

Energy and Water Conservation in AI Data Centers

Z. Qian

ABSTRACT

As artificial intelligence applications become increasingly widespread, data centers' electricity and water consumption is drawing growing attention. This article analyzes 30 scientific and institutional reports to explore how these facilities can meet ever-increasing computing demands while conserving resources. The article compares published estimates with survey findings and uses a simple calculation to illustrate the impact of a lower power usage effectiveness (PUE) ratio. In 2023, electricity consumption at U.S. data centers reached 176 terawatt-hours (TWh); a 2024 national report projects that this figure will reach 325–580 TWh by 2028. Assuming IT power demand remains constant, reducing the PUE from 1.54 to 1.10 would decrease a facility's total electricity consumption by 28.6%. Using a PUE of 1.40 as a baseline, total electricity consumption would decrease by 21.4% while achieving the same goal. The literature advocates improving both computational and facility operational efficiency; however, cooling choices must account for local water supply and the electricity source. Low-carbon electricity can reduce emissions, but it does not reduce electricity demand. The study concludes that resource conservation requires efficient data processing, appropriate cooling systems, and location-specific management of electricity and water resources. Beyond efficiency metrics, absolute consumption data are important for evaluating progress.

KEYWORDS

artificial intelligence, data centers, energy efficiency, water conservation, PUE, sustainable data processing

HOW TO CITE THIS ARTICLE

Z. Qian (2026). Energy and Water Conservation in AI Data Centers. *The Journal of Young Innovators*, 2(2), pp. 52–75.

DOI <https://doi.org/10.67419/jyi.v2i2.18>

READ ONLINE <https://young-innovator.org/issues/articles/ai-data-centers>

Abstract

As artificial intelligence applications become increasingly widespread, data centers' electricity and water consumption is drawing growing attention. This article analyzes 30 scientific and institutional reports to explore how these facilities can meet ever-increasing computing demands while conserving resources. The article compares published estimates with survey findings and uses a simple calculation to illustrate the impact of a lower power usage effectiveness (PUE) ratio. In 2023, electricity consumption at U.S. data centers reached 176 terawatt-hours (TWh); a 2024 national report projects that this figure will reach 325–580 TWh by 2028. Assuming IT power demand remains constant, reducing the PUE from 1.54 to 1.10 would decrease a facility's total electricity consumption by 28.6%. Using a PUE of 1.40 as a baseline, total electricity consumption would decrease by 21.4% while achieving the same goal. The literature advocates improving both computational and facility operational efficiency; however, cooling choices must account for local water supply and the electricity source. Low-carbon electricity can reduce emissions, but it does not reduce electricity demand. The study concludes that resource conservation requires efficient data processing, appropriate cooling systems, and location-specific management of electricity and water resources. Beyond efficiency metrics, absolute consumption data are important for evaluating progress.

Keywords: artificial intelligence, data centers, energy efficiency, water conservation, PUE, sustainable data processing

Introduction

With the rapid development of artificial intelligence and growing reliance on data centers, electricity and water consumption at data centers are also rising rapidly, as this type of infrastructure requires a continuous supply of electricity and cooling. The International Energy Agency (IEA, 2025) estimates that global data center electricity consumption in 2024 was about 415 terawatt-hours (TWh), accounting for about 1.5% of global electricity consumption. In its 2026 update, the agency estimated 2025 consumption at 485 TWh and projected it to reach approximately 950 TWh by 2030 (IEA, 2026). However, these figures include data centers supporting all types of digital activities, not just artificial intelligence. Although AI is a major driver of growth, attributing the sector's entire electricity consumption to AI would overestimate its share of total consumption.

Data centers' water consumption comes not only from on-site cooling but also from the water power plants use to generate electricity. Some water-efficient cooling methods require more electricity, which may increase water consumption during power generation. Therefore, when selecting a cooling system, consider not only the water saved by the data center itself but also the resulting changes in electricity use and associated water consumption to determine whether overall water savings have been achieved (Macknick et al., 2012; Mytton, 2021).

This raises the question of whether artificial intelligence development can advance while reducing energy and water consumption. Based on a literature review, analysis, and comparison, this paper focuses on three areas: computational efficiency, cooling and site selection, and more environmentally friendly power generation methods. We chose these three areas because they affect electricity consumption, water use, and carbon emissions,

respectively. Finally, this paper uses a simple PUE calculation to show how the initial PUE value affects potential energy-saving benefits.

Background

Electricity Use and Facility Efficiency

Data centers house servers, storage devices, networking equipment, and the systems that keep them running. During training, an AI model is developed; during inference, the trained model generates results. Both processes consume electricity. Cooling systems, power conversion, and other auxiliary systems add electricity demand on top of the IT load. Because these components serve different functions, reducing cooling load and lowering energy consumption per computational task are two distinct areas of focus (U.S. Department of Energy [DOE], 2024a).

Power usage effectiveness measures total facility electricity relative to IT electricity over the same period and boundary:
$$\text{PUE} = \frac{\text{Total facility electricity}}{\text{IT equipment electricity}}$$

A PUE value of 1.54 means that for every 1 unit of IT power consumed, an additional 0.54 units are consumed to support the operation of the equipment. The theoretical minimum is 1.00. A lower value indicates less overhead but does not indicate whether the IT equipment is performing useful work efficiently. As the computational load increases, a facility may improve its PUE value while simultaneously increasing total power consumption.

Water Accounting and Environmental Boundaries

Water withdrawal refers to the extraction of water from a water source. Water consumption refers to the portion of that water that does not return to the immediate water environment, for example through evaporation. Direct consumption occurs within the facility;

indirect consumption occurs within the electricity supply chain. When comparing water estimates, this paper distinguishes both from “water withdrawal” (Li et al., 2025; Mytton, 2021).

$$\text{Site WUE} = \frac{\text{Annual on-site water consumption (L)}}{\text{Annual IT electricity (kWh)}}$$

This paper uses a consumption-based definition of site water usage effectiveness (WUE): the annual on-site water consumption divided by annual IT electricity consumption, expressed in liters per kilowatt-hour (L/kWh). A source-inclusive version of this metric also includes electricity-related water consumption. Terms such as “water use” may refer to different quantities; therefore, their accounting boundaries should be checked before comparing values (Mytton, 2021; Shehabi et al., 2024).

Electricity consumption, emissions, and water-use impact are also distinct outcomes. A transition in electricity supply can reduce emissions while electricity consumption remains constant. Wind and solar photovoltaic facilities typically use little water during operation, whereas certain other low-carbon technologies require large amounts of cooling water or involve evaporation from reservoirs. Therefore, relevant comparisons depend on specific power generation and cooling technologies (Macknick et al., 2012)

Methods

Evidence Selection and Scope

This review used online literature search tools such as Google Scholar to identify relevant literature, then searched publisher and institutional websites and research databases. The initial search used four sets of keywords: “data center electricity consumption,” “AI water footprint,” “PUE survey 2025,” and “energy-efficient data center design.” The study initially selected 12 sources and then conducted supplementary searches on five themes: data center electricity

consumption estimates, computational efficiency, cooling systems, water consumption, and environmental reporting. Together, these thematic searches and targeted searches based on titles cited in the existing sources yielded 18 new sources. We selected sources that provided relevant data, research methods, or technical explanations for the aforementioned themes, prioritizing academic papers and original institutional reports.

A total of 30 sources were ultimately selected, including journal articles, conference papers, and institutional reports. To ensure the data were timely, the search was limited to the past 20 years, and the included sources were published between 2011 and 2026. The review is selective rather than systematic; it compares findings by theme and does not pool effect sizes across studies. Key figures were checked against the IEA executive summary, the relevant pages and figures of the LBNL report, and the Uptime survey report, including values, years, units, and accounting boundaries. During verification, the review distinguished between historical estimates and future projections, PUE values under different weighting methods, and water withdrawal versus water consumption. This verification focused on sources that provided key quantitative data; for supplementary literature, we reviewed only publicly available abstracts, authors' versions, or publication information. This review included PUE calculations to show how reducing facility overhead affects total electricity consumption when IT electricity consumption remains constant, and to explain why different starting PUE values yield different energy-saving ratios.

Table 1

Main Themes in the Selected Literature

Evidence group	Representative sources	Contribution to this review
Electricity demand and forecasts	IEA (2025, 2026); Shehabi et al. (2024); Koot and Wijnhoven (2021)	Demand estimates, scenario assumptions, and regional differences.
Computing and facility efficiency	Masanet et al. (2011, 2020); Sze et al. (2017); Uptime Institute (2025)	Distinguishes useful IT work from facility overhead.
AI training and deployment	Strubell et al. (2019); Patterson et al. (2021); Luccioni et al. (2023, 2024)	Connects model choice and life-cycle boundaries to resource requirements.
Water accounting and location	Mytton (2021); Ristic et al. (2015); Siddik et al. (2021); Li et al. (2025)	Separates withdrawal, consumption, and local water stress.
Cooling and electricity supply	Ebrahimi et al. (2014); Macknick et al. (2012); DOE (2024a)	Explains energy and water tradeoffs in engineering choices.
Measurement and reporting	Henderson et al. (2020); Dodge et al. (2022); Mytton and Ashtine (2022)	Supports consistent boundaries and clearer interpretation of estimates.

Calculations and Interpretation

Electricity consumption is expressed in TWh, where 1 TWh is equivalent to 1 billion kWh. Percentage changes are calculated based on published final data. Historical estimates, survey results, and forecast data are presented separately. In the charts, the lines connect reported observed values; no estimates are provided for missing years. In PUE calculations, IT electricity consumption is held constant, and only the facility's overhead is varied.

In energy efficiency scenarios, total electricity consumption is IT electricity consumption multiplied by PUE. With IT electricity consumption held constant, the following relative energy

savings can be derived: Electricity saving fraction = $1 - \frac{\text{Target PUE}}{\text{Starting PUE}}$

The base scenario starts with a PUE value of 1.54 and examines target values ranging from 1.40 to 1.10. The second set of calculations starts with a PUE value of 1.40, which roughly corresponds to the 2023 U.S. average reported by Shehabi et al. (2024, pp. 47–48). These scenarios are intended to illustrate the impact of the baseline value and do not estimate the energy savings achievable across the total national data center inventory.

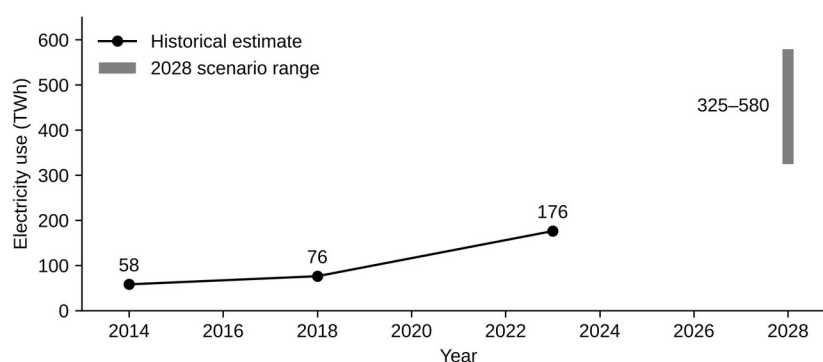
Results

Growth in Electricity Demand

Figure 1 shows estimated electricity consumption by data centers in the United States: 58 TWh in 2014, 76 TWh in 2018, and 176 TWh in 2023. The 2014 data comes from the U.S. Department of Energy (DOE) announcement for this report (DOE, 2024b); the remaining data, as well as the 2028 scenario range, come from Shehabi et al. (2024, pp. 52–53). Between 2018 and 2023, electricity consumption is estimated to have increased by approximately 131.6%. The forecast range for 2028 is 325–580 TWh, which is 84.7%–229.5% higher than the 2023 estimate.

Figure 1

U.S. Data-Center Electricity Use and the 2028 Scenario Range

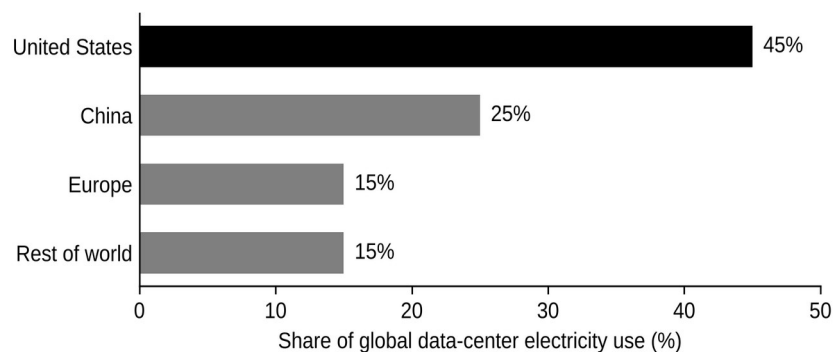


Note. Data from Shehabi et al. (2024, Figure 5.5); the 2014 value of 58 TWh is specified in DOE (2024b). The vertical bar shows the 2024 report’s range of scenarios for 2028.

According to the IEA's 2025 report, the United States accounted for about 45% of global data center electricity demand in 2024, China 25%, and Europe 15%. Figure 2 attributes the remaining 15% to the rest of the world. This concentrated distribution makes site selection critical: even a relatively small global share can place enormous strain on individual power supply systems. The IEA's 2026 outlook projects about 950 TWh for 2030, largely in line with its previous projection of 945 TWh (IEA, 2025, 2026).

Figure 2

Regional Shares of Global Data-Center Electricity Use in 2024



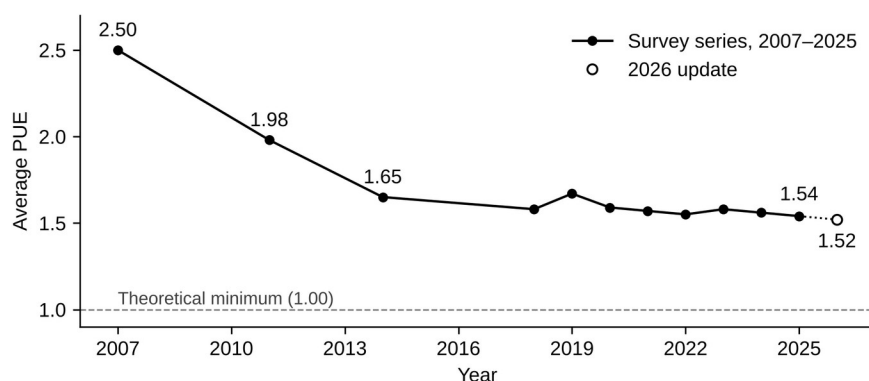
Note. Data from IEA (2025), executive summary. Shares are approximate. The rest of the world is calculated as 100% minus the three reported regional shares.

Facility Efficiency and the Limits of PUE

A series of Uptime Institute surveys shows that the average PUE declined from 2.50 in 2007 to 1.65 in 2014, then changed less and fluctuated year to year. The reported value for 2025 was 1.54, while the updated value for 2026 was 1.52 (Figure 3). The significant decline in the early years suggests substantial facility improvements, while recent data indicate that the pace of improvement in this metric has slowed (Uptime Institute, 2025).

Figure 3

Reported PUE in Uptime Institute Surveys



Note. Data from Uptime Institute (2025, Figure 2), with the 2026 headline value from Davis (2026). The survey samples vary across years.

Davis (2026) reports a PUE of 1.52 and an alternative, capacity-weighted value of 1.36. Since larger facilities are typically more efficient and carry greater weight in this calculation, the capacity-weighted average is lower. This discrepancy illustrates why a consistent weighting method should be used when comparing average PUE values.

The mathematical structure of PUE also introduces another limitation. If server power consumption is reduced through IT optimization while some fixed overhead remains constant, the PUE value may still rise despite a decrease in total power consumption. Conversely, any additional computational load, whether useful or not, can improve the PUE value by spreading fixed overhead over a larger denominator. Therefore, when conducting environmental assessments, absolute power consumption and workload-based performance metrics must be considered in addition to PUE.

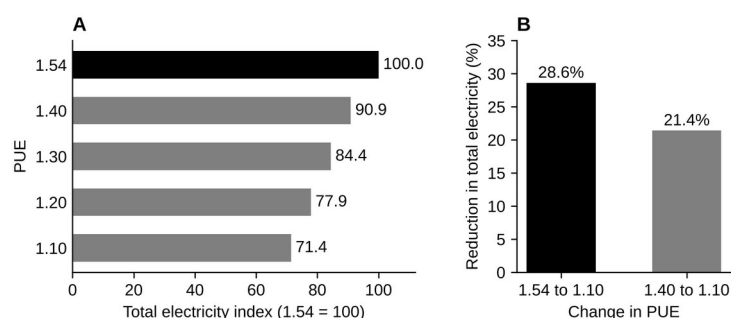
Illustrative Savings and Baseline Sensitivity

Using an initial PUE of 1.54, target values of 1.40, 1.30, 1.20, and 1.10 imply reductions in total power consumption of 9.1%, 15.6%, 22.1%, and 28.6%, respectively, assuming IT power

consumption remains constant. Figure 4A illustrates these results in the form of an index of total facility power consumption.

Figure 4

Illustrative Electricity Savings From Lower PUE



Note. Calculated as $1 - \text{target PUE} / \text{starting PUE}$, holding IT electricity constant. Panel A starts at 1.54; Panel B compares two baselines with a target of 1.10.

If the initial equipment is already relatively efficient, the energy savings achieved by pursuing the same goal will be reduced. Improving efficiency from 1.40 to 1.10 reduces total electricity consumption by 21.4%, whereas using 1.54 as the baseline yields 28.6% savings. The difference between the two is 7.1 percentage points (calculated using unrounded values). Therefore, select a realistic baseline value when evaluating a planned modernization project.

Growth in demand can also offset the gains in energy efficiency. If IT electricity consumption doubles and the PUE decreases from 1.54 to 1.10, total electricity consumption will be $2 \times 1.10 \div 1.54$, or 1.43 times the initial level. Despite improvements in equipment efficiency, electricity consumption would still increase by approximately 42.9%. This hypothetical example illustrates why reducing resource consumption intensity and reducing absolute consumption are two distinct objectives.

Direct and Indirect Water Consumption

Shehabi et al. (2024, pp. 55–57) estimate that direct water consumption by U.S. data centers in 2023 was approximately 66 billion liters, while indirect water consumption resulting from electricity generation amounted to nearly 800 billion liters. The latter figure was calculated based on regional electricity mixes and does not account for building-specific power purchase agreements or on-site energy generation. These figures show that both on-site cooling water consumption and electricity-related water consumption should be estimated. Indirect water consumption is nearly twelve times higher than direct consumption. So, if people consider only the water used to cool data centers, they overlook a large portion of water consumption. Water conservation is thus not just about cooling systems; electricity consumption and its source must also be considered.

Where water is consumed matters, not just the national total. In their study of the United States, Siddik et al. (2021) found that about one-fifth of the direct water footprint occurs in moderately to highly water-stressed river basins. A liter of water consumed in a water-stressed region does not necessarily have the same environmental impacts as a liter consumed in a region with relatively abundant water resources. This shows that site selection must consider not only electricity prices, land availability, and network connectivity, but also water availability. Even if a building is highly water-efficient, constructing it in a water-stressed region can exacerbate local pressure on the water supply.

Li et al. (2025) estimated that direct water consumption for the GPT-3 training simulation in the United States was approximately 700,000 liters. They projected that global water withdrawal driven by artificial intelligence (AI) would reach between 4.2 and 6.6 billion cubic meters by 2027; this figure measures withdrawal rather than consumption. Both figures rely on

model assumptions and cover different statistical scopes; therefore, they are not directly comparable.

Reporting on water consumption remains incomplete. The Uptime Institute (2025, Figure 6, pp. 11–12) reported that 47% of respondents had collected water data relevant to their company’s sustainability. In this figure, this metric is referred to as “water use,” while the accompanying text uses the term “water consumption.” This discrepancy points to reporting gaps but does not prove that all respondents use the same definition of consumption. This issue aligns with the disclosure concerns Mytton (2021) raises. The relatively low share of respondents collecting water data may reflect a greater focus on electricity costs for operations, inadequate water monitoring, and inconsistent reporting requirements across regions. However, these are only possible explanations; the percentage alone is not sufficient to determine the cause. At the very least, this suggests that a large proportion of operators do not systematically collect water data for sustainability reporting.

Discussion

The existing literature supports improving computational efficiency and facility operational efficiency simultaneously. Masanet et al. (2011) categorized potential savings into those related to IT equipment and those related to supporting infrastructure. Subsequent U.S. modeling studies found that improvements in efficiency, along with the transition to larger-scale facilities, are key reasons why service demand and energy consumption have not grown in tandem (Shehabi et al., 2018). These findings provide a basis for a comprehensive examination of multiple aspects of facilities.

For operators, a straightforward starting point is to measure both effective output and energy consumption. When making comparisons, consistency in task types, model quality, and response time requirements should be maintained as much as possible. Sze et al. (2017) elaborated on how hardware design and efficient data movement influence neural network energy consumption. Furthermore, Patterson et al. (2021) note that model selection, processors, and deployment location are key factors influencing training efficiency and carbon emissions.

The computational workload itself also warrants attention. Strubell et al. (2019) demonstrated that model development and training can incur significant resource costs. Schwartz et al. (2020) argued that model evaluation should consider both efficiency and accuracy. In terms of applications, Luccioni et al. (2024) found significant differences in energy consumption across tasks and between specialized and general-purpose models. This suggests that model selection should be based on task requirements rather than blindly assuming that the largest available model must be used.

As the PUE example in the Results showed, improved efficiency does not necessarily reduce total consumption; this indicates that efficiency improvements and demand control must be considered together. Selecting an appropriately sized model reduces the computational effort required to complete the same task, while demand control involves determining which tasks need to be run and how many times. If each computation consumes less power but the number of computations increases significantly as a result, total electricity consumption may still rise.

Decisions regarding cooling require a comprehensive assessment of both energy consumption and water consumption. The U.S. Department of Energy (DOE, 2024a) lists airflow management, appropriate temperature settings, and direct liquid cooling as technical options.

Liquid cooling can remove heat from equipment, but water consumption also depends on how heat is dissipated to the external environment. Dry cooling can reduce reliance on evaporative cooling, but its electricity demand may be higher in hot weather and varies by design and climate. Furthermore, findings from the European Code of Conduct indicate a downward trend in PUE for participating facilities, highlighting the importance of ongoing attention to operational practices (Avgerinou et al., 2017).

Waste heat recovery offers another avenue for utilizing excess heat from data centers. Ebrahimi et al. (2014) note that relatively low temperatures limit the selection of suitable recovery technologies. Therefore, projects need to identify compatible heat consumers and must evaluate the required additional equipment.

When conducting site selection assessments, compare electricity consumption, direct and indirect water consumption, local water stress, and seasonal cooling demand. Ristic et al. (2015) emphasized that the water footprint is not limited to a facility's direct water needs. Although reclaimed water can reduce reliance on potable water resources, evaporation still consumes it. Similarly, a low site WUE may be accompanied by a significant electricity-related water footprint (Mytton, 2021; Siddik et al., 2021).

Clean electricity can mitigate the impact of the electricity that must still be consumed. Macknick et al. (2012) explain why power generation and cooling technologies must be considered together when assessing water demand. Dodge et al. (2022) note that the location and timing of cloud computing deployments influence operational emissions. With this information, operators can schedule flexible workloads but must also consider water availability and service requirements (Li et al., 2025).

Consistent measurement data helps translate these options into actionable decisions. Henderson et al. (2020) proposed tools for tracking energy consumption and CO₂ emissions during machine learning processes. Lannelongue and colleagues (2021) provide a framework for estimating the carbon footprint of computational operations in the absence of detailed measurement data. Their accompanying practical guide recommends reducing unnecessary computational operations and adopting a holistic approach to hardware selection and site selection (Lannelongue et al., 2021).

Building on this recommendation, the first step may be to reduce unnecessary computations and select models sufficient for specific tasks (Luccioni et al., 2024). For teams that can determine model selection and task scheduling, this is typically easier than replacing cooling systems or relocating data centers, and they can make adjustments as part of daily operations. However, if operators merely provide computing facilities to clients, model selection requires client cooperation. The most challenging aspects are relocating facilities and changing power supplies, as these require funding, infrastructure, and local resources, making them difficult to accomplish in the short term.

Table 2

Comparative Assessment of Environmental Protection Strategies

Strategy	Primary outcome	Supporting evidence	Main condition or tradeoff
Computing and facility efficiency	Less electricity per comparable service	Historical models and engineering studies	Absolute demand can still grow; PUE excludes IT productivity.

Strategy	Primary outcome	Supporting evidence	Main condition or tradeoff
Water-conscious cooling and siting	Less direct consumption or local water stress	Water accounting and spatial assessments	Dry cooling, water reuse, and heat rejection have different energy and water consequences.
Cleaner electricity and flexible scheduling	Lower emissions; potentially less indirect water use	Power-sector and cloud-computing studies	Low-carbon power is not always low-water; service and grid constraints matter.

Note. Qualitative comparison based on Masanet et al. (2020), Patterson et al. (2021), Mytton (2021), Siddik et al. (2021), Macknick et al. (2012), Dodge et al. (2022), Li et al. (2025), and DOE (2024a).

Interpreting Changing Estimates

Masanet et al. (2020) estimated that, despite significantly faster growth in computing power, global data center electricity consumption increased by only about 6% between 2010 and 2018. Their key methodological critique is that directly extrapolating electricity consumption based on service demand may overlook improvements in energy efficiency. This warning remains relevant today. However, it does not prove that electricity demand will remain essentially constant in the face of major changes in computing hardware and workloads.

Forecasts depend both on the growth in usage and on the rate of improvement in energy efficiency. Koot and Wijnhoven (2021) demonstrated how changes in these assumptions affect projected electricity demand. As de Vries (2023) and the International Energy Agency (IEA, 2026) note, artificial intelligence development will create additional demand pressures. These studies reach a common conclusion: while past energy-efficiency improvements provide a useful reference, future demand must be reassessed as workloads and hardware evolve.

Differences among various estimates also reflect differences in their evidence bases. Mytton and Ashtine (2022) point out two common issues: limited access to underlying data and reliance on a small number of data providers. Therefore, comparing these projections requires examining the facilities covered, workload assumptions, and source data, rather than evaluating them solely on their results.

Limitations and Future Studies

First, this review is selective; these 30 sources do not constitute a comprehensive sample, and for some supplementary sources, we reviewed only abstracts or authors' versions. Second, some studies draw on the same underlying data, so their findings are not entirely independent. Furthermore, the quantitative analysis focuses primarily on electricity and water consumption during the operational phase, while the manufacturing and disposal stages of equipment also have significant impacts: Luccioni et al. (2023) note that estimates of the carbon footprint of large language models would change if other life-cycle stages were taken into account.

The PUE scenarios assume that IT power consumption remains constant and do not account for factors such as cost, reliability, weather conditions, or equipment replacement. These scenarios are intended to illustrate a relationship rather than predict energy savings achievable through equipment upgrades. National-level projections are scenario ranges, and overall trends cannot reveal the standalone causal contribution of artificial intelligence. Comparisons of water consumption are influenced by differences in definitions, geographic conditions, and methods for attributing water use to electricity generation. Therefore, future research should use standardized reporting periods and clearly defined denominators to enable comparisons across facilities. A useful dataset should link IT energy consumption and total electricity usage to workload metrics, direct water withdrawal and consumption, cooling configurations, local water stress, and power

generation characteristics. Distinguish between annual totals and seasonal peaks. Such data would let researchers investigate whether apparent improvements reduce overall environmental impact or merely shift it to other systems.

Conclusion

As the analysis shows, AI data centers can reduce resource consumption for service delivery by improving computational efficiency, lowering facility overhead, adopting water-efficient cooling solutions, optimizing site selection, and using clean energy. Each strategy yields different outcomes: energy efficiency affects electricity demand; the choice of cooling method affects direct and indirect water consumption; and the electricity supply affects emissions and water consumption per unit of electricity.

An illustrative analysis of PUE reveals the potential and limitations of simple energy-saving reasoning. Assuming IT demand remains constant, reducing PUE from 1.54 to 1.10 would decrease total electricity consumption by 28.6%. However, different baseline assumptions yield different results, and demand increases may offset these improvements. This is also the study's main conclusion: efficiency improvements do not necessarily reduce total energy consumption, as growing computing demand may offset the energy-saving effects. So data center operators and policymakers should focus on absolute electricity and water consumption and factor in water consumption during power generation in their assessments.

References

- Avgerinou, M., Bertoldi, P., & Castellazzi, L. (2017). Trends in data centre energy consumption under the European Code of Conduct for data centre energy efficiency. *Energies*, 10(10), Article 1470. <https://doi.org/10.3390/en10101470>
- Davis, J. (2026, August 6). *The growing PUE advantage of larger data centers*. Uptime Intelligence. <https://intelligence.uptimeinstitute.com/resource/growing-pue-advantage-larger-data-centers>
- de Vries, A. (2023). The growing energy footprint of artificial intelligence. *Joule*, 7(10), 2191–2194. <https://doi.org/10.1016/j.joule.2023.09.004>
- Dodge, J., Prewitt, T., Tachet des Combes, R., Odmark, E., Schwartz, R., Strubell, E., Luccioni, A. S., Smith, N. A., DeCario, N., & Buchanan, W. (2022). Measuring the carbon intensity of AI in cloud instances. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1877–1894). Association for Computing Machinery. <https://doi.org/10.1145/3531146.3533234>
- Ebrahimi, K., Jones, G. F., & Fleischer, A. S. (2014). A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities. *Renewable and Sustainable Energy Reviews*, 31, 622–638. <https://doi.org/10.1016/j.rser.2013.12.007>
- Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248), 1–43. <https://www.jmlr.org/papers/v21/20-312.html>
- International Energy Agency. (2025). *Energy and AI*. <https://www.iea.org/reports/energy-and-ai>

International Energy Agency. (2026). *Key questions on energy and AI*.

<https://www.iea.org/reports/key-questions-on-energy-and-ai>

Koot, M., & Wijnhoven, F. (2021). Usage impact on data center electricity needs: A system dynamic forecasting model. *Applied Energy*, 291, Article 116798.

<https://doi.org/10.1016/j.apenergy.2021.116798>

Lannelongue, L., Grealey, J., Bateman, A., & Inouye, M. (2021). Ten simple rules to make your computing more environmentally sustainable. *PLOS Computational Biology*, 17(9),

Article e1009324. <https://doi.org/10.1371/journal.pcbi.1009324>

Lannelongue, L., Grealey, J., & Inouye, M. (2021). Green algorithms: Quantifying the carbon footprint of computation. *Advanced Science*, 8(12), Article 2100707.

<https://doi.org/10.1002/advs.202100707>

Li, P., Yang, J., Islam, M. A., & Ren, S. (2025). Making AI less “thirsty.” *Communications of the ACM*, 68(7), 54–61. <https://doi.org/10.1145/3724499>

Luccioni, A. S., Viguier, S., & Ligozat, A.-L. (2023). Estimating the carbon footprint of BLOOM, a 176B parameter language model. *Journal of Machine Learning Research*, 24(253), 1–15. <https://www.jmlr.org/papers/v24/23-0069.html>

Luccioni, S., Jernite, Y., & Strubell, E. (2024). Power hungry processing: Watts driving the cost of AI deployment? In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 85–99). Association for Computing Machinery.

<https://doi.org/10.1145/3630106.3658542>

- Macknick, J., Newmark, R., Heath, G., & Hallett, K. C. (2012). Operational water consumption and withdrawal factors for electricity generating technologies: A review of existing literature. *Environmental Research Letters*, 7(4), Article 045802.
<https://doi.org/10.1088/1748-9326/7/4/045802>
- Masanet, E. R., Brown, R. E., Shehabi, A., Koomey, J. G., & Nordman, B. (2011). Estimating the energy use and efficiency potential of U.S. data centers. *Proceedings of the IEEE*, 99(8), 1440–1453. <https://doi.org/10.1109/JPROC.2011.2155610>
- Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, 367(6481), 984–986.
<https://doi.org/10.1126/science.aba3758>
- Mytton, D. (2021). Data centre water consumption. *npj Clean Water*, 4, Article 11.
<https://doi.org/10.1038/s41545-021-00101-w>
- Mytton, D., & Ashtine, M. (2022). Sources of data center energy estimates: A comprehensive review. *Joule*, 6(9), 2032–2056. <https://doi.org/10.1016/j.joule.2022.07.011>
- Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). *Carbon emissions and large neural network training* [Preprint]. arXiv.
<https://doi.org/10.48550/arXiv.2104.10350>
- Ristic, B., Madani, K., & Makuch, Z. (2015). The water footprint of data centers. *Sustainability*, 7(8), 11260–11284. <https://doi.org/10.3390/su70811260>
- Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63. <https://doi.org/10.1145/3381831>

- Shehabi, A., Smith, S. J., Hubbard, A., Newkirk, A., Lei, N., Siddik, M. A. B., Holecek, B., Koomey, J., Masanet, E., & Sartor, D. (2024). *2024 United States data center energy usage report* (LBNL-2001637). Lawrence Berkeley National Laboratory.
<https://doi.org/10.71468/P1WC7Q>
- Shehabi, A., Smith, S. J., Masanet, E., & Koomey, J. (2018). Data center growth in the United States: Decoupling the demand for services from electricity use. *Environmental Research Letters*, 13(12), Article 124030. <https://doi.org/10.1088/1748-9326/aaec9c>
- Siddik, M. A. B., Shehabi, A., & Marston, L. (2021). The environmental footprint of data centers in the United States. *Environmental Research Letters*, 16(6), Article 064017.
<https://doi.org/10.1088/1748-9326/abfba1>
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3645–3650). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1355>
- Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12), 2295–2329.
<https://doi.org/10.1109/JPROC.2017.2761740>
- Uptime Institute. (2025). *Uptime Institute global data center survey 2025*.
<https://datacenter.uptimeinstitute.com/rs/711-RIA-145/images/2025.Annual.Survey.Report.pdf>
- U.S. Department of Energy. (2024a). *Best practices guide for energy-efficient data center design* (Revised July 2024). <https://datacenters.lbl.gov/sites/default/files/2025-07/best-practice-guide-data-center-design.pdf>

U.S. Department of Energy. (2024b, December 20). *DOE releases new report evaluating increase in electricity demand from data centers*. <https://www.energy.gov/articles/doe-releases-new-report-evaluating-increase-electricity-demand-data-centers>