data centers capacity in the EU, Norway, Switzerland and Britain is forecast to hit 35 GW of capacity and 150 TWh of annual consumption by 2030, up from 10 GW and 62 TWh today, according to McKinsey.[29] Europe's data centers are thus expected to account for 5% of the continent's total electricity consumption

24. "Avis d'experts – Les data centers ou centres de données", *op. cit.*
25. A. Green *et al.*, "How Data Centers and the Energy Sector Can Sate AI's Hunger for Power," McKinsey, September 17, 2024, available at: https://www.mckinsey.com.
26. "AI Revolution: Meeting Massive AI Infrastructure Demands", Barclays Research, January 16, 2025.
27. A. Green *et al.*, "How Data Centers and the Energy Sector Can Sate AI's Hunger for Power," *op. cit.*
28. "AI is Poised to Drive 160% Increase in Data Center Power Demand," *op. cit.*
29. A. Granskog *et al.*, "The Role of Power in Unlocking the European AI Revolution," McKinsey, October 24, 2024, available at: https://www.mckinsey.com.

then (vs. 2% today).[30] By 2030, the power needs of European data centers may match the current total consumption of Portugal, Greece, and the Netherlands combined (220 TWh), according to Goldman Sachs, or even 287 TWh in Beyond Fossil Fuels high-demand scenario.[31] Even though these assumptions may be too bullish, such a deployment of data center infrastructure would yield in a capture of 20% of additional renewables projects developed by the EU by 2030, if they were all to be powered with renewable electricity sources. However, some European analysts are more cautious about the demand growth in Europe, and the European Commission expects data centers in the European Union to consume 98.5 TWh in 2030, which is less than the Netherlands.[32] Regarding Ireland, without policy restrictions to access the grid, data centers could consume one-third of the grid's electricity in 2026.[33]

AI and Data Centers: Catalysts of Opportunities

AI is supposed to offer a wide range of possibilities for the economy. Generative AI could help create between $2.6 trillion and $4.4 trillion in economic value throughout the global economy. But achieving just a quarter of this potential by the end of the decade would require between 50 and 60 GW of additional data center infrastructure in the United States alone.[34]

Regarding industry more generally, AI-driven near-term efficiency gains could amount to 30%, according to Siemens' Data Center Segment CTO. But such game-changing gains must cope with the absolute need to educate and train workers in AI uses and possibilities, which will probably take time.

Regarding the energy sector, AI could boost some breakthrough technologies. Coupled with high-performance computing (HPC), AI could help study fusion plasma stabilization and behavior, to anticipate plasma instabilities and so adjust controls within milliseconds. This could make fusion reactions more stable, enabling real-time control. AI could also be used to improve the efficiency of solar cells and the performance of batteries, which require immense computing power, as well as improve the function of nuclear fission reactors (through predictive maintenance). Smart grids, water management, geothermal energy and carbon capture, utilization and storage (CCUS) should also benefit highly from AI and HPC.

30. *Ibid.*

31. J. McArdle and P. Terras, "System Overload: How New Data Centres could Throw Europe's Energy Transition Off Course", Beyond Fossil Fuels, February 2025.

32. Commission Delegated Regulation (EU) 2024/1639 of 14 March 2024 on the first phase of the establishment of a common Union rating scheme for data centres, p.1.

33. "Electricity 2024 – Analysis and Forecast to 2026", *op. cit.*, pp. 32-33.

34. A. Green *et al.*, "How Data Centers and the Energy Sector can Sate AI's Hunger for Power," *op. cit.*

AI may also help Big Oil to bring down its greenhouse gas (GHG) emissions. Abu Dhabi hydrocarbon champion Adnoc is planning to roll out AI agents to act autonomously across all its operations. Adnoc claims to have allocated $23 billion to develop low-carbon technology using AI and is "very optimistic" it could achieve its carbon reduction targets ahead of schedule because of the new technology. However, AI's role in decarbonizing Big Oil is yet to be proven.

Furthermore, at present, the reuse of heat generated by data centers is limited. The potential for recovering waste heat from data centers in France is estimated at around 1 TWh per year,[35] which represents around 6% of the total power data centers consumed in 2024.

There is still little view on data centers' flexibility potential. It appears to the authors that if, theoretically, data center operators can convert their facilities to create more flexible loads for power grids, this is quite rare in practice. The first reason is that when it comes to storage and cloud (~50% of today's data center load), it is difficult to vary or to shift electricity demand. Regarding AI (~10% of total data centers' load), training and inference (i.e., user requests) should be analyzed separately. Regarding training, once a training phase is launched, it cannot be stopped until it is completed, even if the intensity of calculations (for instance, with combined HPC) can be slightly modulated. The only way to accommodate to the grid would be thus to launch training phases at off-peak time. Regarding inference, a dichotomy exists between general AI like large language models (LLM) or widely used AIs like ChatGPT (the chatbot), and specific AI designed for some precise tasks and thus for particular companies. For the former, it is not likely that demand is flexible (i.e., can be shifted or rescheduled). The need for immediate response rules out off-peak rescheduling, especially when considering live voice recognition. Plus, it seems those requests are treated centrally in hyperscalers and would only rarely be transferred to edge data centers connected to less strained grids (up to 10-15% of times). By contrast, for the latter (specific AI), a possible match with the daily generation cycle of renewables could exist. It is possible to imagine a widespread use of specific tasks (like for AI requests) during working hours when solar PVs receive daylight, followed by reduced electricity needs when workers are sleeping and not using devices. It is also easier to scale these solutions as less infrastructure is needed, making them more flexible. The same day-use night-off logic could be applied to software as a service (SaaS) solutions and enterprise servers –that together account for ~20% of today's data center load.

35. "Avis d'experts – Les data centers ou centres de données", *op. cit.*

AI's Challenges and Constraints on Data Centers

Although it offers a wide range of opportunities across all sectors, AI and data centers will both have to overcome several systemic challenges. We can list four factors that are likely to hamper its full deployment: i) grid constraints and queues; ii) the immense need for investments in infrastructure and power generation capacities; iii) the trade war in a highly concentrated market; and iv) the lack of AI-skilled labor forces.

The fast expansion of data centers is already putting grids under pressure. The most notable example is in Ireland, where the state-owned electricity operator imposed a moratorium on centers in Dublin in early 2022 and set out conditions to connect new centers to the grid, including a preference for those generating their own electricity. The operator has "comprehensive" plans to build out the grid, but restrictions on large new connections are expected to continue probably until 2028, or thereabouts. Grid queues are growing around the globe. In West London, historically considered a hub for data centers, new facilities must wait until 2030 to connect to the grid.[36] In the south of Sweden, a strong market for renewables, there is so much demand to get connected that businesses may have to wait years. Demand is so high that large tech companies are having bidding wars over data center sites with ready access to power. In the US, when a member organization of the northeastern US power grid submitted a request to the Energy Regulatory Authority (FERC) to increase the power capacity of an Amazon data center from 300 to 480 megawatts, the FERC rejected the request, arguing that "it could have major consequences for grid reliability." Moreover, considering the potential competition with renewables for grid connections and improvements, bottlenecks in grid deployment and reinforcement are more likely to occur. This is even more likely when it is recalled that a data center takes only 1-2 years to build, compared to grid lead times which are far longer (4 years for a large power transformer; 8-10 years for an underground or overhead transmission line).[37] The conclusions about waiting lines must be tempered to some degree, as data centers usually apply for several locations and grid connections, pushing up grid connection delays everywhere.

Thus, investments in grids are needed, way beyond what was previously necessary to adapt to the growing share of renewable electricity sources in the electricity mixes. In the US of course, but also in Europe. Even though its estimate may be excessive, Goldman Sachs considers Europe needs over €800 billion in spending on transmission and distribution in only the coming decade.[38] This is far above the €584 billion of the European Commission

36. Kao Data, cited in J. Saul *et al.*, "AI is already wreaking havoc on global power systems", *op. cit.*
37. A. Green *et al.*, "How Data Centers and the Energy Sector can Sate AI's Hunger for Power," *op. cit.*
38. "AI is Poised to Drive 160% Increase in Data Center Power Demand," *op. cit.*

roadmap to strengthen the grid.[39] Regarding the other investments needed, augmenting secure or dispatchable power generation capacity is becoming a key topic in the United States. North America's electricity grid faces "critical reliability challenges" as power generation fails to keep pace with surging demand from artificial intelligence.[40] Some States are already under pressure when looking at their peak load growth expectations, and half of the US grid is at risk of power shortfalls in the upcoming ten years, according to the North American Electric Reliability Corporation.[41] If we add the 69 data centers currently demanding access to the grid in Georgia to the State's current peak load (16.5 GW), the future peak load demand will double to reach 31 GW. This gap is astonishing, and the State may fail to create such huge additional power capacities, even with gas as a bridge. To recall, the current trend for the US is to add 6 GW of combined cycle gas turbine (CCGT) output every year, in all its States combined. Until 2030, 47 GW in new thermal power generation capacity (roughly $50bn) is required to support the expansion of data centers.[42] This figure was estimated before the announcement of a $500bn investment in Stargate, which will probably require around 25 GW of uninterrupted power to operate,[43] representing 5 years of additional CCGT in the US, equivalent to 20 AP1000 Westinghouse nuclear reactors. The US will likely thus face difficulties to cope with this additional power demand as nuclear power cannot be built out within a few years, even more so as less than 10 Westinghouse reactors have currently been commissioned in the world. The same Goldman Sachs analysis has reproduced its American trend estimations for Europe,[44] and considers AI will lead Europe to increase its power consumption by 50% between 2023 and 2033. If for the US the expansion seems reasonable, it appears to us unlikely regarding Europe.

The strained labor force is an additional inhibitor, particularly the emerging shortage of electrical trade workers essential to executing these projects. McKinsey estimates anticipate a potential shortage of up to 400,000 workers in the United States, based on the projected data center build-out and expansion of comparable assets requiring similar skills, such as semiconductor fabrication and battery gigafactories.

Moreover, AI is a very concentrated market, especially regarding its key technology component: chips and, more particularly, GPUs. Two figures are eloquent on the matter. While Taiwan Semiconductor has/accounts for more

39. "An EU Action Plan for Grids", Communication from the Commission (EU) 2023/757, November 11, 2023.
40. "2024 Long-Term Reliability Assessment", North American Electricity Reliability Corporation, 2024.
41. *Ibid.*
42. "AI is Poised to Drive 160% Increase in Data Center Power Demand," *op. cit.*
43. In March 2024, Sam Altman whispered an upcoming investment of $100 billion for a 5 GW data center called "Stargate". Now that we know Stargate will be a $500 billion investment, we have multiplied the previous figure by 5.
44. "AI is Poised to Drive 160% Increase in Data Center Power Demand," *op. cit.*

than 90% of the production market for AI-related chips,[45] in terms of sales, Nvidia holds about 88% of the GPU market.[46] Similarly, buyers of Nvidia's Hopper Generation technology are scarce, and Microsoft is leading the race. In 2024, the company bought around 500,000 H-chips, almost as much as Meta, Amazon and Google combined.[47] If data centers only take 1-2 years to build (even though this is increasing), a chip fabrication plant takes 3 to 5 years to set up, meaning the already concentrated market will likely face shortages compared to the global demand in the next years, if production does not become more diverse or ramp up. However, there is a growing number of chipmakers entering the market and specializing in one side (training or inference) of AI technology. But, when we know that a single 5 GW Stargate data center will need 2 million GPUs to work,[48] it is not very likely that the new market entrants will fill the gap between the monumental demand and the constrained offer.

In this context, and in contrast to expectations, Donald Trump's return to the White House and the announced trade war between the US, the EU and China is likely to hamper AI's development and deployment worldwide. The competition for chips has been taking place since the end of 2022. Back then, the US implemented a chip export ban to China regarding Nvidia's technologies. With this, state-of-the-art GPUs like Nvidia's A100 and H100 can no longer be sold to Chinese companies. Yet, China had anticipated this move and smartly stockpiled chips before the US instituted export controls.[49] Moreover, China has three software developers for every two in the US, and there are currently no restrictions on US-based entities related to or owned by Chinese tech companies accessing high-end AI chips through data centers located in the US. Chinese companies are also using the chip void to master technologies, as well as approach and recruit key stakeholders and developers in the Silicon Valley.[50]

On top of the factors that could drive AI-related electricity demand up (or down), the data center boom must cope with two other energy-linked issues: price and the water supply.

The surge in data center demand, combined with the heavy investments previously mentioned, is increasing the likelihood that customers will see their energy prices go up. In the US, the FERC's above-mentioned decision not to give access to the grid was also justified by the possible "major

45. "World Energy Outlook 2024", *op. cit.*, p. 186-189.

46. R. Dow, "Shipments of Graphics Add-in Boards Decline in Q1 of 24 as the Market Experiences a Return to Seasonality," John Peddie Research, June 6, 2024, available at: https://www.jonpeddie.com.

47. G. Smith *et al.*, "FirstFT: Microsoft Buys Twice as Many Nvidia AI Chips as its Rivals in 2024," *Financial Times*, December 18, 2024, available at: https://www.ft.com.

48. "Infrastructure is Destiny, Economic Returns on US Investment in Democratic AI", OpenAI, September 2024.

49. S. Rai *et al.*, "China AI Startup Stockpiled 18 Months of Nvidia Chips Before Ban," *Bloomberg,* November 10, 2023, available at: https://www.bloomberg.com.

50. E. Olcott, "Chinese Tech Groups Build AI Teams in Silicon Valley," *Financial Times,* November 18, 2024, available at: https://www.ft.com.

consequences for consumer costs." To take the example of Nebraska, the costs of adding more electrical generation to support new facilities for Google and Meta are being passed onto residents, with an estimated rate increase of around 2.5% to 3% per year.[51] In Europe, wholesale power prices in Ireland have been a third higher on average this year than in the rest of the continent. There is no denying that these higher prices have been driven by the more than 20% of electricity consumed by data centers that are decoupling increasing demand with the slower growth of power supply capacity.

Cooling is also needed to prevent chips from malfunctioning, which has historically required consuming water. Progress in water use efficiency has been made. From the first generation of owned data centers in the early 2000s to the current generation in 2023, Microsoft has reduced their water intensity (water consumed per kilowatt-hour) by over 80%. However, data centers still require vast amounts of water to cool the storage devices. In the US, the water used by data centers is equivalent to that used by a modern city of 3 million people. The demand for water from the data center industry is already giving rise to serious concerns in some states like Virginia. Alternative approaches do exist, but they require about 5% more energy on average.[52]

51. D. Lee, "Sam Altman's Energy 'New Deal' Is Good for AI. What About Americans?," *Bloomberg*, October 17, 2024, available at: https://www.bloomberg.com.
52. "2024 Environmental Sustainability", Microsoft, 2024.

Dominant Future Trends

Big Tech and Big Oil: Caught between Ambition and the Energy Dilemma

Big Tech is leading the AI boom with products like ChatGPT (OpenAI/Microsoft), Gemini (Google), Alexa (Amazon) and Siri (Apple). All the firms have announced goals to run their data centers entirely on low-carbon energy and to set climate neutrality goals. Amazon and Microsoft aim at powering their data centers only with clean energy by this year, and Google by 2030. Amazon wants to be climate-neutral by 2040. Yet, for now, the path forward remains unclear, and AI is jeopardizing Big Tech's climate targets. Microsoft recently admitted that its AI push is compromising its long-held goal to be carbon-negative by 2030, as its carbon emissions jumped 30% last year. And the company does not seem to be doing all it can to get back on track, as evidenced by its president claiming that "the good AI will do for the world will outweigh its environmental impact."

Moreover, Big Tech's data centers are almost climate-optimal regarding efficiency. Google claims that the average annual power usage effectiveness (PUE) for its global fleet of data centers was 1.1 in 2023. Meta claims 1.08 for the same year. In 2023, Google's operational data centers, on average, exhibited a water usage effectiveness (WUE) of 0.18 L/kWh, while the average data center had a WUE of 1.8L per 1 kWh.[53] This raises concerns about Big Tech's margin of maneuver to keep reducing their carbon emissions in the context of superfast-growing demand for their services. In addition, if Big Tech's data center-related emissions are low thanks to renewable energy certificates, considering emissions from the in-house data centers of Google, Microsoft, Meta and Apple multiplies their GHG emissions by 7.62.[54]

Data centers are in competition with industry as well as the heating, cooling and transport sectors for the security of electricity supply, driving consumer prices up. The top ten corporate buyers of clean energy with power purchase agreements (PPAs) in 2023 included Amazon, Meta, Alphabet and Microsoft. Amazon was even the first buyer of solar and wind-generated electricity in the world. This tendency should continue in the coming years, as these companies need to secure electricity supply for the development of their activities. In the 2024-2025 period, the 86 Meta-supported wind and

53. "Data Center Water Cooling", Sunbird, available at: https://www.sunbirddcim.com.
54. I. O'Brien, "Data Center Emissions Probably 662% Higher than Big Tech Claims. Can it Keep Up the Ruse?," *The Guardian*, September 25, 2024.

solar projects located across 24 US states and 74 countries will add 9,800 MW of renewable energy to local systems.

On top of PPAs for RES, Big Tech is also looking into nuclear power, whether small modular reactors (SMRs) or traditional plants: Sam Altman, OpenAI chairman, also chairs the SMR company Oklo. Amazon, Microsoft and Google have all signed partnerships with nuclear operators to meet their growing electricity needs. Amazon is also to acquire a share in SMR start-up X-energy. For SMR developers, corporate PPAs with these giants are better than what the wholesale electricity market can give them. Regarding already existing nuclear plants, the most well-known move has been the reopening of the Three-Mile Island nuclear power plant for Microsoft, which sealed Constellation's biggest-ever PPA. Likewise, Amazon's cloud computing division agreed to spend $650 million to acquire a data center campus connected to Talen Energy Corp.'s 40-year-old nuclear power plant on the Susquehanna River in Pennsylvania.

Major developments in nuclear energy are more likely to materialize in the future (beyond ten years) than in the present. Yet, Big Tech is calling for a nuclear revival. Open AI, in a document titled "Infrastructure is Destiny", called for the construction of several 5-gigawatt power plants in various states across the US, each to support 5 GW data centers at a cost of around $100 billion.[55] A few months later, this project got a name: Stargate. With several 5 GW data centers, it gives Big Tech even more incentives to develop and acquire big nuclear power plants, as SMRs will not be enough for such titanic projects. In France, energy companies heard this call. As Macron announced on February 9, 2025 on the opening of the AI Action Summit a €109bn investment in AI in the upcoming years, national giant EDF launched a call for expressions of interest from digital companies to whom they would provide four ready-to-use industrial sites connected to the electrical grid and powered by 2 GW of available nuclear plants.[56]

In line with these trends, the US is already implementing policies to accommodate Big Tech, but with paradoxical results. Virginia, the state with the most data centers in the country, forfeited $750 million in tax revenue because of data-center-related incentives in 2023. In Illinois, $468 million in subsidies for the centers created only 339 positions— about $1.4 million per job).[57] This tax-cutting and fast-facilitating agenda is very likely to be reinforced under the new Trump administration. Indeed, on the day following his return to the White House, Donald Trump announced a $500bn investment over four years with an immediate commitment of $100bn for a joint venture called Stargate, uniting OpenAI, software leader

55. "Infrastructure is Destiny, Economic Returns on US Investment in Democratic AI," OpenAI, September 2024.
56. "EDF Supports Digital Companies in Developing New Data Centers in France," EDF press release, February 10, 2025, available at: https://www.edf.fr.
57. J. Whiton, "New Data on Data Center Subsidies, Same Old Problems", *Good Jobs First*, July 25, 2023, available at: https://goodjobsfirst.org.

Oracle and the Japanese investment bank SoftBank, to bolster US dominance in the AI. The first data centers are already under construction in Texas.

All these elements may lead to the unavoidable role of Big Tech in the future of electricity systems, especially if the flexibility issue promised by them is resolved through R&D. If data centers prove to be flexible, with record PPAs in renewables, nuclear power, SMR start-ups –in addition of a large fleet of bidirectional charging electric vehicles–, Big Tech firms may become major stakeholders in managing demand responses, and would play an indispensable role in electricity balances and grid-linked issues and policies.

In early November 2024, Abu Dhabi welcomed an Energy-AI summit where Big Oil recognized that "the rapid rise of artificial intelligence gives the world's largest oil companies a major incentive to increase their investments in renewable energies".[58] It is even said that Shell and BP, which had pulled back from renewables in the past two years to refocus on their core oil and gas businesses, are thinking of exploring renewables again.

Big Oil also welcomes the huge growth of data center electricity demand, that is driving a revival for fossil-fueled power plants, either gas or with CCUS. AI is set to provide a long tailwind for natural gas demand, which is the most likely fuel to fill any gap left by renewable sources, as coal continues to be phased out in the US.[59] Indeed, electricity demand from data centers is outpacing the deployment of renewables in many places. In some countries, including Saudi Arabia, Ireland and Malaysia, the energy required to run all the data centers they plan to build at full capacity exceeds the available supply of renewable energy. Most of Malaysia's data center capacity is not in use yet, but factoring in everything under construction, the amount of electricity used just by data centers will widely exceed the country's total renewable output in 2022, the latest year for which data is available.[60]

The Future of Data Centers

Data centers will pursue their expansion and increase their electricity consumption. Furthermore, they will keep being more and more concentrated, despite the rollout of edge data centers. There is no reason for the US to lose their dominance (or dependence) on data centers. Thanks to the fast construction permitting, tax cuts and deregulation promised by President Trump and his Department of Government Efficiency, the current trend is unlikely to be reversed: 1/3 of data centers were in the US in 2020, rising to 1/2 in 2024. The market will do its job and data centers will be more and more concentrated in the top 15 states, especially in Texas and Virginia

58. M. Moore, "'AI Huge Power Gives Oil Major Incentives to Invest in Renewables,' says Adnoc Boss", *Financial Times*, November 4, 2024, available at: https://www.ft.com.
59. "AI revolution: Meeting massive AI infrastructure demands", Barclays Research, January 16, 2025.
60. J. Saul *et al.*, "AI is Already Wreaking Havoc on GLobal Power Systems," *op. cit.*

where data center electricity consumption in 2030 could top 120 TWh per year in each state (equivalent to the current consumption of all US data centers), based on the EPRI's highest growth scenario. In Virginia, data centers could account for half of the State's electricity consumption.

In Europe, data center power demand will rise in two kinds of countries.[61] The first sort are countries with cheap and abundant power from nuclear, hydro, wind, or solar sources (the Nordic nations, Spain and France). The second kind will include countries with large financial services and tech companies, which offer tax breaks or other incentives to attract data centers (Germany, the UK, and Ireland). Indeed, in the end, regarding AI, it will be the consumers who decide where data centers will be, namely where industry and finance are. And proximity to users is important in very competitive and modern sectors, as even delays of milliseconds can have long-term productivity costs.

Edge data centers will grow steadily by around 20% in the next five years. They will mainly serve for smart-city applications and trading optimization. However, their deployment will be too slow to balance the unreliable constraints on grids caused by hyperscale data centers. Indeed, the rollout of 5G reduces the trade-off between distance and latency, so that non-specialized AI and data centers can now be located anywhere across the world, encouraging both colocation and hyperscalers for the economies of scale they provide. Data centers will keep getting bigger. As enormous data centers nearing 100 MW capacity are already in service, Meta announced in January 2025 that it was planning to build a 2 GW data center which would almost be as large as central Paris, and require at least two nuclear reactors to be powered at full capacity. This trend is even outpacing the already hard-to-follow 0.5 order of magnitude/year trend growth in AI training computing.[62] If this trend continues, we could witness a 10 GW data center working in 2028, which by itself would need the equivalent of average Portugal's power consumption. Given that the first Stargate 5 GW data center should also run in 2028, data centers are on track to continue their trend-based exponential growth.

However, the constraints on the grids may lead to new forms of policies and to a more open use of AI. Chris Lehane, OpenAI's head of policy, opened up the idea that in return for incentives and regulatory ease for grid connection, "it seems to me very reasonable that you should be asking for 20% or 25% of the compute to be available for your public education system, for your public universities—to stand up an AI hub that relates to industries that you already have in your state."[63] We shall see in a couple of years if this proposal is translated into reality.

61. "AI is Poised to Drive 160% Increase in Data Center Power Demand," *op. cit.*

62. L. Aschenbrenner, "Situational Awareness: The Decade Ahead", June 2024.

63. D. Lee, "Sam Altman's Energy 'New Deal' Is Good for AI. What About Americans?", *Bloomberg*, October 17, 2024.

One main topic for 2025 regarding data centers will be frugal AI and *sobriété*. Frugal and specific AI deployment are key features to mitigate the boom in electricity demand from AI-related data centers. By frugal AI, we refer less parameters in the models (through *pruning* notably: i.e., the removal of irrelevant or redundant data to enhance efficiency) and less training, that can be achieved mostly using pre-training models instead of models trained from scratch, delivering the same output with far less training and thus wasted energy.

On *sobriété* –"sober" or energy-efficient activities–, France hosted the AI Action Summit on February 10-11, 2025 (the third international summit on AI, after the AI Safety Summit at Bletchley Park (the UK) in November 2023 and another summit in Seoul in May 2024). Amongst the three main topics addressed were the sustainability of AI, and a political initiative to respond to the greater awareness about the impacts of AI on the environment. France's Ministry of Ecological Transition alongside the Presidency have sought to tackle four aspects of AI's sustainability: frugal AI, specific AI, environmental assessment, recycling and the circular economy.

If the US has an absolute lead in data centers' infrastructure, the Chinese start-up DeepSeek launched its LLM in late January 2025, which has completely shaken the AI world. With the promise of being as "smart" as ChatGPT, DeepSeek has reshuffled the cards of the great geopolitical AI game. Not only does it prove that China can and will compete with the US for AI supremacy (which has led some people to call this a "Sputnik moment"), but it also indicates that quality can be achieved through *sobriété* both in financing and conception. DeepSeek claims to have used only 2,000 chips to train the third version of its model, way less than the 16,000 ones on which ChatGPT was trained, although this figure is questionable and has met with skepticism in the AI world. The DeepSeek model apparently only requires a small amount of energy to run, and its pre-training phase only cost $6mn. Though, this figure –which was largely spread by mass media– only accounts for a small share of the total cost of development. It is estimated the total capital expenditures of the company was $1.6bn with over $500mn in hardware expenses (Hopper and other GPUs chips).[64] Likewise, the declining costs of AI inferences follow the trend of a fourfold to tenfold algorithmic improvement per year, and will continue to do so. Since the release of ChatGPT, costs per computational instance have fallen 1,200 times. The upcoming and continuing AI marvels and mysteries are yet to be unveiled.

64. D. Patel *et al.*, "DeepSeek Debates: Chinese Leadership On Cost, True Training Cost, Closed Model Margin Impacts," *SemiAnalysis*, January 31, 2025, available at: https://semianalysis.com.

Conclusion

The AI Action Summit was marked by a flurry of announcements regarding investments in artificial intelligence, with €109 billion committed in France and €200 billion across the European Union. These investments will notably support the construction of data centers on European soil. However, according to a McKinsey report, the energy demand of data centers in the EU and the UK combined could reach up to 150 TWh by 2030—equivalent to 1/3 of France's total electricity consumption in 2024. As a result, the summit also served as a key platform to discuss the major environmental and energy challenges tied to the rise of these technologies.

While this issue was not explicitly one of the summit's five main themes, it surfaced as an underlying concern throughout most discussions held in Paris on February 10-11, 2025. Promoting the development of sustainable AI emerged as one of the three key priorities identified for future action. The summit also aimed to trigger a global push toward making AI more environmentally sustainable. Yet, the massive energy demands of large-scale AI deployment remain a fundamental challenge, one that prompted several major announcements.

First, France, the United Nations Environment Programme (UNEP), and the International Telecommunication Union launched a coalition for environmentally sustainable AI. This coalition will support and amplify initiatives to reduce AI's energy and environmental footprint, ensuring that its growth aligns with broader energy and climate goals. Additionally, the International Energy Agency introduced the world's first Global Observatory on AI and Energy, designed to better anticipate the energy needs of AI models and data centers, optimize energy systems, and leverage AI for carbon emission reductions and breakthrough innovations in energy applications. Finally, UNEP will publish, in March 2025, a set of guidelines to steer public and private investments toward energy-efficient data centers.

This research aimed to shed light on this debate by presenting credible forecasts on AI-related energy demand and the data centers it requires. It highlighted the low likelihood that ongoing efficiency gains will fully offset the surge in electricity consumption driven by expanding data center usage, due to the rebound effect. AI will trigger higher electricity demand but is unlikely to cause an unstoppable exponential electricity consumption, as its growth will likely be constrained by several factors: grid constraints and queues; additional power generation capacities; the trade war in a highly concentrated chip market; and its adoption rate—shaped by workforce training timelines and the specific ways AI is integrated into industries.

Ultimately, the question remains whether AI will become the next driving force of industry. Its future impact on the energy system will largely depend on the pace at which businesses and workers adopt this technology.

french
institute
of international
relations
since
1979

27 rue de la Procession 75740 Paris Cedex 15 – France

Ifri.org