

# Resource Management Challenges in AI-Driven Data Centers: A Systematic Review of Energy, Water, and Material Constraints

Thelma Posadas-Paredes , Diana Karen Zavala-Vega, César Ramírez-Márquez \*  and José María Ponce-Ortega \* 

Chemical Engineering Department, Universidad Michoacana de San Nicolás de Hidalgo, Francisco J. Múgica S/N, Ciudad Universitaria, Morelia 58060, Mexico; 1804827k@umich.mx (T.P.-P.); 1919682x@umich.mx (D.K.Z.-V.)

\* Correspondence: cesar.ramirez@umich.mx (C.R.-M.); jose.ponce@umich.mx (J.M.P.-O.);

Tel.: +52-443-3223500 (ext. 1277) (J.M.P.-O.)

## Abstract

The rapid expansion of artificial intelligence has intensified the demand for high-performance data centers, leading to unprecedented pressures on energy, water, and material resources. This critical review examines the emerging challenges associated with resource management in AI-driven data centers by focusing on the interplay between computational growth and environmental constraints. The analysis integrates recent advances in energy efficiency, cooling technologies, and hardware design while highlighting the increasing water footprint of thermal management systems and the material implications linked to semiconductor manufacturing and infrastructure scaling. Particular attention is given not only to the trade-offs between performance optimization and sustainability but also to the limitations of current metrics used to assess resource efficiency. The review identifies key gaps in the literature, including the lack of integrated frameworks that simultaneously address energy, water, and material flows. Finally, this review provides insights into pathways for a more sustainable AI infrastructure by synthesizing present-day knowledge, critically evaluating existing strategies, and emphasizing the need for systemic approaches that align technological innovation with resource conservation and long-term environmental resilience.

**Keywords:** artificial intelligence; data centers; energy consumption; water use; critical materials; resource efficiency; sustainability

## 1. Introduction

In recent years, the rapid expansion of emerging technologies such as artificial intelligence (AI) and increasingly accessible digital services has driven the global growth of data centers [1]. The rapid expansion of AI is not only transforming social and economic dynamics but also increasing general reliance, causing significant environmental risks that have originated from its widespread use [2]. These infrastructures, or data centers, are the physical facilities that make up the digital cloud where much of the technology used by society on a daily basis currently operates [3]. Progressively, human dependence on AI (used via our electronic devices) continues to grow. This dependence is poised to produce dire consequences as it increases the consumption of energy, water, and materials whose environmental impacts are often overlooked in the face of innovation and progress [4].

The expansion of data centers is now drawing significant attention due to the rapid increase in resource consumption associated with them. This phenomenon is evident when one considers that between 2005 and 2010 the global energy consumption by data centers



Received: 22 July 2026

Revised: 27 August 2026

Accepted: 2 September 2026

Published: 4 September 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

increased by 56% [5]. In this regard, there are challenges associated with data center energy consumption that must be analyzed in order to meet this electricity demand without jeopardizing energy security or the reliability of the electrical grid. However, there are other concerns, beyond the significant electricity demand, associated with data centers. The digital cloud represents the essential infrastructure on which AI systems operate, and its physical infrastructure depends on thousands of devices, computers, routers, servers, storage systems, and other factors necessary for their operation, creating new environmental challenges that must be addressed [6].

The emphasis on AI in this context is important, as according to the International Energy Agency (IEA), the amount of energy consumed by a single AI search can be up to ten times greater than that of a standard internet search, given that generative AI search necessitates compute-intensive autoregressive neural network inference, whereas conventional keyword-based web search depends on efficient retrieval from a static inverted index [7]. Nevertheless, this consumption multiplier represents only a standard, as it is highly variable and mainly depends on the model parameter scale, hardware efficiency, and query complexity. In this sense, the capacity of data centers must be expanded to ensure that these tools function properly, which simultaneously impacts costs and other resources; for example, it is estimated that approximately 46% of data center operating costs are associated with energy consumption [8]. Ignoring these factors can lead to an overestimation of the contributions of AI in sustainable development and in the assessment of environmental impacts.

On the other hand, there is growing concern regarding the use of water in data centers as cooling systems are consuming enormous amounts to prevent their processors from overheating. Large-scale data centers can have power capacities reaching several hundreds of thousands of kilowatts, and the electricity required to operate these facilities may entail water consumption ranging from 0.18 to 2 L per kWh generated, depending on the generation technology [9]. If we break down these figures by city, the data centers in The Dalles, for example, consumed 1344 million liters of water in 2021, accounting for approximately 29% of the total water consumption in the city [9]. Excessive water use in data centers contributes to water stress, as data centers also consume water indirectly to generate electricity, underscoring the importance of water conservation goals [9]. Furthermore, since a data center requires an enormous number of materials to be built and operated, the rapid growth of AI has intensified the demand for high-performance hardware, such as servers and GPUs, whose manufacture not only demands critical materials but also generates significant waste [10]. While the debates over the impact of AI often focus on its energy consumption, this approach fails to account for the costs of the materials associated with its physical infrastructure; these include the extraction, processing, and manufacturing of materials as well as the volumes of waste and greenhouse gas emissions that these procedures create.

Large quantities of metals and minerals are required to produce servers, graphics processing units, storage systems, and other electronic components, with server replacement being a major contributor to raw material demand [11]. Planned obsolescence, whether caused by hardware failure or software updates, can make existing servers incompatible with newer technologies, increasing the need for hardware replacement [12]. Data center racks constitute a fundamental component of this infrastructure and accommodate servers, power distribution equipment, networking components, storage systems, and cooling-related equipment [13]. As AI workloads continue to increase computational requirements and hardware power densities, the material requirements associated with these components become increasingly relevant to the environmental assessment of AI-driven data centers. A single rack may contain materials such as antimony (Sb), barium (Ba), beryllium (Be), silicon (Si), indium (In), tantalum (Ta), bauxite, niobium (Nb), and gallium (Ga) [14,15].

Several of these materials are considered critical raw materials by the European Commission and have relatively low recycling rates [16], increasing concerns regarding resource depletion and waste generation at the end of their service life. It is estimated that between 200 and 250 million tons of electrical and electronic equipment waste with a low recycling rate is generated every year worldwide, and that this figure will continue to grow at a rate of 3 to 5% yearly [17].

At present, environmental sustainability is a strategic priority for major industries as they strive to reach economic, environmental, and social balance. Although AI is seen as a driving force (crucial for the economic growth of a nation), questions regarding its sustainability and its impact remain unanswered [18]. There remains a lack of public information about the resource usage, infrastructure, and operation details that enable the quantification of the resource demand of AI-driven data centers, which has led to a shortage of quantitative analysis with empirical substantiation that could be used as a basis to develop sustainable resource strategic management [19]. This situation is a cause for concern due to the tension these technologies create between economic growth objectives and environmental sustainability goals.

The development of green data centers has been closely associated with efforts to improve the environmental performance and resource efficiency of data center infrastructure. Early initiatives, including the work of The Green Grid and the U.S. Environmental Protection Agency, established a foundation for assessing and improving data center energy efficiency through metrics such as Power Usage Effectiveness (PUE) and Data Center Infrastructure Efficiency (DCiE) [20]. Subsequent approaches have expanded this perspective to incorporate additional environmental dimensions, including carbon emissions, water consumption, cooling technologies, renewable energy integration, and sustainable materials. In this context, Sánchez and Barrezueta [21] provide an applied example of green data center design in Ecuador, emphasizing cooling-system modernization, renewable energy integration, and the use of environmentally sustainable materials.

The literature on the impact of data centers continues to focus primarily on energy and water consumption. This analytical shortcoming limits the comprehensive understanding of the sustainability of digital infrastructure, given that the UN has developed and incorporated environmental improvements and greenhouse gas emissions as a key component within the sustainable development agenda [22]. Consequently, there is a persisting knowledge gap regarding the comprehensive assessment of data center sustainability, highlighting the need for research studies that must incorporate methodologies to more accurately quantify the environmental impacts associated with the growing expansion of AI and its supporting infrastructure. This study proposes to analyze the impact of AI from a broader perspective, considering not only the energy requirements of data centers but also the environmental implications that arise from water consumption, the use of critical materials, and the generation of waste associated with their infrastructure.

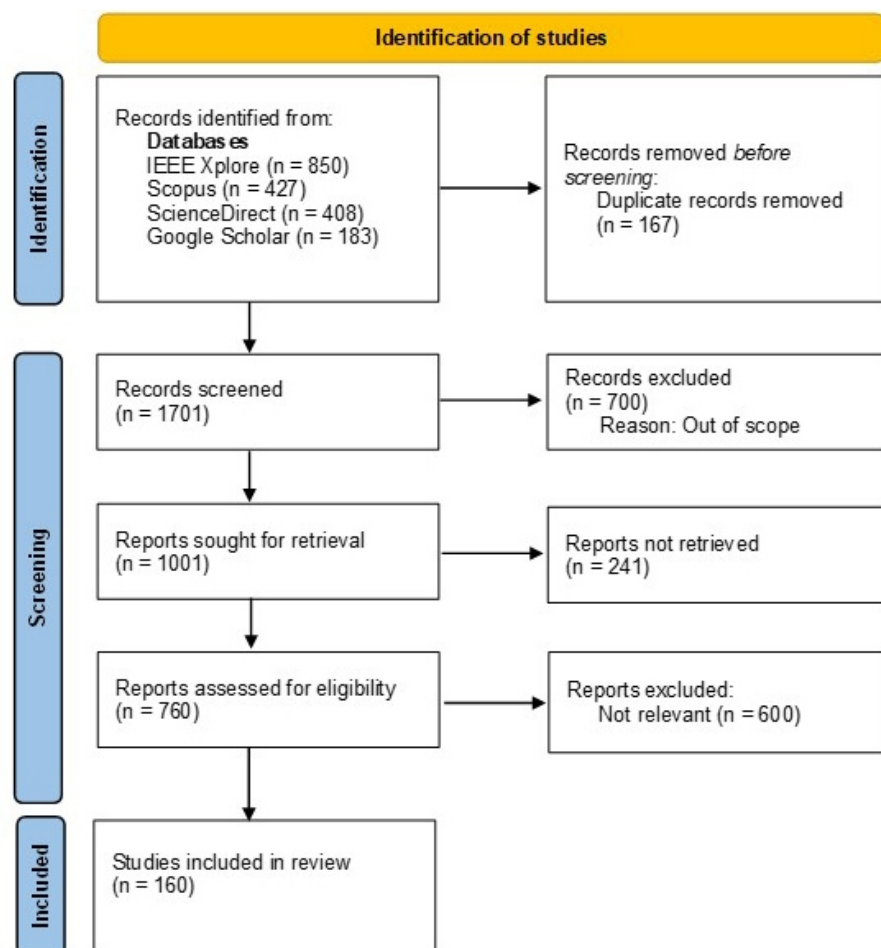
## 2. Methods

The literature search for this systematic review was carried out between April 2026 and July 2026. It includes studies published from 2015 to 2026 taken from four different databases: IEEE Xplore, Scopus, ScienceDirect, and Google Scholar. Additionally, the literature search was conducted in accordance with the PRISMA-S framework, a widely used framework in scientific research to ensure comprehensiveness in reporting, reproducibility, and transparency [23]. The completed PRISMA-S checklist is provided in the Supplementary Materials.

As per the PRISMA-S framework, in the first step, a total of 1868 studies were identified, considering only peer-reviewed studies in the English language published between 2015

and 2026, where 167 duplicates were removed. It is important to highlight that for this identification phase, specific topic-related keywords and Boolean operators were used. Relevant strings applied include: (“AI” OR “artificial intelligence”) AND (“challenges” OR “sustainability”) AND (“resource use” OR “resource use interactions” OR “resource management” OR “water use” OR “energy use” OR “material use” OR “minerals use” OR (“rare earth minerals use” OR “metals use” OR “semiconductors use”) AND (“renewable energies” OR “wind power” OR “solar power”) AND (“resource use indicators” OR “resource use metrics”) AND (“cooling systems” OR “hardware” OR “infrastructure”) AND (“data centers” OR “AI-driven data centers”).

After the identification phase, a title and abstract screening was conducted, where 700 additional studies out of the scope of this review were excluded. Furthermore, from the remaining 1001 studies, 241 studies were impossible to retrieve. Following this, we selected 160 studies to be included in this review, after the 760 remaining studies were subjected to a detailed full-text revision to confirm their eligibility. The PRISMA-S flow diagram (Figure 1) represents the literature selection process. The exclusion criteria applied can be summarized as follows:



*The report numbers are internally consistent:  $850 + 427 + 408 + 183 = 1868$  identified records. After removing 167 duplicate records, 1701 records remained for screening ( $1868 - 167 = 1701$ ). Following the removal of 700 out-of-scope records, 1001 records were left for retrieval ( $1701 - 700 = 1001$ ). Of these, 241 were unretrievable and 600 were not relevant, resulting in 160 records included in the review ( $1001 - 241 - 600 = 160$ ).*

**Figure 1.** PRISMA-S flow diagram.

- Studies published outside the analysis period (2015 to 2026).
- Studies not written in English.

- Unrelated field studies.
- Duplicate and incomplete studies.
- Non-peer-reviewed and unpublished studies.
- Irrelevant studies about resource management challenges in AI-driven data centers.

This systematic review was conducted and reported in accordance with the PRISMA-S guidelines. Based on the final article selection, this systematic review critically discusses the challenges related to resource management in AI-driven data centers. A detailed summary of the studies included in the review, including their research objectives, methodologies, and contributions, is provided in Table S2 of the Supplementary Materials. In this regard, this review article has been organized thematically into the following sections: Section 1: introduction, Section 2: methods, Section 3: energy use in AI-driven data centers, Section 4: water use and cooling systems, Section 5: material demand and supply constraints, Section 6: resource interactions and limitations of current approaches, Section 7: implications for resource management, Section 8: conclusions, and references.

### 3. Energy Use in AI-Driven Data Centers

#### 3.1. Growth in Energy Demand

According to MIT, AI, an energy-intensive technology, has been experiencing a rapid expansion since 2022 [24]. In 2023, powerful tech companies like Microsoft and Alphabet substantially expanded their AI capacity; this sudden growth coincided with the successful introduction of OpenAI's ChatGPT, which reached 100 million users in just 2 months [25]. Similar to Microsoft and Alphabet's expansion, in 2023 NVIDIA, the leading AI-driven chip production company, recorded an increasing demand for AI chips, reporting a record revenue of \$13.5 billion in just a 3-month period [26]. According to the International Energy Agency (IEA), in 2025, data centers account for around 2% to 3% of the global electricity demand [27,28], while projecting that the total global demand could surpass 20% by 2030 [13]. Berkeley Lab studies report that the electricity demand associated with the expansion of AI doubled between 2017 and 2023 [29]. In the U.S., it is noted, electricity consumption had shown an approximate annual growth rate of 7% from 2014 to 2018, but it increased to 18% from 2018 to 2023 [30]. The forecast rates show that this rate will again increase from 13% to 27% between 2023 and 2028 [29]. Research studies indicate that by 2027, NVIDIA could be shipping around 1.5 million AI server units per year, with a consumption of 6.5 to 10.2 kW per unit [29]. Other predictions point to a possible energy demand of  $1 \times 10^{15}$  kWh by 2030, a somewhat higher demand than the base case predicted by IEA [31]. In this regard, the AI-accelerated evolution is generating concerns about the electricity consumption of AI-driven data centers and their potentially related environmental impact. Nonetheless, it is important to highlight that these predictions are highly uncertain, and contrary to the predictions, the rise in AI electricity demand is constrained by the capacity of technology producers and investment capacity.

The IEA reports that the energy consumption of AI hardware manufacturing is lower than the energy consumption associated with AI operation, where 60% of the energy consumption is intended for wafer and semiconductor production while the remaining 40% is destined for auxiliary services [32]. Despite energy improvements in the chip manufacturing process, the complexity of the metal layers in chips raises the manufacturing electricity demand. However, the continuous advancement of chip computer power will reduce the energy demand for manufacturing due to fewer chips being required. Masanet et al. [33] identified that efficiency gains in data centers do not significantly increase energy consumption. Therefore, an increase in energy efficiency could reduce electricity demand, but several studies suggest that a rebound effect, which could lead to a higher electricity

demand, is feasible [26,31,34]. AI efficiency improvements by themselves cannot fully compensate for the increasingly long-term electricity demand of AI. Currently, 70% of chip production is located in East Asia, making this specific geographic area responsible for most of the global chip production demand [35]. A circumstance that could lead to severe resource depletion problems. By optimizing the supply chain of chip manufacturing, production areas, like those of East Asia, could prevent major resource depletion problems, something that otherwise would not be possible without the enhancement of AI resource use transparency.

On the other hand, the electricity consumption needed for AI operations recognizes two main phases: model training and inference. Model training refers to the process by which the model learns from data to generate a prediction, output, or action according to the patterns and relationships identified. Inference refers to the deployment of AI models previously pre-trained, which interpret new data to generate predictions or outcomes [36].

Even though the first studies about the energy required for AI operations stated that the training phase was the most energy-consuming phase [35], Alphabet and other technology companies currently report that inference consumes a greater amount of energy than training [34]. Nonetheless, GPT-3 in 2023 required an energy demand of  $1.287 \times 10^6$  kWh during its entire training phase, while inference only required  $5.64 \times 10^5$  kWh per day [37]. This, according to Argerich & Patino-Martinez [38] and Desislavov et al. [35], can be explained by the fact that even when training is more energy-intensive per execution in comparison with inference, the huge user inference demands over a model's lifetime have made energy demand dominated by the inference phase. In addition, the data suggests that the weight of electricity consumption for each phase depends on factors such as training frequency, the trade-off between model performance and energy consumption, and user inference demand [39]. Other studies have stated that the energy footprint of AI operations could be distributed as follows: (10%) model development and fine-tuning, (30%) model training, (60%) use of the model or inference [40].

Most of the AI sustainability research has focused on the training phase, while inference has received less attention. Verdecchia et al. [37] identified that out of 98 research papers written between 2015 and 2023, 49 focused on the study of AI training, while 17 focused on the inference phase. A possible reason for this might be that training commonly implies long-duration computation where associated power draws are predictable, while inference shows variable loads for a heterogeneous hardware mix, challenging to model but important due to its deployment scales. Additionally, the energy demand for the inference phase depends on several factors that include: the size of the input query and output answer, model size, input and output format (video, image, text, etc.), implementation of algorithmic efficiencies, and hardware characteristics [31].

### 3.2. Energy Efficiency Metrics

Given the increasing energy demand associated with AI-driven data centers, energy-efficient metrics are necessary to monitor energy use, making the assessment of the performance of data centers and the comparison of different technologies possible. Power Utilization Effectiveness (PUE) is the most widely used energy efficiency metric, even when other metrics have been introduced [41]. This metric is defined as the ratio between the total facility energy consumed and the energy consumed only by the information technology (IT), where a perfect efficiency is indicated by a value of 1.0. Values of PUE for typical data centers are between 1.3 and 2.5 [42,43]. These values depend on different factors that include: climate, operational practices, hardware, and software technology [44].

In general, energy efficiency metrics can be classified as follows depending on their relationship with the information technologies (ITs):

- IT-related metrics. These metrics evaluate the energy performance of networking and computing components by assessing the effectiveness of how IT uses energy to accomplish computational tasks.
- Non-IT-related metrics. By measuring the ratio of energy used by non-IT systems to the total energy consumption, these metrics evaluate the efficiency of supporting infrastructure such as cooling systems, power distribution, and building facilities.

Table 1 presents some IT-related and non-IT-related metrics; it is important to identify that each energy efficiency metric faces specific challenges and limitations [45,46]. Major limitations rely on the fact that data centers comprise multiple interconnected subsystems (workloads, power, cooling, environmental controls, etc.) that are influenced by diverse factors (humidity, temperature, hardware, etc.), so applying individual metrics fails to identify specific issues about energy efficiency [47]. Future metric developments should focus on defining techniques for globally uniform calculation metrics across geographic and regulatory boundaries, automating data collection, and predictive analysis.

**Table 1.** IT-related metrics.

| Energy-Efficient Metric                                                                       | Formulation                                     | Units                             | Optimal Value                   | Challenges                                                            | References |
|-----------------------------------------------------------------------------------------------|-------------------------------------------------|-----------------------------------|---------------------------------|-----------------------------------------------------------------------|------------|
| IT-related metrics                                                                            |                                                 |                                   |                                 |                                                                       |            |
| <i>Compute Power Efficiency (CPE).</i><br>Computes output per unit of IT power.               | $\frac{\text{Compute output}}{\text{IT Power}}$ | Computational Operations (Ops)/kW | $+\infty$                       | Difficulty in quantifying computing output across tasks.              | [47–49]    |
| <i>Performance per Watt (PpW).</i> Performance per unit of power.                             | $\frac{\text{Performance}}{\text{Power}}$       | Perf/kW                           | $+\infty$                       | Inaccurate comparisons because of the variety of workloads.           | [50–52]    |
| <i>Server compute Efficiency (ScE).</i><br>Server compute output per energy used.             | $\frac{\text{Output}}{\text{Server Power}}$     | Ops/kW                            | $+\infty$                       | Complexity in isolating server-level outputs.                         | [53,54]    |
| <i>Data Center Workload Power Efficiency (DWPE).</i> Workload processed per unit of IT power. | $\frac{\text{Workload}}{\text{IT Power}}$       | Task/kW                           | $+\infty$                       | Difficulties in establishing workload uniformity between platforms.   | [55,56]    |
| <i>Average Power Consumption (APC).</i><br>Average IT equipment power consumption over time.  | $\frac{\text{Total IT Power}}{\text{Time}}$     | kW                                | The lower the value, the better | Power usage tracking is challenging because of workload fluctuations. | [57,58]    |

Table 1. Cont.

| Energy-Efficient Metric                                                                                                   | Formulation                                           | Units                  | Optimal Value | Challenges                                                                  | References |
|---------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------|------------------------|---------------|-----------------------------------------------------------------------------|------------|
| Energy Wastage Ratio (EWR). Energy wasted per unit of computational work.                                                 | $\frac{IT\ Energy}{Work\ Output}$                     | kWh/Task               | 0             | Technical difficulties in identifying energy waste.                         | [59,60]    |
| IT Equipment Energy Efficiency (ITTE). IT hardware efficiency.                                                            | $\frac{Rated\ Performance}{Rated\ Power}$             | %                      | 1             | Efficiency variation according to types of workloads                        | [45,46]    |
| Operational System Workload Efficiency (OSWE). System workload output per total energy.                                   | $\frac{Workload}{Total\ System\ Energy}$              | %                      | 1             | Defining system boundaries in shared environments can be challenging.       | [46,61]    |
| Non-IT-related metrics                                                                                                    |                                                       |                        |               |                                                                             |            |
| Power Usage Effectiveness (PUE). Ratio between the total facility energy consumed and the energy consumed only by the IT. | $\frac{Total\ power}{IT\ equipment\ energy}$          | Ratio                  | 1             | It ignores IT use in favor of infrastructural energy.                       | [39,62]    |
| Total-Power Usage Effectiveness (TUE). Total resource usage (internal and external energy use).                           | $PUE * \frac{Total\ IT\ Energy}{Compute\ Energy}$     | Ratio                  | 1             | Difficulties in capturing overall energy usage.                             | [63,64]    |
| Data Center infrastructure Efficiency (DciE). Infrastructure energy ratio.                                                | $\frac{IT}{Total\ facility\ energy}$                  | Ratio                  | 1             | Absence of new perspectives without additional measurements.                | [48,65]    |
| Carbon Usage Effectiveness (CUE). Carbon emissions per IT energy.                                                         | $\frac{CO_2}{IT\ equipment\ energy}$                  | kgCO <sub>2</sub> /kWh | 0             | Lifecycle emissions and offset strategies are not taken into consideration. | [66,67]    |
| Data Center energy Productivity (DCeP). Energy productivity.                                                              | $\frac{Useful\ work}{Total\ facility\ energy}$        | Output/kW              | 1             | Lacks standards and uses abstract services to measure productivity.         | [68,69]    |
| Data Center green Efficiency (DCgE). Green energy efficiency.                                                             | $\frac{Renewable\ energy}{Total\ energy} \times DCiE$ | Ratio                  | 1             | Accessing reliable green energy data is difficult.                          | [45]       |

Table 1. Cont.

| Energy-Efficient Metric                                               | Formulation                                      | Units     | Optimal Value | Challenges                                                                                         | References |
|-----------------------------------------------------------------------|--------------------------------------------------|-----------|---------------|----------------------------------------------------------------------------------------------------|------------|
| Data Hall Utilization Efficiency (DH-UE). IT floor space utilization. | $\frac{\text{Active floor}}{\text{Total hall}}$  | Ratio     | 1             | Distorted by cooling techniques and layout heterogeneity.                                          | [71,72]    |
| Data Hall Utilization Rate (DH-UR). Rack utilization rate.            | $\frac{\text{Deploy racks}}{\text{Total racks}}$ | Ratio     | 1             | Reflects the use of space but not thermal or power efficiency.                                     | [70,71]    |
| Power Delivery Efficiency (PDE). Power loss assessment.               | $\frac{\text{IT power}}{\text{Input power}}$     | Ratio     | 1             | Isolating energy losses in converters and UPS (Uninterruptible Power Supply) systems is difficult. | [46,73]    |
| Data Center Performance Efficiency (DCPE). Performance per energy.    | $\frac{\text{Output}}{\text{Facility energy}}$   | Output/kW | 1             | Complex benchmarking is necessary for a variety of services.                                       | [74,75]    |

### 3.3. Energy Management Strategies

The use of fossil fuels, which commonly supply the increasing power demand of AI-driven data centers, leads to significant greenhouse gas emissions, contributing to environmental degradation and climate change. This situation has raised global ecological concerns and has sparked a need for sustainable solutions.

It is important to note that hardware efficiency directly impacts electricity consumption [76]. Thus, one approach to reducing the power consumption of AI-driven data centers could be to replace obsolete equipment with new technologies optimized for minimal energy consumption without compromising computing performance. Another factor that greatly contributes to the energy demand of AI-driven data centers is the cooling system, which can contribute up to 40% of the total energy demand [77]. Given that AI-driven data centers generate substantial operational heat, traditional air-cooling systems are insufficient [78]. It is for this reason that new cooling technologies are critical to keep energy use and costs down. In this sense, innovative cooling techniques such as immersion cooling, liquid cooling, hot aisle containment, and heat reuse can reduce overall power demand, facilitating the integration of renewable energy sources such as solar and wind power [79]. Mavisclara et al. [80] reported that liquid cooling can reduce cooling energy consumption by 20% compared with air-cooling methods. Similarly, Microsoft has demonstrated that immersion cooling allows for a higher computing capacity without an increase in energy demand [81]. Another cornerstone strategy for power demand reduction is the use of virtualization technologies. These technologies allow for multiple virtual machines and workloads to run on a single server, thus decreasing the need for additional hardware and optimizing resource use by multi-task scheduling and load shifting to geo-distributed data centers with power availability [82].

However, even if the power consumption in AI-driven data centers is reduced, the continued reliance on fossil fuel-based energy for power generation remains an environmental concern. In view of this, renewable energy emerges as a promising alternative by allowing data centers to optimize their energy mix according to resource availability at a specific location, reducing grid dependence and potentially lowering energy costs while also supporting advanced cooling systems crucial for AI workloads. The International Renewable Energy Agency (IRENA) indicated that from 2010 to 2023, solar photovoltaic (PV) technology costs decreased by 80%, making on-site and off-site generation economically feasible at hyperscale for AI-driven data centers [83]. Google in Nevada corroborated this claim by obtaining 90% carbon-free energy for its data centers through off-site solar PV generation [19]. Furthermore, the scalability and modularity of solar PV installations make them suitable for data centers of varying scales [27]. For data centers located in regions where hydropower is available, it can provide a more stable and dispatchable electricity supply that complements intermittent solar and wind generation. In Texas, Amazon has been able to obtain 50% of its power demand from wind farms, while Microsoft in Sweden obtained almost 100% of renewable energy for its data centers from hydropower [82].

Barriers to the application of innovative cooling techniques and renewable energy sources include a high initial investment cost, even though operation will become cost-effective in the long term due to lower operational costs, expertise gaps, and the handling of environmental concerns that involve technology waste and land use for large-scale solar and wind farms [83]. However, the intermittency of solar and wind power poses a major challenge in the implementation of these techniques due to their dependence on weather and time, making it difficult to match energy production with the variable demand of data centers [27]. To achieve the success of data centers, a careful location assessment must be performed to ensure the optimization and integration of these energy sources. It is also necessary to have flexible storage technologies, such as batteries, that store wind and solar power to be used either in low or intermittent production periods or complementary non-intermittent renewable energies such as nuclear, hydropower, or geothermal. Microgrids are another promising alternative to integrate renewable energy sources in data centers at a local level. Microgrids consist of localized energy distribution systems that can operate in conjunction with or independently from the core power grid [84]. Jones [85] demonstrated that microgrids co-localizing non-intermittent renewable energies like geothermal with hybrid battery storage provide optimal cost efficiency, rapid implementation, and essential reliability to manage the extreme load spikes of generative AI workloads. Nonetheless, a life cycle assessment is needed to develop a holistic approach to AI-driven data center energy consumption, which would take into account the manufacturing, transportation, operational, and end-of-life disposal processes in order to balance economic, operational, environmental, and regulatory considerations.

Currently, AI is being used to optimize data center operations. Among these operations, AI predicts workloads, manages resources, operates cooling systems, enhances the overall efficiency of data centers, and reduces energy waste [86]. AI enables the efficient use of resources, including energy, based on real-time demand by predicting peak loads and automatically adjusting resource allocation to meet requirements, thereby preventing waste during low-demand periods [5]. AI technologies also allow cooling systems to be automated and optimized. By analyzing temperature data, cooling requirements are predicted, and cooling mechanisms are dynamically adjusted to achieve maximum efficiency [78]. Data centers can also implement intelligent lighting and ventilation systems to enhance resource efficiency. Figure 2 presents several of the strategies discussed to reduce fossil fuel-based power demand.

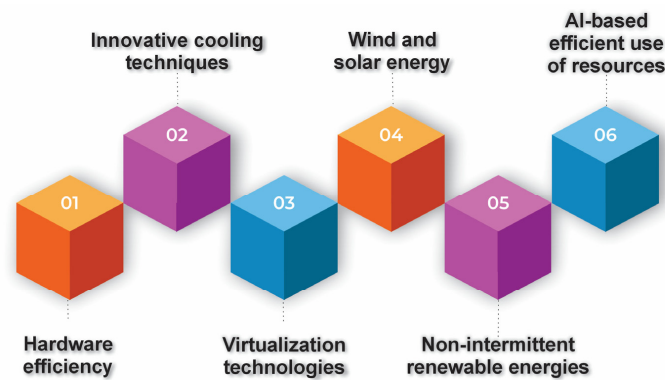


**Figure 2.** Strategies for reducing fossil fuel-based power demand.

## 4. Water Use and Cooling Systems

### 4.1. Cooling Technologies

Cooling technologies are critical in the operation of data centers. As computational complexity and power density continue to increase, heat generation within these facilities rises significantly, requiring more efficient cooling systems. Earlier studies of conventional data center servers reported that, although most servers operated at only around 20% of their capacity, data centers could still draw between 60% and 100% of their maximum power from the grid due to infrastructure inefficiencies, even when idle [78]. As computing infrastructure has evolved toward more demanding and energy-intensive workloads, particularly those associated with AI, these challenges have become increasingly relevant because higher computational densities result in greater heat generation and cooling requirements. Traditional cooling infrastructure, which relies primarily on mechanical vapor compression air conditioning systems, accounts for 30% to 50% of the total energy consumption of the facility [87].

To optimize airflow distribution and efficiency, various alternative methods have been implemented. These air-cooling systems have been conventionally classified into four main configurations: room-level, row-level, rack-level, and chip-level cooling [3]. Although design parameters and geographical techniques vary from country to country, incorporating systems such as thermosiphons or free cooling, all configurations are designed to maintain a stable thermal environment.

#### 4.1.1. Air Cooling Configurations and Implementation

The distinction between room-level, row-level, and rack-level systems lies primarily in their airflow distribution mechanisms, although they provide comparable net cooling capacities [88]. Row- and rack-level configurations offer greater efficiency by minimizing the distance that air must travel to the cooling unit, thereby reducing fan power consumption and supporting thermal loads of up to 30 kW per rack [89].

- Row-based cooling: Uses units installed between the racks or suspended above them. Cool air is directed into a contained aisle, while hot exhaust air from the rear of the racks is captured and recirculated directly to the cooling unit. This configuration minimizes heat loss and is considered ideal for data centers with total capacities of less than 200 kW [90].
- Chip-level cooling: This configuration directs heat dissipation directly to the component. However, its widespread adoption remains limited due to the high implementation and maintenance costs involved [91].

Historically, air-cooling systems have been the predominant technology, continuously evolving to maintain adequate operating conditions in the face of increasing power densities. However, the increasing heat flux generated by high-performance processors is approaching the practical limits of conventional air-based cooling [92]. At the chip level, heat fluxes exceeding approximately  $0.1 \text{ kW/cm}^2$  can challenge the ability of air cooling to effectively remove heat within the processor footprint [93,94]. This limitation becomes increasingly relevant as the power consumption of individual processors and GPUs continues to rise, with high-performance GPUs potentially approaching power levels of approximately 2 kW [95]. These trends are accelerating the transition toward liquid cooling technologies for high-density computing and AI workloads [79]. Figure 3 summarizes the main cooling technologies currently employed in data centers, highlighting the transition from traditional air-cooling systems toward liquid- and immersion-cooling approaches capable of managing the increasing thermal loads associated with high-performance computing and AI workloads.

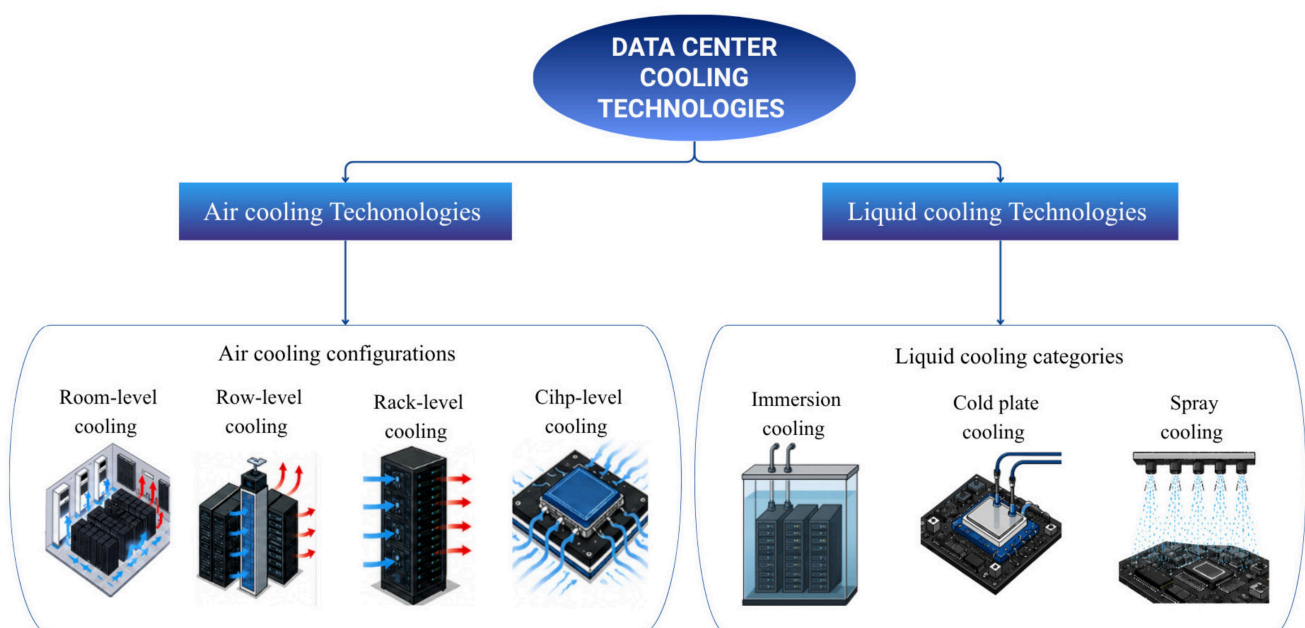


**Figure 3.** Overview of cooling technologies.

#### 4.1.2. Transition to Liquid Cooling Technologies

As power densities rapidly increase, air cooling faces structural limitations due to the low thermal conductivity and specific heat capacity of air. Attempting to mitigate high thermal loads using air results in prohibitively high power consumption by fans, excessive noise, and enormous surface areas [96]. This operational challenge is exacerbated in tropical regions, where the hot, humid ambient air requires a high, continuous supply of electrical power to operate auxiliary equipment such as chillers, pumps, and fans [88]. As a result, liquid cooling systems are emerging as a highly efficient alternative for high-density data centers. While an air-cooled rack typically supports a power density of around 20–30 kW, liquid cooling technologies can substantially increase rack-level power density, with advanced systems supporting values of approximately 150 kW per rack [88]. As AI workloads continue to increase processor power and rack-level power density, future data center architectures are expected to target power levels approaching 1000 kW per rack, further reinforcing the need for advanced liquid cooling technologies [97,98].

Liquid cooling technologies can be classified into three categories based on their heat dissipation method [99]:

- Immersion cooling: Server components are fully or partially submerged in a dielectric fluid, which absorbs heat directly through advanced heat transfer processes [92].
- Direct-to-chip cooling: This method relies on direct heat conduction between a high-thermal-conductivity plate through which a liquid coolant circulates and the components that generate the most heat (such as CPUs or GPUs). Secondary server components are subsequently cooled through localized air circulation [100].
- Spray cooling: This method uses specialized nozzles to atomize and spray the coolant directly onto the hot surfaces of the components, achieving rapid heat dissipation [101].

As summarized in Table 2, the transition to liquid cooling is not merely an optional advancement, but rather a thermodynamic constraint imposed by the evolution of AI hardware.

**Table 2.** Comparison of data center cooling strategies.

| Refrigeration Settings            | Flow Distribution Mechanism/Technology                                                                                                  | Power Density Limits                                                                                                                 | Benefits and Efficiency Features                                                                                                                | Benefits and Efficiency Features                                                                                           | References   |
|-----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|--------------|
| Room-Level Cooling                | Air conditioning system based on mechanical vapor compression refrigeration.                                                            | Typically, low density (<20 kW/rack).                                                                                                | A predominant, widely adopted, and historically standardized traditional technology.                                                            | It consumes between 30% and 50% of the total energy. It suffers significant heat loss due to the distance the air travels. | [100,102]    |
| In-Line Cooling                   | Units installed between the racks or suspended from the ceiling, blowing air into an aisle and capturing the exhaust air from the back. | Supports up to 30 kW per rack.                                                                                                       | It reduces the air path length and minimizes heat loss. Ideal for data centers with a capacity of less than 200 kW.                             | It requires a greater number of localized units compared to the central system.                                            | [103,104]    |
| Rack Cooling                      | Air distribution units attached to or directly integrated into the frame structure.                                                     | Supports up to 30 kW per rack.                                                                                                       | High efficiency by drastically reducing the airflow distance and lowering the power required by the fans.                                       | Greater complexity in the technology floor's cabling layout and dependence on rack modularity.                             | [105,106]    |
| Chip-Level Cooling (Air)          | Microarchitectures designed to maintain operating conditions directly within the processor.                                             | Focused on high density of individual components.                                                                                    | Localized response capability in the face of sudden increases in power density.                                                                 | Rarely used in practice due to the high costs associated with its implementation and structural maintenance.               | [107,108]    |
| Natural Systems and Thermosiphons | Integration of natural cooling (free-cooling) cycles or thermosiphons to ensure thermal stability.                                      | Varies depending on the architectural design.                                                                                        | Designed to maintain a stable temperature by reducing reliance on mechanical compression.                                                       | Highly dependent on the temperature and humidity conditions of the country or region where the equipment is operated.      | [109,110]    |
| Immersion Cooling (Liquid)        | Total or partial submersion of the servers in a heat-absorbing dielectric fluid.                                                        | Advanced systems can support substantially higher rack power densities, with representative values of approximately 150 kW per rack. | Optimized heat absorption through advanced heat transfer processes; this enables greater computing power without increasing energy consumption. | High initial investment; requires evaluating fluid compatibility and hardware maintenance procedures.                      | [92,111,112] |

Table 2. *Cont.*

| Refrigeration Settings          | Flow Distribution Mechanism/Technology                                                                   | Power Density Limits                                                                                                                 | Benefits and Efficiency Features                                                                                                                | Benefits and Efficiency Features                                                                                                                                                          | References   |
|---------------------------------|----------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| Direct-to-chip Cooling (Liquid) | Direct contact between a plate with circulating coolant and the latest-generation components (CPU/GPU).  | Advanced systems can support substantially higher rack power densities, with representative values of approximately 150 kW per rack. | Heat dissipation is focused on the areas of highest thermal stress; the remaining secondary components are cooled using localized airflow.      | It requires hybrid (liquid/air) systems and a precise internal piping design to prevent the risk of leaks.                                                                                | [91,113,114] |
| Spray Cooling (Liquid)          | Use of specialized nozzles that spray the coolant onto hot surfaces.                                     | Advanced systems can support substantially higher rack power densities, with representative values of approximately 150 kW per rack. | Accelerated heat dissipation via phase-change heat transfer or direct liquid film heat transfer.                                                | Mechanical complexity in the spray nozzles and strict control of fluid distribution.                                                                                                      | [115]        |
| Evaporative Cooling/Economizers | Direct outside air injection (free cooling) or evaporative precooling of the intake airflow using water. | Typically limited to low and moderate densities (<25 kW/rack) when operating autonomously.                                           | Significant net energy savings (OPEX) and a reduction in PUE by reducing or eliminating the use of mechanical compression; high sustainability. | Significantly reduced efficiency in tropical regions due to the hot, humid ambient air; requires rigorous water treatment and advanced filtration to prevent corrosion and contamination. | [116,117]    |

Consequently, the accumulated scientific evidence suggests that the search for a single, universal, and definitive cooling system is not feasible. The paradigm of maximum efficiency does not lie in the adoption of a single technological approach (whether air-based or liquid-based), but rather in the convergence toward hybrid systems.

#### 4.2. Water Consumption

The quantification of carbon and water footprints is typically assessed by first analyzing location and energy consumption, since both the electricity mix and the carbon intensity of the energy used to power data centers vary according to location. Direct water consumption refers to the water used in cooling systems, while indirect consumption refers to the water consumption embedded in the electricity purchased to operate the data centers [9]. Among the main pathways of direct water consumption, evaporative cooling plays a particularly important role, as water is used to remove heat from IT equipment and is partially lost through evaporation. Additional water may also be required for blowdown and system maintenance. Other cooling configurations, such as air-cooled and closed-loop liquid cooling systems, can reduce direct on-site water consumption, although their effectiveness depends on system design and operating conditions. It is estimated that in 2023, 560 billion liters of water were consumed in data centers. Out of the 560 billion liters of water, 373 billion liters were used for indirect consumption, and 140 billion liters were used for direct consumption. The remaining 47 billion liters are believed to have been used in the manufacturing of servers [118].

Table 3 presents data on water and electricity consumption for some of the world's largest companies and highlights the lack of consistency in data disclosure by the largest companies within the sector. The direct water consumption values reported in Table 3 refer to water consumed at the data center facilities and therefore exclude the water consumption embedded in purchased electricity. In terms of water efficiency, direct consumption shows

significant variations ranging from 0.17 to 1.61 L/kWh [118]. This suggests that water consumption does not depend exclusively on electricity demand, but also on cooling technology, on local climate conditions, on the extent of water reuse, and on the measures implemented by each company. The results show that the actual water footprint of a data center may be underestimated when only the operations of the facility are considered [119,120]. Furthermore, the study shows that the PUE (a measure of a data center’s energy efficiency) and WUE (a measure of water use efficiency during data center operations) indicators are not necessarily correlated; for example, Google and Meta report very similar PUE values, yet their WUE values differ, suggesting that greater energy efficiency does not necessarily imply a lower water demand.

**Table 3.** Reported direct and indirect water consumption and electricity consumption in the sustainability reports of some of the world’s largest companies. Copied from “The Carbon and Water Footprints of Data Centers and What This Could Mean for AI” [118].

| Reporting Metrics          | Scope of Reporting           | Unit  | Company (Year of Publication of the Most Recent Environmental Report) and Reporting Period |                                  |                                  |                                    |                                    |
|----------------------------|------------------------------|-------|--------------------------------------------------------------------------------------------|----------------------------------|----------------------------------|------------------------------------|------------------------------------|
|                            |                              |       | Amazon (2025)                                                                              | Apple (2025)                     | Baidu (2025)                     | Google (2025)                      | Meta (2025)                        |
|                            |                              |       | CY 2024                                                                                    | Fiscal Year 2024                 | CY 2024                          | CY 2024                            | CY 2024                            |
| Electricity consumption    | Total Company                | kWh   | -                                                                                          | 3,776,000<br>$3.776 \times 10^9$ | 1,145,028<br>$1.145 \times 10^9$ | 32,106,200<br>$32.106 \times 10^9$ | 18,423,634<br>$18.424 \times 10^9$ |
|                            | Total number of data centers | kWh   | -                                                                                          | 2,515,000<br>$2.515 \times 10^9$ | -                                | 30,825,600<br>$30.826 \times 10^9$ | 18,061,781<br>$18.062 \times 10^9$ |
|                            | Average PUE for data centers | N/A   | 1.15                                                                                       | -                                | 1.20                             | 1.09                               | 1.08                               |
| Direct water consumption   | Total Company                | L     | -                                                                                          | 3,406,871<br>$3.407 \times 10^9$ | 1,465,129<br>$1.465 \times 10^9$ | 30,794,325<br>$30.794 \times 10^9$ | 3,123,000<br>$3.123 \times 10^9$   |
|                            | Total number of data centers | L     | -                                                                                          | -                                | -                                | 29,477,002<br>$29.477 \times 10^9$ | 2,974,000<br>$2.974 \times 10^9$   |
|                            | Average WUE of data centers  | L/kWh | 0.15                                                                                       | -                                | 1.61                             | -                                  | 0.19                               |
| Indirect water consumption | Total Company                | L     | -                                                                                          | -                                | -                                | -                                  | 72,207,000<br>$72.207 \times 10^9$ |
|                            | Total number of data centers | L     | -                                                                                          | -                                | -                                | -                                  | -                                  |
| Implicit water intensity   | Total Company (direct)       | L/kWh | -                                                                                          | 0.90                             | 1.28                             | 0.96                               | 0.17                               |
|                            | Total Company (indirect)     | L/kWh | -                                                                                          | -                                | -                                | -                                  | 3.92                               |
|                            |                              |       | Microsoft (2025)                                                                           | Oracle (2025)                    | Tesla (2025)                     | Tencent (2025)                     |                                    |
|                            |                              |       | Fiscal Year 2024                                                                           | CY 2023                          | CY 2024                          | CY 2024                            |                                    |
|                            |                              |       |                                                                                            |                                  |                                  |                                    |                                    |
| Electricity consumption    | Total Company                | kWh   | 29,829,540<br>$29.830 \times 10^9$                                                         | 3,623,170<br>$3.623 \times 10^9$ | 2,355,046<br>$2.355 \times 10^9$ | 6,429,610<br>$6.430 \times 10^9$   |                                    |
|                            | Total number of data centers | kWh   | -                                                                                          | -                                | -                                | -                                  |                                    |
|                            | Average PUE for data centers | N/A   | 1.16                                                                                       | -                                | -                                | 1.29                               |                                    |

Table 3. Cont.

|                            |                              |       | Microsoft<br>(2025)              | Oracle<br>(2025)               | Tesla (2025) | Tencent<br>(2025)                  |
|----------------------------|------------------------------|-------|----------------------------------|--------------------------------|--------------|------------------------------------|
|                            |                              |       | Fiscal Year<br>2024              | CY 2023                        | CY 2024      | CY 2024                            |
| Direct water consumption   | Total Company                | L     | 5,807,000<br>$5.807 \times 10^9$ | 866,069<br>$0.866 \times 10^9$ | -            | 10,370,588<br>$10.370 \times 10^9$ |
|                            | Total number of data centers | L     | -                                | -                              | -            | -                                  |
|                            | Average WUE of data centers  | L/kWh | 0.30                             | -                              | -            | -                                  |
| Indirect water consumption | Total Company                | L     | -                                | -                              | -            | -                                  |
|                            | Total number of data centers | L     | -                                | -                              | -            | -                                  |
| Implicit water intensity   | Total Company (direct)       | L/kWh | 0.19                             | 0.24                           | -            | 1.61                               |
|                            | Total Company (indirect)     | L/kWh | -                                | -                              | -            | -                                  |

Regarding the operational water impact, the estimated indirect water footprint of data centers due to their electricity demand reached 383 billion liters in 2018. In contrast, the indirect water footprint attributed to the water supply and wastewater treatment companies serving these facilities was several orders of magnitude lower, at 0.45 billion liters [121]. Nationwide in developed countries it is estimated that 1 kWh of energy consumption by a data center requires a weighted average of 7.1 L of water to support its operations [122].

#### 4.3. Energy–Water Trade-Offs

The analysis of operational trade-offs between energy and water has driven the development of innovative strategies aimed at resource conservation, which include the integration of recycled water into thermal management systems and the recovery of waste heat for reinjection into energy networks. However, the crux of this debate lies in the fact that reducing Power Usage Effectiveness (PUE) through evaporative or adiabatic cooling systems (which harness the latent heat of vaporization to reduce the mechanical compression load) lowers the electricity consumption of the infrastructure but can significantly increase Water Usage Effectiveness (WUE) at the local (on-site) level [115]. Conversely, the transition process to utilize dry cooling systems completely eliminates the direct dependence of a facility on water but degrades thermal efficiency in hot climates, increasing the power consumption of fans and compressors; this shifts the water footprint to the thermal power plants in the regional power grid (off-site) [110]. Importantly, improving liquid-cooling efficiency can also help mitigate the energy–water trade-off by enabling operation at higher coolant inlet temperatures. Higher coolant temperatures increase the potential for free cooling and can reduce reliance on mechanical chillers and compressors, while facilitating the use of closed-loop dry coolers. Recent NVIDIA AI infrastructure designs, such as the Vera Rubin platform, are engineered to operate with coolant inlet temperatures of up to 45 °C. Under suitable climatic conditions, this operating temperature enables the use of ambient air and closed-loop dry coolers, reducing the need for mechanical cooling and, consequently, cooling-related energy and water consumption [123]. Therefore, the sustainability of a digital infrastructure requires a co-dependent optimization approach, in which thermal design decisions are not evaluated in isolation but rather done by minimizing the combined consumption of water and equivalent energy [9,119].

Recent quantitative models show that, if the water usage patterns recorded in 2024 are projected, the expansion of computing infrastructure by 2030 will require a net incremental global water capacity of between 2.638 and 5.493 billion liters per day (BLD) [9]. This volume of direct water withdrawal is equivalent to adding a demand load comparable to that of major global metropolitan areas such as New York City (3.785 BLD). Even under an optimistic technological scenario, assuming a 10% annual compound reduction in sectoral consumption intensity, the remaining water pressure would range between 0.859 and 2.286 BLD, which helps us grasp the magnitude of the problem when considering only this aspect of it [121]. Now, imagine a scenario where the correlations between necessary resources, costs, location, and environmental damages are analyzed. We must assess whether the current situation facing society and the world is truly viable in order to provide a response that is as close to reality as possible [121].

## 5. Material Demand and Supply Constraints

### 5.1. Materials in Data Center Infrastructure

Data centers comprise a complex infrastructure that houses thousands of servers as well as network devices, storage systems, processing units, cooling systems, power supplies, backup power systems, monitoring sensors, connection cables, and many other essential components that enable the growing AI computing demands in AI-driven data centers [12]. Regardless of this, the lack of detailed public data concerning the components' design and volume makes it difficult to verify which materials are being used in data centers and their impact on the growing demand for AI materials.

In addition to common bulk materials like concrete and steel, the construction of an AI-driven data center requires a wide variety of metals and minerals that include germanium, gallium, indium, metallurgical silicon, tantalum, copper, platinum-group metals, rare earth metals, silver, and gold, among many others [124]. It is important to highlight that the International Renewable Energy Agency (IRENA) recognizes critical minerals as those produced within a limited number of countries or whose extraction is demanding [125]. Once extracted, these metals and raw materials are refined to achieve the specific magnetic, electronic, mechanical, and optical characteristics needed for the technological infrastructure of a data center. Nonetheless, the mining of these resources has created severe environmental impacts that include soil degradation, air pollution, and water contamination [126]. Mining alone accounts for approximately 7% of global greenhouse gas emissions [127], and 90% of mining operations use between 340 and 6270 L of water per ton of ore processed, while surface mining exhibits approximately a 35% loss of vegetation cover [128].

Semiconductor chips, integrated circuits, or microchips are the fundamental hardware that powers AI systems. To manufacture these semiconductor chips, high-purity metals are needed to enhance chip performance by enabling thermal stability, excellent conductivity, and corrosion resistance [129]. According to Columbia University, semiconducting materials include arsenic, aluminum, gallium, indium, germanium, molybdenum, lithium, tantalum, silicon, tungsten, cerium, erbium, dysprosium, gadolinium, europium, lanthanum, praseodymium, neodymium, terbium, yttrium, and thulium [130]. Table 4 shows some of the materials commonly used in AI-driven data centers along with their principal applications, as well as the primary sectors competing for the use of such material.

**Table 4.** Materials commonly used in AI-driven data centers.

| Material                                                               | Main Applications                                                                                                                                                                                      | Competing Demand Sectors                                                                                                          |
|------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| Copper                                                                 | Because of its durability and conductivity, it is commonly used in data cables and networking, power distribution systems, and cooling infrastructure for pipes and heat exchangers [131].             | Industrial machinery, building, defense, electric vehicles and renewable energy interconnection [132].                            |
| Silicon                                                                | High-purity silicon is the most broadly used semiconductor for high-speed memory and processors [130]                                                                                                  | Automotive semiconductors, lithium-ion battery anodes, solar photovoltaics, fuel cells, consumer and power electronics [133,134]. |
| Gallium                                                                | In the form of semiconductors such as gallium nitride and gallium arsenide, gallium is used in radio frequency components and high-frequency and high-efficiency power converters [14].                | LED illumination, solar cells, 5G/telecommunications, and defense [135,136].                                                      |
| Aluminium                                                              | Given its weight and exceptional thermal conductivity, it is used for structural components (server racks, mounting frameworks, and casing) and cooling systems (heat sinks and cooling plates) [124]. | Grid conductors, construction, transportation (aerospace, marine and automotive), packaging, and defense [137,138].               |
| Rare earth elements (neodymium, praseodymium, dysprosium, and terbium) | In particular, neodymium, praseodymium, dysprosium, and terbium are used in high-performance magnets in cooling fan motors, optical components, and precision actuators [139].                         | Wind turbine generators, consumer electronics, defense, and electric vehicle traction motors [140,141].                           |

It is important to highlight that the production of solar and wind energy, renewable energy sources that can potentially supply data centers, also depends on minerals and metals. The IEA has identified that, in comparison to a natural gas plant of the same capacity, an onshore solar power plant requires nine times more mineral resources [142]. According to the projections developed by Amoah et al. [132] for 2005 to 2030, AI-driven data centers' mineral demand for copper could increase from 576,000 tons to 1.88 million tons, for aluminum from 22,000 to 74,000 tons, for iron from 24,000 to 82,000 tons, and for lithium carbonate from 3300 to 20,000 tons over the same time period. Amoah et al. [132] also identified that the key constraint for these minerals is their availability and the capacity for mining extraction to meet the substantial demands for building and powering AI-driven data centers; additionally, for rare earth elements, the weaknesses lie in midstream refining and processing limitations, along with geopolitical concentration. It is important to keep in mind that the increase in demand for AI is not only driving up the mining of metals and minerals but also their environmental repercussions.

## 5.2. Supply Chain Issues

The supply of critical metals and minerals required to fulfill the AI infrastructure demand is not only dependent on their extreme geographical concentration and market logistics, but it is also influenced by geopolitics. Nowadays, China dominates the production market of raw materials, such as gallium, antimony, germanium, and other rare earth elements, necessary to generate AI; Chile and the Democratic Republic of the Congo are

major contributors to the world's copper supply; the United States, Canada, and Malaysia produce smaller shares of silicon and rare earth elements; Australia is one of the main producers of lithium; the Democratic Republic of Congo is the leading producer of cobalt, and South Africa produces most of the platinum-group metals [124].

The United States leads the world in the manufacturing of AI components from metal and mineral raw materials, as it possesses the advanced manufacturing technology required. NVIDIA, Intel, and AMD are examples of American-based companies with the technology required for chip design. NVIDIA, alone, collaborates with a wide variety of refining, mining, and smelting companies, and it acquires mineral raw materials from 259 companies worldwide [143]. Nonetheless, over 70% of semiconductor manufacturing is outsourced by East Asian companies such as South Korea's Samsung Foundry and Taiwan's TSMC [31].

There are geopolitical fluctuations, such as export restrictions and fluctuations, that could restrict global access to the computing abilities of AI. In 2023–2024, China imposed a series of export restrictions on gallium, germanium, graphite, tungsten, and bismuth as a response to U.S. technology sanctions [124]. As a result, gallium prices outside of China doubled in just six months. Meanwhile, export restrictions on AI processors, such as NVIDIA's A100 and H100, have been enforced by the United States, thereby denying Chinese consumers access to state-of-the-art hardware [143]. The fact that the concentration of computing power and raw material supply chains relies on a few economies creates dependencies that limit AI development and risk sustainability.

### 5.3. End of Life and Waste

The continuously growing demand for raw materials associated with AI-driven data centers is also creating a great amount of disposable AI hardware. Many data center components have a short service life cycle, which requires them to be replaced after a period of around two to five years [144]. The removal of these minerals and metals from the supply chain increases the extraction demand for these raw materials, and even when some materials are recovered, their disposal represents a potential environmental and human health risk due to the presence of toxic elements such as arsenic, mercury, and lead in them, which can seep into the water and soil.

To ensure that potentially dangerous components are handled safely, particular methods to collect, sort, shred, and disassemble are implemented to segregate these materials. Regarding mechanical separation techniques, magnetic separation is used with ferrous metals, while with non-ferrous metals, eddy current separation is implemented. Additionally, pyrometallurgical (smelting electronic scrap to extract metals) or hydrometallurgical (chemical leaching to recover metals) processes are also used for further refinement. Bioleaching, which uses microorganisms to recover metals from crushed e-waste, is another emerging research area [145]. Figure 4 summarizes the key steps in AI-driven data center waste treatment.

High-income countries implement formal collection procedures even when around 40% of disposals are collected [146]. Meanwhile, many people in low-income countries are being exposed to dangerous chemicals, such as furans and dioxins, where informal and harsh recycling methods are implemented to extract valuable metals from components or where disposals are burned. Although it is known that recycling generates emissions and consumes energy, the life cycle emissions of an AI infrastructure decrease when materials are recovered in comparison to those of virgin extraction. For example, the recycling of aluminum saves around 95% of the energy that is needed to obtain it from raw sources, while the recycling of copper saves around 85% [147]. In a similar way, recycling rare earth elements can partially mitigate the environmental damage that is caused by their mining [148].

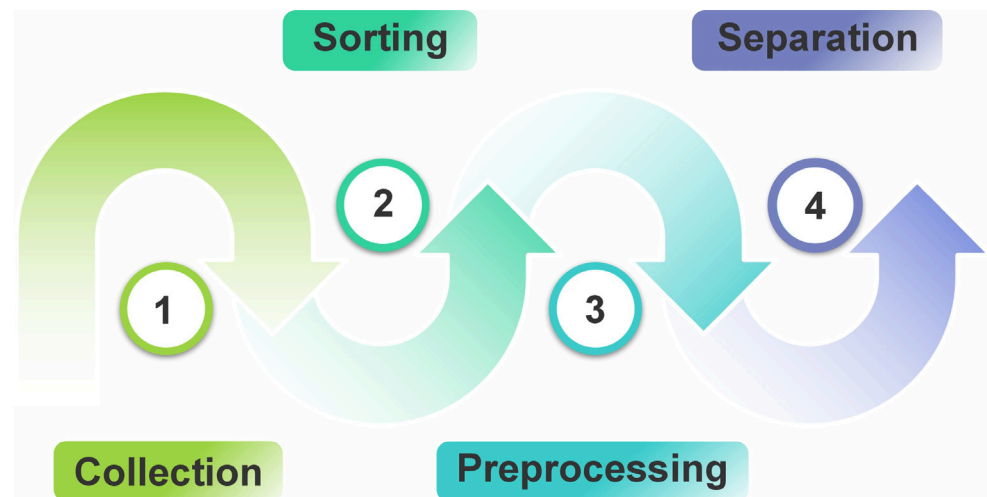


**Figure 4.** Key steps for AI-driven data center waste treatment.

Nevertheless, the recycling of raw materials used in AI-driven data centers presents several challenges, as some metals are difficult to separate or are present only in trace quantities. In the case of lithium-ion batteries, careful handling is required to prevent accidents [145]. Furthermore, recycling also depends on economic factors; subsidies are essential when the cost of recycling exceeds the recovered market value of materials. In this sense, expanding recycling infrastructure and developing more efficient recycling methods is crucial. Recent recycling advances use supercritical fluid extraction or innovative solvent techniques to recycle circuit boards [147]. The circular economy also surges as a powerful strategy to counterbalance recycling challenges by focusing on creating products with a longer lifespan, refurbishing, reusing, and applying a disassembly-friendly design [149].

## 6. Resource Interactions and Limitations of Current Approaches

As analyzed in previous sections, AI-driven data centers require significant amounts of water, energy, and materials. The increasing proliferation of AI creates a surge in demand for these resources, limiting their availability for other vital applications. Water is the cornerstone of every living system on our planet, a resource that is essential and crucial to human societies and ecosystems [150]. Because water and energy resources are fundamental to food production and processing, their depletion could lead to food shortages for the world's growing population and threaten the well-being of all living systems.

Studies have also identified the interdependence between the hydric, energetic, and material resources used for AI-driven data centers [151]. Water enables hydropower generation that is used as a steam generator for thermal and nuclear power plants, and that is essential in AI cooling systems and in the extraction and processing of mineral resources and hydrocarbons. Energy resources are essential in the collection, distribution, sanitation, and treatment of water as well as in the extraction and processing of mineral resources and in the operation of AI-driven data centers. Metal and mineral resources are key for the implementation of renewable energies as well as for the implementation of AI and water management due to them providing the structural foundation for treatment infrastructure.

Because the use of any of these resources will affect or constrain other resources, no single element (water, energy, and materials) can adequately explain AI. In order to improve energy efficiency, some facilities change from air-cooling systems to evaporative ones, an approach that increases water usage. On the other hand, water-free cooling options, like closed-loop chillers, can raise energy consumption by 10 to 15% [29]. Given these trade-offs, some authors [151,152] propose that it is necessary to change from efficiency to sufficiency.

This means that the efficiency indicator must be complemented by absolute consumption limits in accordance with the resource use limitations within ecological boundaries. For instance, licenses for data centers could be granted based on carbon intensity targets and water consumption limits in relation to the hydrological capacity of the region. In the same way, sufficiency also implies coordinating and migrating geographical workloads according to favorable conditions, such as cooler climatic conditions to decrease cooling requirements and an oversupply of renewables [153]. Additionally, policies that deal with resource recovery are fundamental to enhance circular economy principles and sufficiency, given that circularity promotes a decrease in resource pressure. Heat recovery from operations in data centers, minerals and metals recovery from electronic waste, and ultra-pure water recycling in fabrication facilities are examples of circularity strategies.

Nonetheless, one of the major challenges concerning the resource demand of AI-driven data centers is the lack of public information on resource usage or on infrastructure and operation details that enable the calculation of resource consumption. Given the absence of an international framework on the environmental cost of AI technologies, most of the available data on resource usage is self-reported by AI corporations, while public databases (e.g., the IEA and United Nations Environment Program (UNEP) focus on national-level reports instead of focusing on individual infrastructures, ignoring cross-resource and cumulative effects [31,154]. All of this makes it difficult for researchers and policymakers to adequately assess the real ecological impact of AI-driven data centers, and even when studies publicize resource demand within AI-driven data centers, a shortage of quantitative analysis with empirical substantiation in this regard persists.

There are also identified limitations about the metrics used to assess resource impact. One of these limitations lies in the fact that AI-driven data centers are integrated by multiple subsystems (cooling, power, workloads, environmental controls, etc.), making them highly complex enterprises. Consequently, the existing individual metrics cannot adequately capture the relation of all the different factors associated with resource usage, making holistic issues hard to identify. The lack of standardization of these metrics impedes them from being uniformly implemented worldwide, which is something that could facilitate resource usage comparison. For example, ISO environmental standards or science-based target frameworks could enhance comparability between firms; it is also important to take into account the complete supply chain of AI-driven data centers from extraction to disposal, rather than only considering their operational environmental impact [155].

## 7. Implications for Resource Management

The rapid growth of AI-powered data centers is transforming how natural resources are managed. However, the urgent need to rethink government policies and resource management strategies remains a contentious issue within the sector [156]. The central problem lies in the fact that individual assessments overlook the deep interdependence among these factors: water enables energy generation, energy drives water cycles, and minerals underpin both the infrastructure and future solutions for this system. This underscores the need to adopt new holistic and systemic approaches to resource allocation, linking these strategies to the principles of the circular economy in order to establish a strategy capable of mitigating the material and water footprint of digital infrastructure [157].

The adoption of circular economy strategies will not only help reduce environmental impacts but also strengthen operational resilience by reducing dependence on raw materials and improving material availability for the future expansion of the sector [158]. In addition, circular economy strategies will help reduce the generation of electronic waste, strengthen the resilience of supply chains, and contribute to the sustainability of AI infrastructure.

As previously mentioned, data centers are multidimensional in nature as they integrate, among other things, complex cooling, workloads, and environmental control subsystems. Therefore, there is a growing trend toward the development of more robust and composite indicators that will accurately reflect the trade-offs between various consumption factors. There is an urgent need to internationally standardize these metrics in order to facilitate performance measurement and cross-border comparisons to mitigate global supply disruptions.

One factor that is often overlooked, but which constitutes one of the biggest challenges facing this sector, is the spatiotemporal dimension of its impacts. Resource consumption by AI data centers is highly asymmetric and dynamic, as it fluctuates depending on workloads, weather conditions, energy source availability, and so on. Since management strategies cannot be based on global averages, it is essential for new systemic approaches to incorporate criteria for operational flexibility, which will allow for the distribution of workloads to locations with lower water stress.

The current optimization efforts are hampered by a lack of transparency and public data, which significantly limit future improvements. Therefore, it is essential to incorporate a quantitative environmental assessment to guide comprehensive research that will allow for the precise regulation of the externalities of AI, ensuring that technological progress does not jeopardize human well-being or the planet's resilience and that such measures should be implemented from the earliest stages of planning [159]. Sustainable data center management should not be limited to the use of individual resources, but it should integrate strategies that simultaneously optimize the use of water, energy, minerals, and so on. In this regard, organizations such as the European Commission promote a comprehensive approach based on energy efficiency, the circular economy, the use of waste heat, and the use of renewable energy, and focus not only on these aspects but also on expanding traditional environmental assessment criteria [160].

Taken together, these implications make it clear that the future management of AI-based data centers will need to evolve toward integrated approaches of resource management. Under the proposed approach, the sustainability of digital infrastructure will depend not only on technological development but also on the capacity of environmental and social regulatory bodies to ensure that the growth of AI occurs within the limits of resource availability and global sustainability goals, without hindering the development of the sector or adversely affecting other sectors in the process.

## 8. Conclusions

Artificial intelligence has become one of the main drivers of digital transformation; however, its rapid expansion has also increased the demand for energy, water, and other critical materials at a scale that challenges conventional approaches to infrastructure sustainability. This review has shown that evaluating AI-driven data centers exclusively through their electricity consumption only provides a partial understanding of their environmental implications. The growing dependence on water-intensive cooling technologies, the increasing extraction of critical minerals required for the advanced hardware, and the accumulation of electronic waste reveal that the sustainability of AI infrastructure is governed by a complex network of interdependent resource flows rather than by a single environmental dimension.

The analyzed evidence also shows that technological improvements alone are unlikely to guarantee sustainable long-term development. Although advances in hardware efficiency, cooling technologies, renewable energy integration, virtualization, and AI-based operational optimization have substantially improved resource efficiency, these gains coexist with a continuous increase in computational demand that may offset a significant fraction of

the expected environmental benefits. Future assessments must move beyond isolated efficiency metrics toward integrated evaluation frameworks capable of capturing energy, water, material consumption, and life cycle impacts simultaneously in order to provide a more realistic representation of the environmental footprint of AI infrastructure.

The unavoidable trade-offs among resource management strategies have been a recurring theme throughout the reviewed literature. Measures designed to reduce electricity consumption may increase water demand, while alternatives intended to minimize water use frequently require additional energy inputs. Likewise, the expansion of renewable energy generation reduces operational carbon emissions but simultaneously increases dependence on critical minerals whose extraction, processing, and geopolitical concentration introduce new sustainability challenges. These interactions demonstrate that resource optimization cannot be performed independently for each resource and instead requires a systems-level perspective capable of balancing competing environmental objectives within ecological constraints.

Another identified major limitation is the lack of transparency and methodological harmonization regarding resource reporting. Current environmental assessments rely heavily on self-reported corporate data while existing performance indicators frequently evaluate individual operational aspects without adequately representing the interactions among infrastructure subsystems or the complete life cycle of AI technologies. This fragmentation limits meaningful comparisons across facilities, complicates policy development, and hinders the establishment of internationally consistent sustainability benchmarks. Addressing these deficiencies will require standardized reporting methodologies, more comprehensive environmental indicators, and broader disclosure practices that improve the reliability of sustainability assessments.

The review further highlights that resource management in AI-driven data centers should evolve from a paradigm centered exclusively on efficiency toward a paradigm that incorporates sufficiency, circularity, and long-term resource resilience. Strategies such as extending the hardware lifetime, recovering critical materials from electronic waste, reusing water, utilizing waste heat, and geographically optimizing workloads illustrate that reducing environmental pressures depends not only on technological innovation but also on systemic resource management throughout the entire infrastructure life cycle.

Ultimately, AI sustainability should not be evaluated solely on its computational capabilities or on its economic value, but also on its capacity to operate within the environmental limits imposed by finite natural resources. As AI continues to expand across virtually every sector of society, ensuring that digital infrastructure develops in harmony with energy availability, water security, material supply, and circular resource management will become an essential prerequisite to achieve genuinely sustainable technological progress.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/resources15090115/s1>. Table S1: PRISMA-S Checklist for Resource Management Challenges in AI-Driven Data Centers: A Critical Review of Energy, Water, and Material Constraints. Table S2: Summary of the studies included in the systematic review, including research objectives, methodologies, and contributions to the review. Reference [161] is cited in the Supplementary Materials.

**Author Contributions:** Conceptualization, T.P.-P., D.K.Z.-V., C.R.-M. and J.M.P.-O.; methodology, T.P.-P., D.K.Z.-V., C.R.-M. and J.M.P.-O.; validation, T.P.-P., D.K.Z.-V., C.R.-M. and J.M.P.-O.; formal analysis, T.P.-P., D.K.Z.-V., C.R.-M. and J.M.P.-O.; investigation, T.P.-P., D.K.Z.-V., C.R.-M. and J.M.P.-O.; resources, C.R.-M. and J.M.P.-O.; data curation, T.P.-P., D.K.Z.-V., C.R.-M. and J.M.P.-O.; writing—original draft preparation, T.P.-P., D.K.Z.-V., C.R.-M. and J.M.P.-O.; writing—review and editing, T.P.-P., D.K.Z.-V., C.R.-M. and J.M.P.-O.; visualization, T.P.-P., D.K.Z.-V., C.R.-M. and J.M.P.-O.;

supervision, C.R.-M. and J.M.P.-O.; project administration, C.R.-M. and J.M.P.-O.; funding acquisition, C.R.-M. and J.M.P.-O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Secretaría de Ciencia, Humanidades, Tecnología e Innovación.

**Data Availability Statement:** All data generated or analyzed during this study are included in this published article.

**Acknowledgments:** The authors acknowledge the financial support from SECIHTI (Mexico) and CIC-UMSNH.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

|        |                                                                    |
|--------|--------------------------------------------------------------------|
| AI     | Artificial Intelligence                                            |
| APC    | Average Power Consumption                                          |
| CPE    | Compute Power Efficiency                                           |
| CPU    | Central Processing Unit                                            |
| CUE    | Carbon Usage Effectiveness                                         |
| DCeP   | Data Center Energy Productivity                                    |
| DCgE   | Data Center Green Efficiency                                       |
| DCPE   | Data Center Performance Efficiency                                 |
| DciE   | Data Center Infrastructure Efficiency                              |
| DH-UE  | Data Hall Utilization Efficiency                                   |
| DH-UR  | Data Hall Utilization Rate                                         |
| DWPE   | Data Center Workload Power Efficiency                              |
| EWR    | Energy Wastage Ratio                                               |
| GPU    | Graphics Processing Unit                                           |
| IEA    | International Energy Agency                                        |
| IEEE   | Institute of Electrical and Electronics Engineers                  |
| IRENA  | International Renewable Energy Agency                              |
| ISO    | International Organization for Standardization                     |
| IT     | Information Technology                                             |
| ITTE   | IT Equipment Energy Efficiency                                     |
| MGD    | Million Gallons per Day                                            |
| MIT    | Massachusetts Institute of Technology                              |
| OpenAI | OpenAI                                                             |
| Ops    | Computational Operations                                           |
| OSWE   | Operational System Workload Efficiency                             |
| PDE    | Power Delivery Efficiency                                          |
| PRISMA | Preferred Reporting Items for Systematic Reviews and Meta-Analyses |
| PUE    | Power Usage Effectiveness                                          |
| PV     | Photovoltaic                                                       |
| PpW    | Performance per Watt                                               |
| ScE    | Server Compute Efficiency                                          |
| UNEP   | United Nations Environment Programme                               |
| UPS    | Uninterruptible Power Supply                                       |
| WUE    | Water Usage Effectiveness                                          |

## References and Note

1. Ferriani, F.; Gazzani, A.G. Powering the Digital Economy: The Global Expansion of Data Centres and Its Energy Implications. *SSRN Electron. J.* **2026**, *1*, 1–28. [[CrossRef](#)]
2. Amado, A.B. Ethics and Public Values as a Guide for Action in the Face of the Expansion of Artificial Intelligence: Why Should Public Administrations Lead the Social Governance of Disruptive Technologies? *Claridades* **2024**, *16*, 105–138. [[CrossRef](#)]

3. Cai, S.; Gou, Z. Towards Energy-Efficient Data Centers: A Comprehensive Review of Passive and Active Cooling Strategies. *Energy Built Environ.* **2026**, *7*, 206–226. [[CrossRef](#)]
4. Dogru, A.K.; Keskin, B.B. AI in Operations Management: Applications, Challenges and Opportunities. *J. Data Inf. Manag.* **2020**, *2*, 67–74. [[CrossRef](#)]
5. Koomey, J.G. Worldwide Electricity Used in Data Centers. *Environ. Res. Lett.* **2008**, *3*, 034008. [[CrossRef](#)]
6. Andrae, A.S.G. New perspectives on internet electricity use in 2030. *Eng. Appl. Sci. Lett.* **2020**, *3*, 19–31. [[CrossRef](#)]
7. Duarte, A.H.F.; Santiago, N.L. Analysis of the Impact on Energy Demand Resulting from the Use of Artificial Intelligence. *Mex. Yearb. Glob. Aff.* **2025**, *3*, 209–230. [[CrossRef](#)]
8. Vafamehr, A.; Khodayar, M.E.; Manshadi, S.D.; Ahmad, I.; Lin, J. A Framework for Expansion Planning of Data Centers in Electricity and Data Networks Under Uncertainty. *IEEE Trans. Smart Grid* **2019**, *10*, 305–316. [[CrossRef](#)]
9. Lei, N.; Lu, J.; Shehabi, A.; Masanet, E. The Water Use of Data Center Workloads: A Review and Assessment of Key Determinants. *Resour. Conserv. Recycl.* **2025**, *219*, 108310. [[CrossRef](#)]
10. Lei, N.; Masanet, E. Climate- and Technology-Specific PUE and WUE Estimations for U.S. Data Centers Using a Hybrid Statistical and Thermodynamics-Based Approach. *Resour. Conserv. Recycl.* **2022**, *182*, 106323. [[CrossRef](#)]
11. Valero, A.; Valero, A.; Calvo, G.; Ortego, A. Material Bottlenecks in the Future Development of Green Technologies. *Renew. Sustain. Energy Rev.* **2018**, *93*, 178–200. [[CrossRef](#)]
12. Bizo, D.; Ascierto, R.; Lawrence, A.; Davis, J. Uptime Institute global data center survey 2021, *UII-511*.
13. Cai, S.; Gou, Z. A Comprehensive Analysis of Green Building Rating Systems for Data Centers. *Energy Build.* **2023**, *284*, 112874. [[CrossRef](#)]
14. Schneider, L. Supply Chain Considerations for AI Implementation—Critical Minerals, Data Centers, and the Environment. In *Artificial Intelligence for Sustainability and Global Security*; IFIP Advances in Information and Communication Technology; Springer: Cham, Switzerland, 2026; Volume 780, pp. 117–126. [[CrossRef](#)]
15. Andreu, J.F.; Morron Lingl, C.; Delgado, A.V.; Torralba, J.T.; Ríos Vera, L. Big Data on a Dead Planet: The Digital Transition's Neglected Environmental Impacts. *Bruss. Left Eur. Parliam.* **2022**, *2*, 1–20.
16. Graedel, T.E.; Reck, B.K.; Miatto, A. Alloy Information Helps Prioritize Material Criticality Lists. *Nat. Commun.* **2022**, *13*, 150. [[CrossRef](#)]
17. Guillaume, B.; Benjamin, D.; Vincent, C. Review of the Impact of IT on the Environment and Solution with a Detailed Assessment of the Associated Gray Literature. *Sustainability* **2022**, *14*, 2457. [[CrossRef](#)]
18. Kar, A.K.; Choudhary, S.K.; Singh, V.K. How Can Artificial Intelligence Impact Sustainability: A Systematic Literature Review. *J. Clean. Prod.* **2022**, *376*, 134120. [[CrossRef](#)]
19. Gopi, K.; Sharma, A.; Rani, M.R.J.; Kamath, K.P.; Manickam, T.; Thangam, D.; Ravindran, K.; Chavadi, C.; Pol, N. Strategies to Achieve Carbon Neutrality and Foster Sustainability in Data Centers. In *Computational Intelligence for Green Cloud Computing and Digital Waste Management*; IGI Global Scientific Publishing: Hershey, PA, USA, 2024; pp. 109–126. [[CrossRef](#)]
20. Smith, S.; Hubbard, A.; Newkirk, A.; Ganeshalingam, M.; Holecek, B.; Sartor, D.; Mills, M.; Shehabi, A. United States Data Center Energy Usage Report: 2025 Update. In *Energy Analysis Division*; Lawrence Berkeley National Laboratory: Berkeley, CA, USA, 2026. [[CrossRef](#)]
21. Sánchez, P.M.M.; Barrezueta, L.D.R. Centros de Datos Verdes en Ecuador: Una Estrategia para Disminuir la Emisión de CO<sub>2</sub> en los Centros de Datos Ecuatorianos. *Ser. Científica Univ. Cienc. Inform.* **2023**, *16*, 1–18.
22. Rodríguez, A.; Tarragó, J.C.P.; Zúñiga, K.M.; Llor, L.V.V. Evaluación Formativa de Los Procesos Cognitivos Con Paradigma Constructivista Mediante Mapa Cognitivo Difuso. *Ser. Científica Univ. Cienc. Inform.* **2021**, *14*, 130–142.
23. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*, n71. [[CrossRef](#)]
24. Markelius, A.; Wright, C.; Kuiper, J.; Delille, N.; Kuo, Y.-T. The Mechanisms of AI Hype and Its Planetary and Social Costs. *AI Ethics* **2024**, *4*, 727–742. [[CrossRef](#)]
25. Patterson, D.; Gonzalez, J.; Holzle, U.; Le, Q.; Liang, C.; Munguia, L.-M.; Rothchild, D.; So, D.R.; Texier, M.; Dean, J. The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink. *Computer* **2022**, *55*, 18–28. [[CrossRef](#)]
26. De Vries, A. The Growing Energy Footprint of Artificial Intelligence. *Joule* **2023**, *7*, 2191–2194. [[CrossRef](#)]
27. Han, J.; Han, K.; Han, T.; Wang, Y.; Han, Y.; Lin, J. Data-Driven Distributionally Robust Optimization of Low-Carbon Data Center Energy Systems Considering Multi-Task Response and Renewable Energy Uncertainty. *J. Build. Eng.* **2025**, *102*, 111937. [[CrossRef](#)]
28. Coskun, A.K.; Acun, F.; Clark, Q.; Hankendi, C.; Wilson, D.C. Data Center Demand Response for Sustainable Computing: Myth or Opportunity? In *Proceedings of the 2024 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Valencia, Spain, 25–27 March 2024. [[CrossRef](#)]

29. Shehabi, A.; Newkirk, A.; Smith, S.J.; Hubbard, A.; Lei, N.; Siddik, M.A.B.; Holecek, B.; Koomey, J.; Masanet, E.; Sartor, D. 2024 *United States Data Center Energy Usage Report*; Lawrence Berkeley National Laboratory Publications: Berkeley, CA, USA, 2024. [CrossRef]
30. Johnson, J.X.; Wade, C.; Blackhurst, M.; DeCarolis, J.F.; de Queiroz, A.R.; Posen, I.D.; Jaramillo, P. Power System Costs and Emissions from Data Center and Cryptocurrency Mining Expansion in the United States. *Environ. Res. Lett.* **2026**, *21*, 114007. [CrossRef]
31. International Energy Agency IEA-Data. Available online: <https://www.iea.org/data-and-statistics> (accessed on 10 July 2026).
32. Spencer, T.; Singh, S. What the Data Centre and AI Boom Could Mean for the Energy Sector. Available online: <https://www.iea.org/commentaries/what-the-data-centre-and-ai-boom-could-mean-for-the-energy-sector> (accessed on 6 June 2026).
33. Masanet, E.; Lei, N.; Koomey, J. To Better Understand AI's Growing Energy Use, Analysts Need a Data Revolution. *Joule* **2024**, *8*, 2427–2436. [CrossRef]
34. Noland, J.K.; Hjelmeland, M.; Korpas, M. Will Energy-Hungry AI Create a Baseload Power Demand Boom? *IEEE Access* **2024**, *12*, 110353–110360. [CrossRef]
35. Desislavov, R.; Martínez-Plumed, F.; Hernández-Orallo, J. Trends in AI Inference Energy Consumption: Beyond the Performance-vs-Parameter Laws of Deep Learning. *Sustain. Comput. Inform. Syst.* **2023**, *38*, 100857. [CrossRef]
36. Lin, L.; Wijayawardana, R.; Rao, V.; Nguyen, H.; Gribga, E.W.; Chien, A.A. Exploding AI Power Use: An Opportunity to Rethink Grid Planning and Management. In Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems (E-Energy '24), Singapore, Singapore, 4–7 June, 2024; Volume 1, pp. 434–441. [CrossRef]
37. Verdecchia, R.; Sallou, J.; Cruz, L. A Systematic Review of Green AI. *WIREs Data Min. Knowl. Discov.* **2023**, *13*, e1507. [CrossRef]
38. Argerich, M.F.; Patiño-Martínez, M. Measuring and Improving the Energy Efficiency of Large Language Models Inference. *IEEE Access* **2024**, *12*, 80194–80207. [CrossRef]
39. Newkirk, A.C.; Fernandez, J.; Koomey, J.; Latif, I.; Strubell, E.; Shehabi, A.; Samaras, C. Empirically-Calibrated H100 Node Power Models for Accurate AI Training Energy Estimation. *Environ. Res. Energy* **2025**, *2*, 045016. [CrossRef]
40. Electric Power Research Institute (EPRI). *Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption*; Electric Power Research Institute (EPRI): Palo Alto, CA, USA, 2024.
41. Flores-Martin, D.; Lemus-Prieto, F.; Rico-Gallego, J.A. PULSE: A Modular Framework for Predictive Energy Efficiency in Heterogeneous Data Centers. *SoftwareX* **2025**, *31*, 102313. [CrossRef]
42. Isazadeh, A.; Ziviani, D.; Claridge, D.E. Global Trends, Performance Metrics, and Energy Reduction Measures in Datacom Facilities. *Renew. Sustain. Energy Rev.* **2023**, *174*, 113149. [CrossRef]
43. Avgerinou, M.; Bertoldi, P.; Castellazzi, L. Trends in Data Centre Energy Consumption under the European Code of Conduct for Data Centre Energy Efficiency. *Energies* **2017**, *10*, 1470. [CrossRef]
44. Tariq, B. Next-Generation Energy-Efficient Data Centers: Technologies, Trends, and Ai-Driven Optimisation for Sustainable Computing. *Int. J. Innov. Stud.* **2025**, *9*, 1590–1607. [CrossRef]
45. Safari, A.; Sorouri, H.; Rahimi, A.; Oshnoei, A. A Systematic Review of Energy Efficiency Metrics for Optimizing Cloud Data Center Operations and Management. *Electronics* **2025**, *14*, 2214. [CrossRef]
46. Reddy, V.D.; Setz, B.; Rao, G.S.V.R.K.; Gangadharan, G.R.; Aiello, M. Metrics for Sustainable Data Centers. *IEEE Trans. Sustain. Comput.* **2017**, *2*, 290–303. [CrossRef]
47. Belady, C.L.; Malone, C.G. Metrics and an Infrastructure Model to Evaluate Data Center Efficiency. *Proc. ASME InterPack Conf. IPACK* **2007**, *1*, 751–755. [CrossRef]
48. Kumar, R.; Khatri, S.K.; Diván, M.J. Efficiency Measurement of Data Centers: An Elucidative Review. *J. Discret. Math. Sci. Cryptogr.* **2020**, *23*, 221–236. [CrossRef]
49. Shaikh, A.; Uddin, M.; Elmagzoub, M.A.; Alghamdi, A. PEMC: Power Efficiency Measurement Calculator to Compute Power Efficiency and CO<sub>2</sub> emissions in Cloud Data Centers. *IEEE Access* **2020**, *8*, 195216–195228. [CrossRef]
50. Hanumaiah, V.; Desai, D.; Gaudette, B.; Wu, C.-J.; Vruthula, S. STEAM: A Smart Temperature and Energy Aware Multicore Controller. *ACM Trans. Embed. Comput. Syst.* **2014**, *13*, 151. [CrossRef]
51. Fargo, F.; Tunc, C.; Al-Nashif, Y.; Hariri, S. Autonomic Performance-per-Watt Management (APM) of Cloud Resources and Services. In CAC '13: ACM Cloud and Autonomic Computing Conference; Association for Computing Machinery: New York, NY, USA, 2013. [CrossRef]
52. Khargharia, B.; Hariri, S.; Yousif, M.S. An Adaptive Interleaving Technique for Memory Performance-per-Watt Management. *IEEE Trans. Parallel Distrib. Syst.* **2009**, *20*, 1011–1022. [CrossRef]
53. Denisenko, B.; Tyanutov, M.; Nikiforov, I.; Ustinov, S. Algorithm for Calculating TCO and SCE Metrics to Assess the Efficiency of Using a Data Center. *Proc. SPIE* **2023**, *12564*, 7–13. [CrossRef]
54. Li, Z.; Yang, Y. RRect: A Novel Server-Centric Data Center Network with High Power Efficiency and Availability. *IEEE Trans. Cloud Comput.* **2020**, *8*, 914–927. [CrossRef]

55. Swain, S.R.; Parashar, A.; Singh, A.K.; Lee, C.N. An Intelligent Virtual Machine Allocation Optimization Model for Energy-Efficient and Reliable Cloud Environment. *J. Supercomput.* **2025**, *81*, 237. [\[CrossRef\]](#)
56. Wilde, T.; Auweter, A.; Patterson, M.K.; Shoukourian, H.; Huber, H.; Bode, A.; Labrenz, D.; Cavazzoni, C. DWPE, a New Data Center Energy-Efficiency Metric Bridging the Gap between Infrastructure and Workload. In *Proceedings of the 2014 International Conference on High Performance Computing & Simulation (HPCS)*; IEEE: New York, NY, USA, 2014; pp. 893–901. [\[CrossRef\]](#)
57. Lee, S.; Lee, D.; Choi, P.; Park, D. Accuracy–Power Controllable LiDAR Sensor System with 3D Object Recognition for Autonomous Vehicle. *Sensors* **2020**, *20*, 5706. [\[CrossRef\]](#)
58. Aravanis, A.I.; Voulkidis, A.; Salom, J.; Townley, J.; Georgiadou, V.; Oleksiak, A.; Porto, M.R.; Roudet, F.; Zahariadis, T. Metrics for Assessing Flexibility and Sustainability of Next Generation Data Centers. In *Proceedings of the 2015 IEEE Globecom Workshops (GC Wkshps)*; IEEE: New York, NY, USA, 2015. [\[CrossRef\]](#)
59. Grishina, A.; Chinnici, M.; De Chiara, D.; Rondeau, E.; Kor, A.L. Energy-Oriented Analysis of HPC Cluster Queues: Emerging Metrics for Sustainable Data Center. In *Applied Physics, System Science and Computers III*; Lecture Notes in Electrical Engineering; Springer: Cham, Switzerland, 2019; Volume 574, pp. 286–300. [\[CrossRef\]](#)
60. Grishina, A.; Chinnici, M.; Kor, A.-L.; Rondeau, E.; Georges, J.-P.; De Chiara, D. Data Center for Smart Cities: Energy and Sustainability Issue. In *Big Data Platforms and Applications*; Computer Communications and Networks; Springer: Cham, Switzerland, 2021; pp. 1–36. [\[CrossRef\]](#)
61. Wesolowski, L.; Acun, B.; Andrei, V.; Aziz, A.; Dankel, G.; Gregg, C.; Meng, X.; Meurillon, C.; Sheahan, D.; Tian, L.; et al. Datacenter-Scale Analysis and Optimization of GPU Machine Learning Workloads. *IEEE Micro* **2021**, *41*, 101–112. [\[CrossRef\]](#)
62. Chen, X.; Wang, X.; Colacelli, A.; Lee, M.; Xie, L. Electricity Demand and Grid Impacts of AI Data Centers: Challenges and Prospects. *arXiv* **2025**. [\[CrossRef\]](#)
63. Martinez, D.; Sickinger, D.; Andersen, A. “How-to” Guide for Transitioning from Air to Liquid-Cooled High Performance Computing Systems. In *Proceedings of the 2024 IEEE International Conference on Cluster Computing Workshops (CLUSTER Workshops)*; IEEE: New York, NY, USA, 2024; pp. 9–18. [\[CrossRef\]](#)
64. Patterson, M.K.; Poole, S.W.; Hsu, C.-H.; Maxwell, D.; Tschudi, W.; Coles, H.; Martinez, D.J.; Bates, N. TUE, a New Energy-Efficiency Metric Applied at ORNL’s Jaguar. In *Supercomputing*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2013; Volume 7905, pp. 372–382. [\[CrossRef\]](#)
65. Hernandez, L.; Hernandez, H.; Orozco, M.; Piñeres, G.; Garcia-Guilianny, J. Measurement of Energy Efficiency Metrics of Data Centers. Case Study: Higher Education Institution of Barranquilla. In *Software Engineering Perspectives in Intelligent Systems*; Advances in Intelligent Systems and Computing; Springer: Cham, Switzerland, 2020; Volume 1295, pp. 23–35. [\[CrossRef\]](#)
66. Almhdi, E.; Miskeen, G. Power and Carbon Footprint Evaluation and Optimization in Transitioning Data Centres. *Wadi Alshatti Univ. J. Pure Appl. Sci.* **2025**, *3*, 221–229. [\[CrossRef\]](#)
67. Cao, Z.; Zhou, X.; Hu, H.; Wang, Z.; Wen, Y. Toward a Systematic Survey for Carbon Neutral Data Centers. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 895–936. [\[CrossRef\]](#)
68. Grishina, A.; Chinnici, M.; De Chiara, D.; Guarnieri, G.; Kor, A.L.; Rondeau, E.; Georges, J.P. DC Energy Data Measurement and Analysis for Productivity and Waste Energy Assessment. In *Proceedings of the 2018 IEEE International Conference on Computational Science and Engineering (CSE)*; IEEE: New York, NY, USA, 2018; pp. 1–11. [\[CrossRef\]](#)
69. Isaev, E.A.E.; Kornilov, V.V.; Grigori’ev, A.A. Data Center Efficiency Model: A New Approach and the Role of Artificial Intelligence. *Math. Biol. Bioinform.* **2023**, *18*, 215–227. [\[CrossRef\]](#)
70. Anghel, I.; Pop, C.B.; Cioara, T.; Salomie, I.; Vartic, I. A Swarm-Inspired Technique for Self-Organizing and Consolidating Data Centre Servers. *Scalable Comput. Pract. Exp.* **2013**, *14*, 69–82. [\[CrossRef\]](#)
71. Procaccianti, G.; Routsis, A. Energy Efficiency and Power Measurements: An Industrial Survey. In *Proceedings of the ICT for Sustainability 2016*; Atlantis Press: Dordrecht, The Netherlands, 2016; pp. 69–78. [\[CrossRef\]](#)
72. Anghel, I.; Cioara, T.; Salomie, I. Context-Aware and Reinforcement Learning-Based Load Balancing System for Green Clouds. In *Resource Management for Big Data Platforms*. *Computer Communications and Networks*; Springer: Cham, Switzerland, 2016; pp. 129–144. [\[CrossRef\]](#)
73. Fiandrino, C.; Kliazovich, D.; Bouvry, P.; Zomaya, A.Y. Performance and Energy Efficiency Metrics for Communication Systems of Cloud Computing Data Centers. *IEEE Trans. Cloud Comput.* **2017**, *5*, 738–750. [\[CrossRef\]](#)
74. Shally; Sharma, S.K.; Kumar, S. Measuring Energy Efficiency of Cloud Datacenters. *Int. J. Recent Technol. Eng.* **2019**, *8*, 5428–5433. [\[CrossRef\]](#)
75. Levy, M.; Raviv, D. A Novel Framework for Data Center Metrics Using a Multidimensional Approach A Novel Framework for Data Center Metrics Using a Multidimensional Approach. In *Proceedings of the LACCEI International Multi-Conference for Engineering, Education and Technology*, Boca Raton, FL, USA, 19–21 July 2017; pp. 19–21. [\[CrossRef\]](#)
76. Rahaman, A.; Noor, K.N.; Abir, T.A.; Rana, S.; Ali, M. Design and Analysis of Sustainable Green Data Center with Hybrid Energy Sources. *J. Power Energy Eng.* **2021**, *9*, 76–88. [\[CrossRef\]](#)

77. Chauhan, M.A. Technological Advancements and Environmental Impacts of Air Source Heat Pumps for Sustainable Cooling and Waste Heat Recovery in California Data Centres. *Eur. J. Adv. Eng. Technol.* **2025**, *2025*, 22–28. [CrossRef]
78. Ebrahimi, K.; Jones, G.F.; Fleischer, A.S. A Review of Data Center Cooling Technology, Operating Conditions and the Corresponding Low-Grade Waste Heat Recovery Opportunities. *Renew. Sustain. Energy Rev.* **2014**, *31*, 622–638. [CrossRef]
79. Capozzoli, A.; Primiceri, G. Cooling Systems in Data Centers: State of Art and Emerging Technologies. *Energy Procedia* **2015**, *83*, 484–493. [CrossRef]
80. Mavisclara, O.A.; Oshobugie, I.I.; Olufunmi, A.T.; Bolaji, A.A.; Olawale, A.O.; Anulika, N.C.; Samuel, O.K. The AI-Driven Energy Surge: A Comprehensive Review of Sustainable Power Solutions for Data Centers. *Glob. J. Eng. Technol. Adv.* **2025**, *24*, 328–344. [CrossRef]
81. Al Kez, D.; Foley, A.M.; Laverty, D.; Del Rio, D.F.; Sovacool, B. Exploring the Sustainability Challenges Facing Digitalization and Internet Data Centers. *J. Clean. Prod.* **2022**, *371*, 133633. [CrossRef]
82. Chidolue, O.; Ohenhen, P.E.; Umoh, A.A.; Ngozichukwu, B.; Fafure, V.A.; Ibekwe, K.I. Green Data Centers: Sustainable Practices for Energy-Efficient IT Infrastructure. *Eng. Sci. Technol. J.* **2024**, *5*, 99–114. [CrossRef]
83. Ghazanfari-Rad, S.; Ebneyousef, S. A Survey of Renewable Energy Approaches in Cloud Data Centers. In *Proceedings of the 2023 8th International Conference on Technology and Energy Management (ICTEM)*; IEEE: New York, NY, USA, 2023; pp. 1–7. [CrossRef]
84. Xu, Y.-P.; Liu, R.-H.; Tang, L.-Y.; Wu, H.; She, C. Risk-Averse Multi-Objective Optimization of Multi-Energy Microgrids Integrated with Power-to-Hydrogen Technology, Electric Vehicles and Data Center under a Hybrid Robust-Stochastic Technique. *Sustain. Cities Soc.* **2022**, *79*, 103699. [CrossRef]
85. Jones, E.C., Jr.; Jones, E.C., Sr. Megawatts to Zettaflops: A Techno-Economic Framework for Grid-Tied Behind-the-Meter Architectures in AI Data Centers. *Electricity* **2026**, *7*, 43. [CrossRef]
86. Hua, H.; Li, Y.; Wang, T.; Dong, N.; Li, W.; Cao, J. Edge Computing with Artificial Intelligence: A Machine Learning Perspective. *ACM Comput. Surv.* **2023**, *55*, 1–35. [CrossRef]
87. Zhou, R.; Zhang, T.; Catano, J.; Wen, J.T.; Michna, G.J.; Peles, Y.; Jensen, M.K. The Steady-State Modeling and Optimization of a Refrigeration System for High Heat Flux Removal. *Appl. Therm. Eng.* **2010**, *30*, 2347–2356. [CrossRef]
88. Kong, R.; Zhang, H.; Tang, M.; Zou, H.; Tian, C.; Ding, T. Enhancing Data Center Cooling Efficiency and Ability: A Comprehensive Review of Direct Liquid Cooling Technologies. *Energy* **2024**, *308*, 132846. [CrossRef]
89. Lu, T.; Lü, X.; Remes, M.; Viljanen, M. Investigation of Air Management and Energy Performance in a Data Center in Finland: Case Study. *Energy Build.* **2011**, *43*, 3360–3372. [CrossRef]
90. Cho, J.; Woo, J. Development and Experimental Study of an Independent Row-Based Cooling System for Improving Thermal Performance of a Data Center. *Appl. Therm. Eng.* **2020**, *169*, 114857. [CrossRef]
91. Heydari, A.; Gharaibeh, A.R.; Tradat, M.; Soud, Q.; Manaserh, Y.; Radmard, V.; Eslami, B.; Rodriguez, J.; Sammakia, B. Experimental Evaluation of Direct-to-Chip Cold Plate Liquid Cooling for High-Heat-Density Data Centers. *Appl. Therm. Eng.* **2024**, *239*, 122122. [CrossRef]
92. Kuncoro, I.W.; Pambudi, N.A.; Biddinika, M.K.; Widiastuti, I.; Hijriawan, M.; Wibowo, K.M. Immersion Cooling as the next Technology for Data Center Cooling: A Review. *J. Phys. Conf. Ser.* **2019**, *1402*, 44057. [CrossRef]
93. Joshi, Y.; Wei, X.; Woodruff, G.W. Micro and Meso Scale Compact Heat Exchangers in Electronics Thermal Management-A Review. In *Proceedings of the Fifth International Conference on Enhanced, Compact and Ultra-Compact Heat Exchangers: Science, Engineering and Technology*; Shah, R., Ishizuka, M., Rudy, T., Wadekar, V., Eds.; Engineering Conferences International: Hoboken, NJ, USA, 2005; pp. 1–18.
94. Xu, M.; Wang, S.; Hua, X.; Ke, C.; Yu, G.; Yang, Z.; Wen, H. Advanced Chip-Level Thermal Management Technologies for High-Power Integrated Processors: A Review. *Energies* **2026**, *19*, 3304. [CrossRef]
95. NVIDIA Corporation. Unlocking Real-Time Trillion-Parameter Models. Available online: <https://www.nvidia.com/en-us/data-center/gb200-nv172/> (accessed on 25 August 2026).
96. Shrigondekar, H.; Lin, Y.-C.; Wang, C.-C. Investigations on Performance of Single-Phase Immersion Cooling System. *Int. J. Heat Mass Transf.* **2023**, *206*, 123961. [CrossRef]
97. Carlini, S. AI Factory Cooling vs Cloud Data Centers: Why Liquid Cooling Is Essential for High-Density AI Workloads. Available online: <https://blog.se.com/datacenter/2026/06/12/ai-factory-cooling-vs-cloud-data-centers-why-liquid-cooling-is-essential-for-high-density-ai-workloads/> (accessed on 25 August 2026).
98. Vertenten, J. Scaling Up: Liquid Cooling for Higher Density Racks. Available online: <https://www.jabil.com/blog/liquid-cooling-higher-density-racks.html> (accessed on 25 August 2026).
99. Chen, H.; Li, D. Current Status and Challenges for Liquid-Cooled Data Centers. *Front. Energy Res.* **2022**, *10*, 952680. [CrossRef]
100. Yu, Y.; Wang, L. Solid Sorption Heat Pipe Coupled with Direct Air Cooling Technology for Thermal Control of Rack Level in Internet Data Centers: Design and Numerical Simulation. *Int. J. Heat Mass Transf.* **2019**, *145*, 118714. [CrossRef]
101. Kandasamy, R.; Ho, J.Y.; Liu, P.; Wong, T.N.; Toh, K.C.; Chua, S.J. Two-Phase Spray Cooling for High Ambient Temperature Data Centers: Evaluation of System Performance. *Appl. Energy* **2022**, *305*, 117816. [CrossRef]

102. Ding, T.; Chen, X.; Cao, H.; He, Z.; Wang, J.; Li, Z. Principles of Loop Thermosyphon and Its Application in Data Center Cooling Systems: A Review. *Renew. Sustain. Energy Rev.* **2021**, *150*, 111389. [CrossRef]
103. Xu, S.; Zhang, H.; Wang, Z. Thermal Management and Energy Consumption in Air, Liquid, and Free Cooling Systems for Data Centers: A Review. *Energies* **2023**, *16*, 1279. [CrossRef]
104. Abbas, A.M.; Huzayyin, A.S.; Mouneer, T.A.; Nada, S.A. Effect of Data Center Servers' Power Density on the Decision of Using in-Row Cooling or Perimeter Cooling. *Alex. Eng. J.* **2021**, *60*, 3855–3867. [CrossRef]
105. Moazamigoodarzi, H.; Tsai, P.J.; Pal, S.; Ghosh, S.; Puri, I.K. Influence of Cooling Architecture on Data Center Power Consumption. *Energy* **2019**, *183*, 525–535. [CrossRef]
106. Li, X.; Li, M.; Zhang, Y.; Han, Z.; Wang, S. Rack-Level Cooling Technologies for Data Centers—A Comprehensive Review. *J. Build. Eng.* **2024**, *90*, 109535. [CrossRef]
107. Wang, Q.; Ozguc, S.; Bonner, R.W. System-Level Thermal Analysis of Two-Phase Direct-to-Chip Cooling for Data Centers with Chilled Facility Water. *J. Electron. Packag.* **2026**, *148*, 40904. [CrossRef]
108. Li, Z.; Luo, H.; Jiang, Y.; Liu, H.; Xu, L.; Cao, K.; Wu, H.; Gao, P.; Liu, H. Comprehensive Review and Future Prospects on Chip-Scale Thermal Management: Core of Data Center's Thermal Management. *Appl. Therm. Eng.* **2024**, *251*, 123612. [CrossRef]
109. Rahma, A.G.; Abadassi, F.; Akyol, J.; Marouf, A.; François, P.; Lawniczak, F.; Ghenaïm, A.; Hoarau, Y.; Bongs, C.; Kauffeld, M.; et al. Two-Phase Thermosiphon Cooling in Data Centers: A Comprehensive Review of Refrigerants for Practical Selection. *J. Eng. Appl. Sci.* **2025**, *72*, 254. [CrossRef]
110. Zhang, H.; Shao, S.; Tian, C.; Zhang, K. A Review on Thermosiphon and Its Integrated System with Vapor Compression for Free Cooling of Data Centers. *Renew. Sustain. Energy Rev.* **2018**, *81*, 789–798. [CrossRef]
111. Ding, B.; Xia, F.; He, L.; Huang, Y. A Novel Concept for Performance Enhancement of Immersion-Cooled Data Center Servers. *Int. Commun. Heat Mass Transf.* **2024**, *159*, 107980. [CrossRef]
112. Haghshenas, K.; Setz, B.; Bloesch, Y.; Aiello, M. Enough Hot Air: The Role of Immersion Cooling. *Energy Inform.* **2023**, *6*, 14. [CrossRef]
113. Wu, Z.; Zhang, G.; Lu, S.; Leng, P.; Yu, Y.; Deng, J.; Huang, W. A Comprehensive Review of Cold Plate Liquid Cooling Technology for Data Centers. *Chem. Eng. Sci.* **2025**, *310*, 121525. [CrossRef]
114. Li, J.; Luo, X.; Wang, M.; Chen, C.; Yang, C.; Zhang, Z.; Xin, G. Study of Flow and Heat Transfer Characteristics of Tandem Cold Plates for Data Center Cooling. *Case Stud. Therm. Eng.* **2025**, *73*, 106590. [CrossRef]
115. Kim, M.-H.; Ham, S.-W.; Park, J.-S.; Jeong, J.-W. Impact of Integrated Hot Water Cooling and Desiccant-Assisted Evaporative Cooling Systems on Energy Savings in a Data Center. *Energy* **2014**, *78*, 384–396. [CrossRef]
116. Han, Z.; Sun, X.; Wei, H.; Ji, Q.; Xue, D. Energy Saving Analysis of Evaporative Cooling Composite Air Conditioning System for Data Centers. *Appl. Therm. Eng.* **2021**, *186*, 116506. [CrossRef]
117. Kim, J.H.; Shin, D.U.; Kim, H. Data Center Energy Evaluation Tool Development and Analysis of Power Usage Effectiveness with Different Economizer Types in Various Climate Zones. *Buildings* **2024**, *14*, 299. [CrossRef]
118. de Vries-Gao, A. The Carbon and Water Footprints of Data Centers and What This Could Mean for Artificial Intelligence. *Patterns* **2026**, *7*, 101430. [CrossRef] [PubMed]
119. International Energy Agency (IEA). *Energy and AI*; International Energy Agency: Paris, France, 2015. Available online: <https://www.iea.org/reports/energy-and-ai> (accessed on 19 July 2026).
120. Jiang, F.; Duan, C.; Chen, B. Facility-Level Energy-Driven Water Footprint and Scarcity Implications of Chinese Data Centers: A Bottom-Up Analysis and Scenario-Based Projection. *Appl. Energy* **2025**, *399*, 126522. [CrossRef]
121. Han, Y.; Li, P.; Wierman, A.; Ren, S. Small Bottle, Big Pipe: Quantifying and Addressing the Impact of Data Centers on Public Water Systems. *arXiv* **2026**, arXiv:2603.02705. [CrossRef]
122. Siddik, M.A.B.; Shehabi, A.; Marston, L. The Environmental Footprint of Data Centers in the United States. *Environ. Res. Lett.* **2021**, *16*, 064017. [CrossRef]
123. Parker, J. Hotter Than a Hot Tub: The 45 °C Breakthrough to Cool AI's Biggest Machines | NVIDIA Blog. Available online: <https://blogs.nvidia.com/blog/liquid-cooling-ai-factories/> (accessed on 25 August 2026).
124. Stacciarini, J.H.S.; Gonçalves, R.J.D.A.F. Data Centers, Critical Minerals, Energy, and Geopolitics: The Foundations of Artificial Intelligence. *Soc. Nat.* **2025**, *37*, e77215. [CrossRef]
125. IRENA; CPI. Global Landscape of Renewable Energy Finance 2023. In Proceedings of the International Renewable Energy Agency (IRENA) Assembly, Abu Dhabi, United Arab, 10–12 January 2023; p. 132.
126. Hosseini, M.; Gao, P.; Vivas-Valencia, C. A Social-Environmental Impact Perspective of Generative Artificial Intelligence. *Environ. Sci. Ecotechnol.* **2025**, *23*, 100520. [CrossRef]
127. Joseph, R. Environmental Issues in Mining: A Comprehensive Review of Challenges and Strategies for Mitigation and Rehabilitation. *Min. Rev.* **2025**, *31*, 51–71. [CrossRef]

128. Emekwisia, C.C.; Ogunsiji, T.O.; Bauna, F.K.; Otsabomhe, G.O.; Akinbamilowo, O.O.; Olagunju, A.R.; Akagu, E.E. Comparative Analysis of Environmental Impacts Associated with Surface and Underground Mining Operations. *S. Asian Res. J. Eng. Technol.* **2025**, *7*, 121–125. [[CrossRef](#)]
129. Liu, F.; Wang, B. Research Progress of Critical Metals for Artificial Intelligence Chips Based on Nonlinear Properties and Material Innovation. *Int. J. High Speed Electron. Syst.* **2025**, *35*, 2540891. [[CrossRef](#)]
130. Bang, M.; Cho, S.-J.; Jin, T.; Park, J. Assessing Supply Chain Risks of Critical Minerals in the Semiconductor Industry of South Korea. *J. Korean Soc. Miner. Energy Resour. Eng.* **2025**, *62*, 283–301. [[CrossRef](#)]
131. Valdivia, A. The Supply Chain Capitalism of AI: A Call to (Re)Think Algorithmic Harms and Resistance through Environmental Lens. *Inf. Commun. Soc.* **2025**, *28*, 2118–2134. [[CrossRef](#)]
132. Amoah, M.; Brown, M.; Simon, A.; Bazilian, M.; Matisek, J. Mineral Demand from AI Data Centers: Infrastructure Intensity, Processing Bottlenecks, and Supply Competition. *Resour. Policy* **2026**, *119*, 105970. [[CrossRef](#)]
133. Wang, S.; Song, X.; Xu, J.; Wang, J.; Yu, L. Flexible Silicon for High-Performance Photovoltaics, Photodetectors and Bio-Interfaced Electronics. *Mater. Horiz.* **2025**, *12*, 1106–1132. [[CrossRef](#)]
134. Petersen, K.E. Silicon as a Mechanical Material. *Proc. IEEE* **1982**, *70*, 420–457. [[CrossRef](#)]
135. Li, Z.; Han, X.; Guo, X.; Ma, L.; Sun, J.; Wen, Y.; Guo, Y. Gallium-Based Liquid Metals: From Properties to Applications. *Nanomaterials* **2026**, *16*, 471. [[CrossRef](#)]
136. Zhuang, J.; Zhu, M.; Zhang, H.; Wang, W.; Lu, X. Gallium-Based Semiconductor Integrated Circuits: Past, Present, and Future for Emerging Optoelectronic Devices. *Adv. Funct. Mater.* **2026**, *36*, e27869. [[CrossRef](#)]
137. Sidhu, A. Critical Minerals: Global Aluminum Supply & Demand. *SSRN Electron. J.* **2023**, *1*, 1–3. [[CrossRef](#)]
138. Stojanovic, B.; Bukvic, M.; Epler, I. Application of Aluminum and Aluminum Alloys in Engineering. *Appl. Eng. Lett.* **2018**, *3*, 52–62. [[CrossRef](#)]
139. Frost, K.; Sousa, I.; Larson, J.; Jin, H.; Hua, I. Environmental Impacts of a Circular Recovery Process for Hard Disk Drive Rare Earth Magnets. *Resour. Conserv. Recycl.* **2021**, *173*, 105694. [[CrossRef](#)]
140. Zhao, T.Y.; Li, W.L.; Kelebek, S.; Choi, Y.; Wu, C.Q.; Zhang, W.J.; Wang, C.Y.; Zhao, Z.W.; Sadri, F. A Comprehensive Review on Rare Earth Elements: Resources, Technologies, Applications, and Prospects. *Rare Met.* **2025**, *44*, 7011–7040. [[CrossRef](#)]
141. Balam, V. Rare Earth Elements: A Review of Applications, Occurrence, Exploration, Analysis, Recycling, and Environmental Impact. *Geosci. Front.* **2019**, *10*, 1285–1303. [[CrossRef](#)]
142. Stacciarini, J.H.S.; Gonçalves, R.J.d.A.F. Global Geography of the Energy Transition and Mineral Extraction. *SocArXiv* **2025**.
143. Berahab, R. *The Energy Economics of Artificial Intelligence in a Fractured Global System*; Research Papers and Policy Papers on Commodities and Energy; Policy Center for the New South: Salé, Morocco, 2025.
144. Wang, P.; Zhang, L.-Y.; Tzachor, A.; Chen, W.-Q. E-Waste Challenges of Generative Artificial Intelligence. *Nat. Comput. Sci.* **2024**, *4*, 818–823. [[CrossRef](#)]
145. Glöser-Chahoud, S.; Schultmann, F. Potential Contribution of Secondary Materials to Overall Supply—The Example of the European Cobalt Cycle. *Mater. Sci. Forum* **2019**, *959*, 11–21. [[CrossRef](#)]
146. Hu, H.; Zeng, X.; Zheng, K.; Zeng, Z.; Dai, C.; Huo, X. Risk Assessment and Partitioning Behavior of PFASs in Environmental Matrices from an E-Waste Recycling Area. *Sci. Total Environ.* **2023**, *905*, 167707. [[CrossRef](#)]
147. Hossain, R.; Sahajwalla, V. Current Recycling Innovations to Utilize E-Waste in Sustainable Green Metal Manufacturing. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2024**, *382*, 20230239. [[CrossRef](#)]
148. Schreiber, A.; Marx, J.; Zapp, P.; Hake, J.-F.; Voßenkaul, D.; Friedrich, B. Environmental Impacts of Rare Earth Mining and Separation Based on Eudialyte: A New European Way. *Resources* **2016**, *5*, 32. [[CrossRef](#)]
149. Vishwakarma, S.; Kumar, V.; Arya, S.; Tembhare, M.; Rahul; Dutta, D.; Kumar, S. E-Waste in Information and Communication Technology Sector: Existing Scenario, Management Schemes and Initiatives. *Environ. Technol. Innov.* **2022**, *27*, 102797. [[CrossRef](#)]
150. D’amore, G.; Di Vaio, A.; Balsalobre-Lorente, D.; Boccia, F. Artificial Intelligence in the Water–Energy–Food Model: A Holistic Approach towards Sustainable Development Goals. *Sustainability* **2022**, *14*, 867. [[CrossRef](#)]
151. Hlabisa, S. The Ecology of Artificial Intelligence: Energy, Water, Materials, and Land Limits of Digital Systems. *Carbon Neutr. Syst.* **2025**, *1*, 19. [[CrossRef](#)]
152. Vergallo, R.; Cruz, L.; Errico, A.; Mainetti, L. On the Effectiveness of the ‘Follow-the-Sun’ Strategy in Mitigating the Carbon Footprint of AI in Cloud Instances. *arXiv* **2025**, arXiv:2506.10990. [[CrossRef](#)]
153. Xiao, T.; Nerini, F.F.; Matthews, H.D.; Tavoni, M.; You, F. Environmental Impact and Net-Zero Pathways for Sustainable Artificial Intelligence Servers in the USA. *Nat. Sustain.* **2025**, *8*, 1541–1553. [[CrossRef](#)]
154. United Nations Environmental Programme UNEP—Publications. Available online: <https://www.unep.org/> (accessed on 10 July 2026).
155. Kaack, L.H.; Donti, P.L.; Strubell, E.; Kamiya, G.; Creutzig, F.; Rolnick, D. Aligning Artificial Intelligence with Climate Change Mitigation. *Nat. Clim. Change* **2022**, *12*, 518–527. [[CrossRef](#)]

156. Naser, A.; Gastón, C. El Gobierno Electrónico en la Gestión Pública. CEPAL 2014. Available online: <https://hdl.handle.net/11362/7330> (accessed on 8 July 2026).
157. Valencia, R. del DANE Sistema de Indicadores Ambientales, de Sostenibilidad y Economía Circular—SIEC. Available online: <https://www.dane.gov.co/index.php/estadisticas-por-tema/ambientales/economia-circular/sistema-de-consulta-de-informacion> (accessed on 9 July 2026).
158. Google Centros de Datos. Economía Circular—Google Sostenibilidad. Available online: [https://sustainability.google/intl/es\\_es/stories/circular-economy/](https://sustainability.google/intl/es_es/stories/circular-economy/) (accessed on 9 July 2026).
159. Castro, K.; Edwar, S. Buenas Prácticas para la Descarbonización de Edificios de Centros de Datos Universitat Politècnica de Catalunya 2024. Available online: <https://hdl.handle.net/2117/414011> (accessed on 9 July 2026).
160. Comisión Europea. Nube Verde y Centros de Datos Verdes | Configurar El Futuro Digital de Europa. Available online: <https://digital-strategy.ec.europa.eu/es/policies/green-cloud> (accessed on 9 July 2026).
161. Rethlefsen, M.L.; Kirtley, S.; Waffenschmidt, S.; Ayala, A.P.; Moher, D.; Page, M.J.; Koffel, J.B. PRISMA-S: An extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Syst. Rev.* **2021**, *10*, 39. <https://doi.org/10.1186/s13643-020-01542-z>.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.