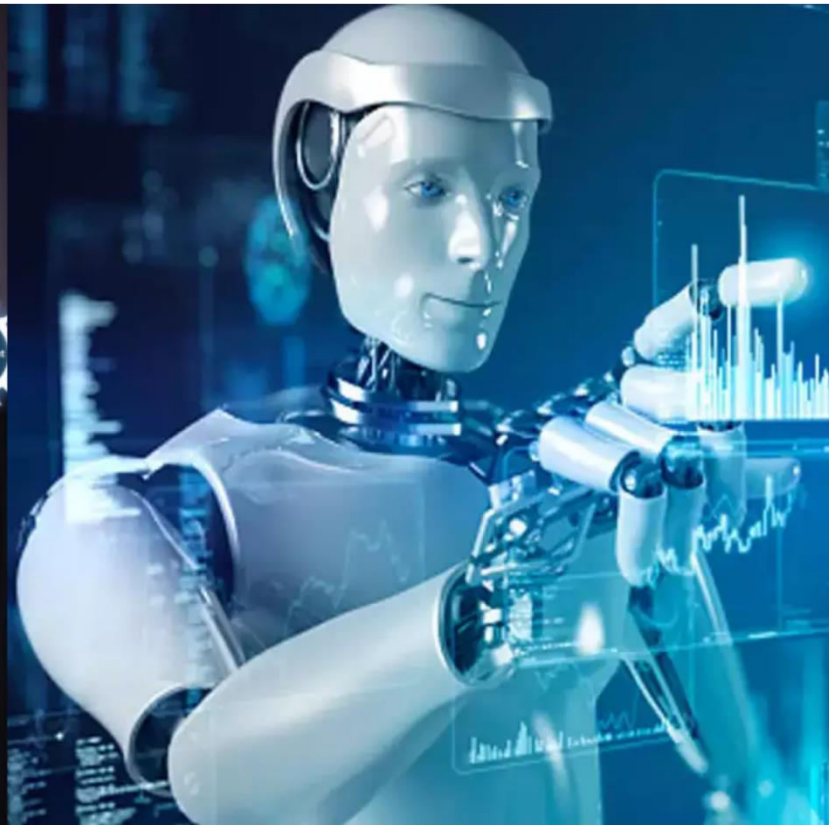
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 2, February 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402015|

Power Consumption and Heat Dissipation in AI Data Centers: A Comparative Analysis

Krishna Chaitanya Sunkara, Krishnaiah Narukulla

Cloud Infrastructure Engineering, Oracle Corporation, Raleigh, USA

ORCID:0009-0009-6159-4280

Core Platform Engineering, Roku, San Jose, USA

ORCID:0009-0006-5602-1021

ABSTRACT: The increasing computational demands of artificial intelligence (AI) workloads have significantly escalated energy consumption in data centers. AI-driven applications, including deep learning, natural language processing, and autonomous systems, require substantial computing power, primarily provided by Graphics Processing Units. These GPUs, while enhancing computational efficiency, contribute to significant power consumption and heat generation, necessitating advanced cooling strategies. This study provides a quantitative assessment of AI-specific hardware power usage, focusing on the NVIDIA H100 GPU. The analysis compares AI data center energy consumption to the average US household power usage, demonstrating that a single AI rack consumes approximately 39 times the energy of a typical household. Additionally, a scalability analysis estimates that approximately 87 new hyper-scale data centers consume the electricity as much as consumed by New York City. This emphasizes that with rapid growth of AI Data Centers, the large-scale deployment could lead to an unprecedented rise in global energy demand. Furthermore, the study evaluates the impact of heat dissipation on cooling requirements, highlighting the need for energy- efficient cooling solutions, including liquid and immersion cooling techniques. Future research directions include energy- efficient AI models, renewable energy integration, sustainable AI accelerator designs, and intelligent workload optimization to mitigate the environmental impact of large-scale AI adoption. This research provides critical insights for designing more sustainable AI-driven data centers while maintaining high-performance computing efficiency.

KEYWORDS: AI data centers, power consumption, heat dissipation, energy efficiency, data center cooling, GPU computing, urban energy impact, sustainable AI, high-performance computing, hyper-scale infrastructure, thermal management, workload optimization, carbon footprint reduction, renewable energy integration, AI accelerators.

I. INTRODUCTION

Artificial Intelligence (AI) and Machine Learning (ML) applications have significantly reshaped industries, necessitating high-performance computing infrastructure. The exponential growth in AI training models has led to an unprecedented surge in energy consumption within data centers.

According to a 2023 report by the International Energy Agency [1], data centers contribute to nearly 1% of global electricity demand, with AI workloads expected to increase this share substantially. Tech giants such as Google, Amazon, and Microsoft are actively investing in AI-specific chips and optimizing data center cooling to mitigate energy footprints. However, research on quantifying the power and cooling demands of AI hardware remains limited.

This study provides a quantitative assessment of power usage in AI-specific hardware, particularly the NVIDIA H100 GPU, and evaluates its impact on urban energy infrastructures.

As AI data centers expand, the impact on power grids becomes a major concern. This study aims to:

- Compare power usage of AI infrastructure with household electricity consumption.
- Evaluate scalability and sustainability in AI data centers.
- Analyze heat dissipation and cooling requirements.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402015|

II. DEFINITION OF KEY TERMS

A. Power and Energy

Power (P): Power is the rate at which energy is consumed by a rack or an electrical device in a data center. It's likethe amount of energy consumed in each moment of time
Relevance: It tells us the peak load that should be supported and defines the capacity of power infrastructure.
Units: Watts or KiloWatts (kW)
Example: Typically, a rack with 42U size consumes around 10kW and if it is an AI rack (GPU), consumes around 30kW at any given moment of time
Energy (E): Total amount of power consumed over time by a rack or electrical device
Relevance: It has direct correlation with operational costs, electricity costs, carbon footprint and cooling
Units: kWh - kilo watt hour
Example: If all servers run continuously for 24 hr in a rack, then the amount of energy consumed is
$10kW \times 24kW = 240kWh$

B. BTU and Heat Dissipation

British Thermal Units (BTU) measure the amount of heat required to raise the temperature of one pound of water by
one degree Fahrenheit. In data centers, 1 watt (W) of power consumption produces 3.412 BTU/hr of heat, making BTU a critical metric for cooling system design.

III. RELATED WORK

Several studies have examined power consumption in data centers. In IEEE AI 2023 proceedings [2], the authors analyzed the impact of AI workloads on cloud infrastructure, finding that GPU-based data centers consume up to 4x more power than traditional CPU-based systems.
A recent study in Green AI 2022 [3] proposed AI model optimizations, such as low-bit quantization and model pruning,
to reduce energy usage. However, these solutions do not address cooling overheads, which contribute significantly to total power consumption.
This paper extends prior research by focusing on AI rack-level energy demand and proposing potential solutions to minimize power inefficiencies.

IV. METHODOLOGY

This study follows a quantitative approach, utilizing:
• NVIDIA's H100 GPU technical specifications [4].
• US Energy Information Administration (EIA) power data [5].
• Empirical energy consumption datasets from AI data centers [6].
• New York Grid Monitor electricity demand data [7].

V. POWER CONSUMPTION ANALYSIS

A. *AI Data Center Energy Demand*
A typical AI rack consists of multiple servers, each containing multiple GPUs.

$$P_{\text{rack}} = 48 \times 700W = 33.6kW \quad (1)$$

Considering 40% additional overhead for CPUs, memory, and networking:

$$P_{\text{total}} = 1.4 \times 33.6kW = 47.04kW \quad (2)$$

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402015|

B. *Comparison With Household Power Consumption*

The average US household electricity consumption is 10,632 kWh per year [5].

$$P_{\text{house}} = \frac{10,632}{365 \times 24} = 1.215kW \tag{3}$$

Ratio of AI rack consumption to a household:

$$\frac{P_{\text{rack}}}{P_{\text{house}}} = \frac{47.04}{1.215} \approx 38.7 \tag{4}$$

VI. SCALABILITY AND GRID IMPACT

A. AI Data Hall Energy Demand

Usually, a medium hyper scale data hall is around 35, 000 square feet, and AI rack is close to 7 square feet (server racks are usually 24 inches width and 42 inches depth which is 2*feet* × 3.50*feet* foot print). Most of the cases only 65% of the data hall floor is allocated for rack installation. Of that only 50% is used for racks deployments. Rest of the floor space needs to accommodate other infra such as chillers, staircases, meeting rooms, MMRs, doors, rack inventory room, aisles for walkways (typically 24 inches wide and 14 feet long). That said, we have ≈ 1625 racks in a data hall as calculated in equation 5 which is maximum, but the real value is even lower which is always limited by cooling capacity, power capacity, and raised floors.

$$N_{\text{racks}} = \frac{35,000 \times 0.65 \times 0.50}{7} \approx 1,625 \tag{5}$$

$$P_{\text{hall}} = 1,625 \times 47.04kW \approx 76.5MW \tag{6}$$

B. AI Data Centers vs. New York City's Power Demand

We now derive what number of AI Data Centers require the power consumed as much as New York City's peak electricity. The peak demand for New York City is 19,938 MW [7].

$$N_{\text{halls}} = \frac{19,938MW}{P_{\text{hall}}} \tag{7}$$

$$N_{\text{halls}} = \frac{19,938MW}{76.5MW} \approx 261 \tag{8}$$

Assuming each data center has 3 halls:

$$N_{\text{centers}} = \frac{261}{3} \approx 87 \tag{9}$$

C. Comparing with CPU based Data Centers

AI workloads rely heavily on GPUs, which consume more power than conventional CPU-based data centers. Table I presents a comparison between AI GPU racks and traditional CPU servers.

TABLE I
POWER CONSUMPTION: AI GPUS VS. CPU DATA CENTERS

Hardware	Power per Unit	Total Power (1 Rack)
NVIDIA H100 GPU	700W	47.04 kW
Intel Xeon CPU	150W	6.75 kW

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402015|

From Table I, it is evident that AI GPU racks consume nearly 7 times more power than CPU-based systems, highlighting the need for power-efficient AI hardware.

VII. HEAT DISSIPATION CONSIDERATIONS

A. Thermal Output of AI Racks

The power consumed by AI hardware is almost entirely converted into heat. The **thermal design power (TDP)** of an NVIDIA H100 GPU is 700W, which is dissipated as heat. Using the **standard thermal conversion factor** of 3.412 BTU/hr per watt, the heat output per GPU is calculated as:

$$Q_{\text{GPU}} = 700 \times 3.412 = 2,388.4 \text{ BTU/hr} \tag{10}$$

For an entire AI rack containing **48 GPUs**, the total heat dissipation is:

$$Q_{\text{rack}} = 2,388.4 \times 48 = 114,643.2 \text{ BTU/hr} \tag{11}$$

B. Cooling Requirements

To maintain operational efficiency and prevent overheating, AI data centers rely on effective cooling strategies. This section details the key parameters affecting airflow-based cooling.

1. Mass Flow Rate of Air: The required airflow is determined based on the **density of air at standard atmospheric conditions**. The **mass flow rate of air** is derived using the following equation:

$$M_{\text{air}} = V \times \rho \tag{12}$$

where:
M_{air} is the **mass flow rate of air** (lb/hr)
V is the **volume flow rate** of air in cubic feet per minute (CFM)
ρ is the **density of dry air at room temperature**

For a typical AI rack, the airflow volume is assumed to be **500 CFM**, and the density of air at room temperature is **0.075 lb/ft³**. Converting to an hourly flow rate:

$$M_{\text{air}} = 500 \times 60 \times 0.075 \tag{13}$$

$$M_{\text{air}} = 2,250 \text{ lb/hr} \tag{14}$$

2. Temperature Rise Calculation: The **temperature rise** due to heat dissipation is determined using the fundamental heat transfer equation:

$$Q = M_{\text{air}} \times C_p \times T_{\text{rise}} \tag{15}$$

where:
Q = **Total heat dissipated** (BTU/hr)
M_{air} = **Mass flow rate of air**
C_p = **Specific heat capacity of air** (BTU/lb°F),
ΔT = **Temperature rise (°F)**. (lb/hr)

Rearranging to solve for T_{rise}:

$$T_{\text{rise}} = \frac{Q}{M_{\text{air}} \times C_p} \tag{16}$$

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402015|

Substituting the known values:

$$T_{\text{rise}} = \frac{114,643.2}{2,250 \times 0.24.} \tag{17}$$

$$T_{\text{rise}} = \frac{114,643.2}{540} \tag{18}$$

$$T_{\text{rise}} \approx 21.2^{\circ}\text{F}. \tag{19}$$

This result indicates that if the only cooling mechanism used is **airflow-based cooling**, the air temperature in an AI rack will **increase by 21.2°F** as it absorbs the dissipated heat.

C. Need for Advanced Cooling Solutions

A **21.2°F increase per AI rack** requires **additional cooling measures** beyond basic airflow, including:

1. *Liquid Cooling:* Liquid cooling systems utilize **direct coolant circulation** around GPUs, effectively transferring heat away from the hardware and significantly reducing reliance on high airflow rates.
2. *Immersion Cooling:* Immersion cooling involves submerging GPUs in a **dielectric fluid**, allowing for superior heat dissipation while minimizing energy consumption from traditional air-cooling methods.
3. *Hot and Cold Aisle Containment:* The implementation of hot and cold aisle containment ensures that **heated air is efficiently exhausted** while preventing its recirculation into the cold air supply. This technique optimizes airflow management in hyperscale data centers.

VIII. CONCLUSION AND FUTURE RESEARCH

The rapid adoption of artificial intelligence (AI) and deep learning workloads has led to a significant increase in energy consumption across hyperscale data centers. This study quantitatively assessed power usage in AI-specific hardware, particularly the NVIDIA H100 GPU, and compared it to the average US household energy consumption. Our findings indicate that a single AI rack consumes approximately **39 times** the power of an average household, demonstrating the energy-intensive nature of AI infrastructure.

Moreover, our scalability analysis shows that a large-scale deployment of close to **87 new hyperscale data centers** consume the electricity equivalent to New York City electricity consumption, which necessitates and emphasizes the optimization of resources imminently. Additionally, heat dissipation challenges in AI racks necessitate **advanced cooling techniques** to prevent thermal runaway and maintain operational efficiency.

Despite these insights, this study is only the beginning of a broader research avenue. Future work should focus on the following areas:

- **Energy-efficient AI model training:** Optimization techniques such as quantization, pruning, and federated learning could significantly reduce the power footprint of AI workloads.
- **Renewable energy integration:** Investigating how AI data centers can leverage solar, wind, and geothermal energy to offset their carbon footprint.
- **AI-driven cooling solutions:** Implementing machine learning-based **adaptive cooling systems** to dynamically manage heat dissipation and airflow optimization.
- **Sustainable hardware design:** Evaluating emerging **low-power AI accelerators**, such as neuromorphic chips and optical computing, as alternatives to traditional GPUs.
- **Regulatory and policy implications:** Understanding the role of governmental policies in enforcing sustainable data center growth while maintaining AI innovation.

By addressing these challenges, future research can contribute to the development of **greener, more sustainable AI-driven data centers** that balance computational performance with environmental responsibility.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 2, February 2025

|DOI: 10.15680/IJIRSET.2025.1402015|

REFERENCES

[1] International Energy Agency (IEA), "Tracking clean energy progress 2023," Paris, France, 2023, license: CC BY 4.0, Accessed: 2024-02-25. [Online].
[2] Available: https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks#tracking
[3] A. Johnson and M. Patel, "Ai data center energy consumption: Trends and challenges," in Proceedings of IEEE AI Conference. IEEE, 2023, pp. 45–52.
[4] J. Smith and R. Doe, "Green ai: Optimizing artificial intelligence for sustainability," IEEE Transactions on Sustainable Computing, vol. 7, no. 4, pp. 125–138, 2022.
[5] NVIDIA Corporation, "H100 gpu architecture whitepaper," 2024. [Online]. Available: https://www.nvidia.com/en-us/data-center/h100/
[6] U.S. Energy Information Administration, "Annual energy outlook 2024," 2024. [Online]. Available: https://www.eia.gov/outlooks/aeo/
[7] Columbia Energy Policy Institute, "Projected electricity demand growth from generative ai," 2024. [Online]. Available: https://www.energypolicy. columbia.edu/
[8] U.S. Energy Information Administration, "New york electricity grid monitor," 2024. [Online]. Available: https://www.eia.gov/electricity/gridmonitor/ dashboard/electric overview/regional/REG-NY

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details