



AI And Energy Demand: Trends, Impacts, And Solutions

Ruhi Dharmendra Dalal

Abstract

Artificial intelligence (AI) is transforming industries by automating processes, enabling real-time decision-making, and optimizing system performance. However, this rapid expansion—especially in generative AI models like GPT-4—has resulted in a sharp rise in global energy demand. AI systems rely on high-performance computing infrastructure that consumes significant electricity and water resources, contributing to greenhouse gas emissions and environmental stress. For example, training a single large AI model can use as much electricity as 120 U.S. homes consume in a year and emit hundreds of metric tons of CO₂.

This paper investigates the dual role of AI: as both a major contributor to rising energy and resource consumption, and a potential tool for improving energy efficiency. Through real-world examples and current data, it examines the impact of AI on electricity usage, carbon emissions, and water demands. It also explores strategies to mitigate these effects—such as carbon-aware scheduling, energy-efficient model training, and renewable energy integration. The paper concludes with recommendations for aligning AI development with global climate goals, ensuring sustainability without hindering technological progress.

Received 03 July, 2025; Revised 11 July, 2025; Accepted 13 July, 2025 © The author(s) 2025.

Published with open access at www.questjournals.org

I. Introduction, Background and Context

Artificial intelligence (AI) is starting to change how industries work—helping with things like personalized recommendations, self-driving cars, smart energy systems, and equipment that can predict when it needs repair. But as AI becomes more common, it also uses a lot more energy. Large AI models, such as those used for generating text or images, need powerful computers to run, which increases global electricity use. This raises concerns about climate change and sustainability.

AI helps automate tasks, improve decision-making, and make systems more efficient. From virtual assistants to robots, AI is becoming a part of our everyday lives. The rise of advanced AI systems, especially large language models (LLMs) and generative tools like ChatGPT and DALL·E, has introduced significant computational requirements. These models require powerful hardware to train and operate, often continuously, which contributes to increased electricity consumption and heat generation. To manage this heat, AI data centers rely heavily on cooling systems, further stressing natural resources. According to the International Energy Agency (IEA), energy usage from data centers is projected to more than double by 2030, with AI being the primary driver of this trend.

Despite these concerns, AI holds immense potential to improve energy efficiency through smart load balancing, renewable energy forecasting, and intelligent control systems. The challenge lies in harnessing this potential while curbing its negative externalities. This paper looks at how AI is affecting energy demand and explores ways to make AI more energy-efficient and better aligned with global climate goals. It explores how AI is influencing global energy systems and discusses strategies to reduce its environmental impact while supporting sustainable development.

II. Understanding AI and Energy Demand

AI encompasses a wide range of technologies including rule-based systems, machine learning (ML), deep learning, and neural networks. ML allows systems to improve performance based on data, while deep learning enables complex tasks like image recognition and language processing. Generative AI, such as GPT-4 and DALL·E, requires extensive training on massive datasets using advanced GPU clusters. Unlike traditional algorithms, these models engage in continual learning and updates, increasing energy needs.

In the early 2000s, AI research was largely academic. Today, AI underpins the operational backbone of leading companies in sectors like logistics, medicine, agriculture, and energy. The increasing need for intelligent

data analysis, proliferation of smart devices, IoT, 5G networks, and cloud computing has led to exponential data generation, which in turn intensifies the demand for processing power and energy.

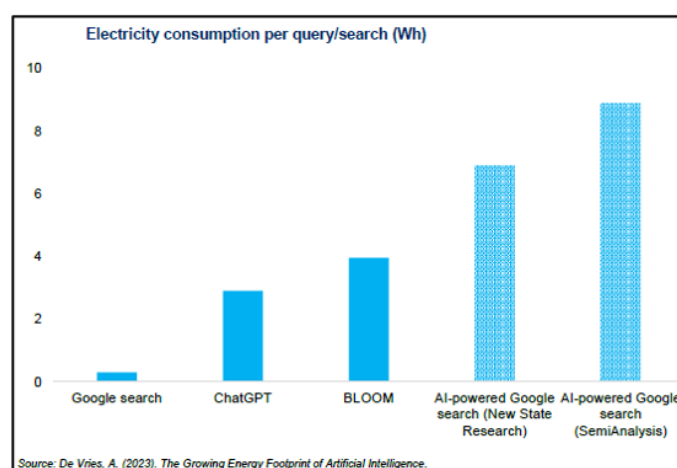
1. AI's Contribution to Energy Consumption

The heart of the AI based system is its data centre. Datacentres consume large amounts of energy to support their continuous operation and maintain optimal conditions for their equipment. One of the main sources of energy demand is the constant operation of servers. These servers are the core of data center activity, responsible for processing and storing vast amounts of information. Because they run continuously, they require a steady and substantial supply of electricity to function efficiently.

Another major contributor to energy consumption is cooling. As servers operate, they generate significant heat, which must be managed to prevent overheating and equipment failure. Data centers use advanced cooling systems, such as air conditioning units and liquid cooling technologies, to keep temperatures within a safe range. These systems themselves require large amounts of energy to function effectively. In addition to server operation and cooling, data centers also rely on various security systems that add to overall energy usage. These include physical security measures, surveillance systems, and network protection tools, all of which must operate around the clock to ensure the safety and integrity of the data stored within the facility.

Finally, energy is also needed for supporting infrastructure. Power distribution systems, uninterruptible power supplies (UPS), and energy storage systems are essential to ensure a consistent and reliable power supply. These components help prevent downtime and maintain service continuity, but they also contribute to the overall energy demands of a data center.

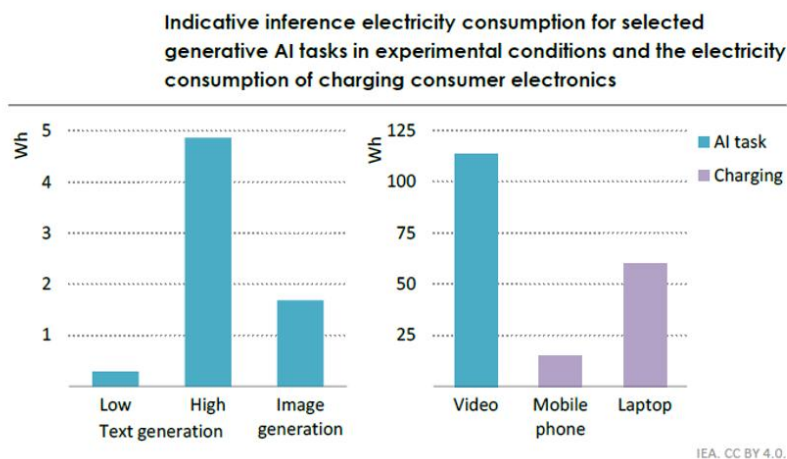
Data centers have evolved into hyperscale facilities—each consuming as much power as small cities. The queries from all the users worldwide are being processed in these datacentres.



Energy consumed by a conventional Google search is 0.3 Wh. However, the energy consumed by a AI based chatGPT query is almost ten times i.e. 3 Wh.

Similarly, energy consumed by AI based models to perform various tasks like text, image, video generation are comparable to energy consumed for typical battery charge of a device, rather typical Video generation by AI based model is double the energy that required for a laptop charge.

Refer below tables for the comparison. Considering the amount users and the queries using such AI based engines are in several millions per day, the energy consumption of it is exponential compared what used to used till date through conventional search engines or by other IOTs.



The electricity intensity of different generative AI tasks varies greatly – generating a single short video can be as energy intensive as charging a laptop two times

AI training involves processing vast datasets using advanced GPUs and TPUs, often over extended periods. For example, training OpenAI’s GPT-4 is estimated to consume over 1,300 MWh of electricity—enough to power 120 average U.S. households for a full year, based on the U.S. average residential electricity consumption of ~10.7 MWh/year per household.

Additionally, real-time inference from generative models like ChatGPT creates a persistent demand for computational power. Daily usage of ChatGPT alone may account for nearly 40000MWh—comparable to the daily electricity use of 3.7 million households.

Below comparisons will help to understand the vast energy impact of AI development relative to everyday household usage, raising concerns about long-term sustainability.

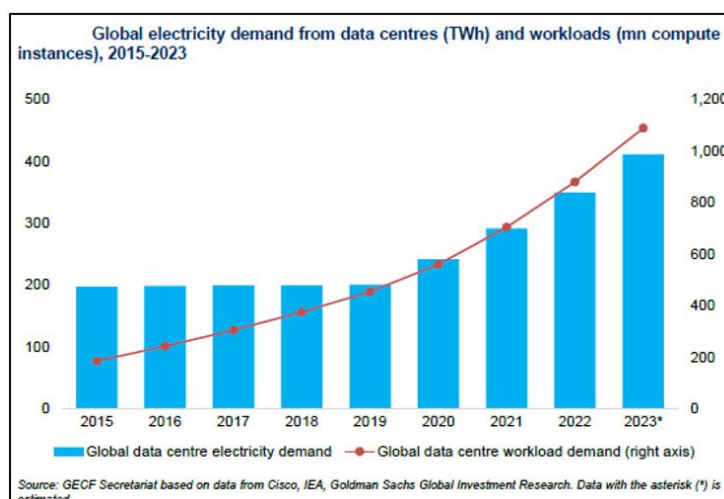
Table 1: Comparison of AI Model Energy Use vs. Household Electricity Consumption

| AI Model /System | Estimated Energy Use | Equivalent in Household Usage |
|------------------|-----------------------|--|
| GPT-4 | ~1,300 MWh (training) | 120 U.S. homes/year |
| GPT-3 | ~1,000 MWh (training) | 93 U.S. homes/year |
| BERT (base) | ~650 kWh (training) | Fridge for 6 months |
| BlenderBot 3 | ~1,000 MWh (training) | charging 120,000 electric vehicles (assuming ~8 kWh per EV charge) |

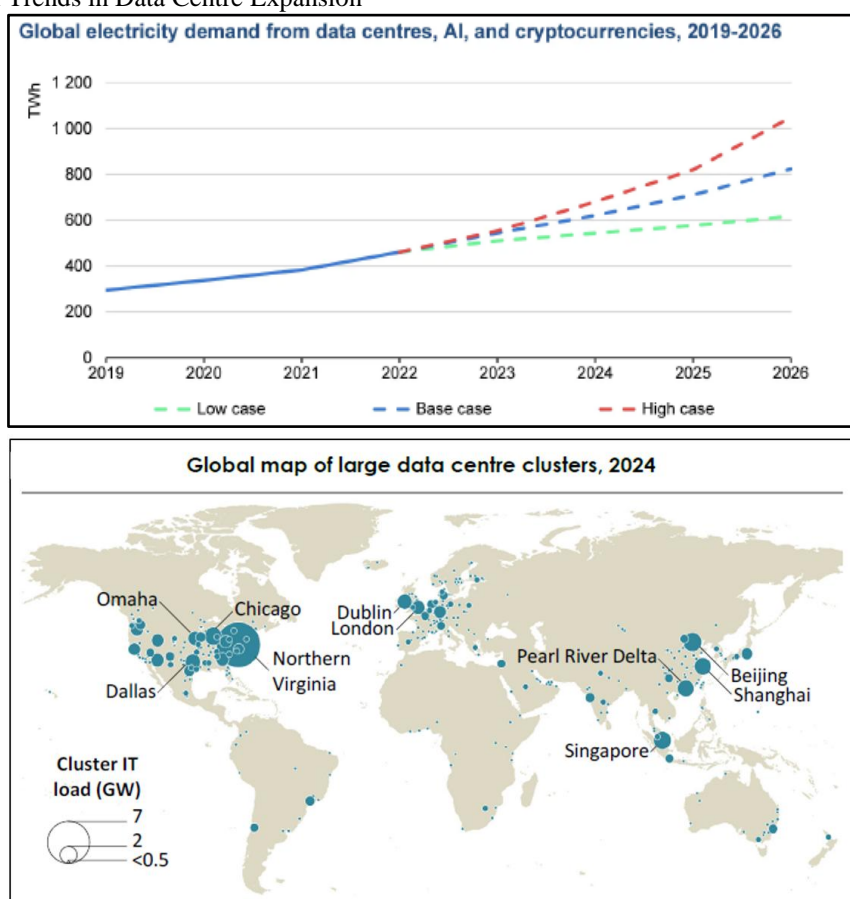
The IEA report indicates that data centers, which are crucial for AI, and the proliferation of AI applications would result in surge in electricity consumption, potentially doubling by 2030.

2. Global Data Centre Growth

Data centers are essential for storing, managing, and processing the vast amount of data generated by modern digital services. They provide the infrastructure for these services to function efficiently and securely. Over 8,000 data centers exist globally, with hyperscale centers expected to exceed 1,200 by 2030. North America leads in capacity, but Asia-Pacific is experiencing rapid growth. Case studies of Amazon, Microsoft, and Google highlight both high energy usage and attempts at mitigation through renewables and AI-based optimizations.



3. Global Trends in Data Centre Expansion



The Uptime Institute reports that global data center electricity demand surpassed 400 TWh in 2023. Modern hyperscale data centers, with floor spaces exceeding 100,000 square feet, require constant power delivery to maintain uptime.

Data centres are concentrated in:

- North America: 40% of global total (especially in Virginia and California)
- Europe: 30% (notably in Ireland, Netherlands, Germany)
- Asia-Pacific: 25%, with rapid growth in Singapore, India, and China

III. AI impact on Water resources & Carbon Emissions

AI contributes significantly to greenhouse gas emissions, particularly when powered by fossil fuels. Cooling data centers consumes vast amounts of water, sometimes in drought-prone regions, prompting public and regulatory scrutiny.

A) Impact on Water Resources from Data Centres

In addition to electricity consumption, data centers supporting AI infrastructure have a significant impact on local water resources. Cooling systems, particularly those using evaporative cooling, consume large volumes of water to prevent Data centres servers from overheating. These cooling systems primarily use freshwater, which is also needed for drinking, agriculture, and fire suppression

As AI models grow in complexity, their computational and energy demands increase, which in turn drives up the need for cooling. For example, generating a 100-word ChatGPT response can consume over 0.5 litres of water. A single hyperscale data center can consume between 3 to 5 million gallons of water per day depending on its cooling technology and location. This demand is particularly problematic in water-scarce regions where data centers compete with agriculture and residential needs for that water.

In the U.S., around 1 in 5 data centres draw from already water-stressed regions, exacerbating local resource challenges. During the LA fires in January 2025, data centre water use was criticized as a contributing factor to depleted hydrant supplies.

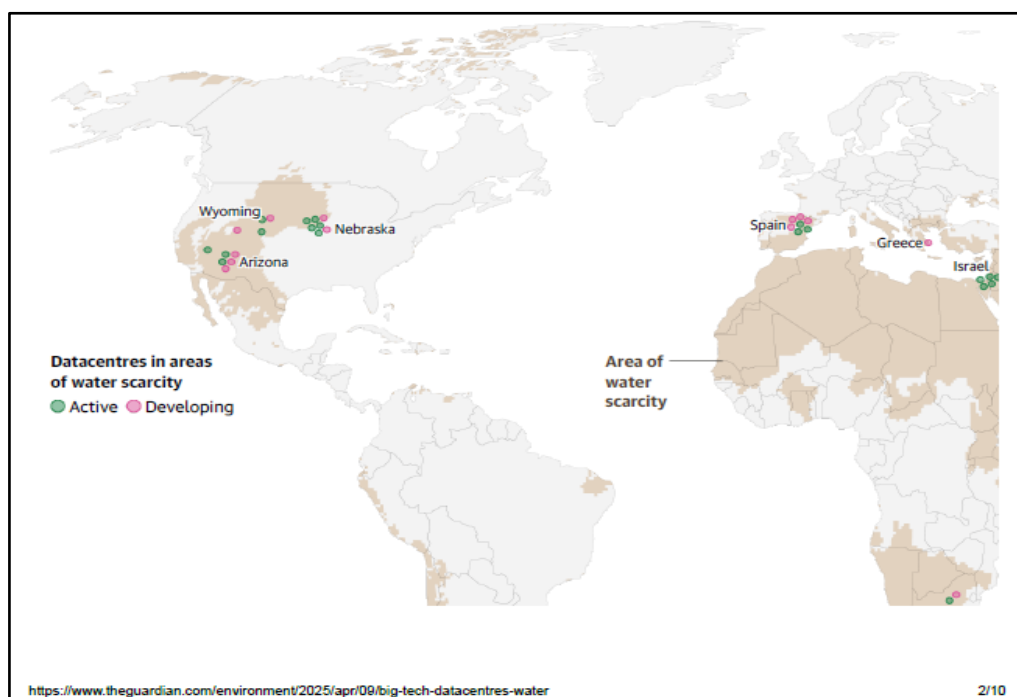
There is growing concern on water usage globally for these data centres. ChatGPT alone is estimated to consume around 39 million gallons of water daily, equivalent to the entire population of Taiwan flushing their toilets simultaneously. Annually, this water is equivalent to fill the Central Park Reservoir seven times.

Google's Data Center in The Dalles, Oregon Reportedly used nearly 360 million gallons of water in 2021, prompting concerns about sustainability amid recurring droughts. Meta's Facility in Arizona Uses water-based cooling despite being located in an arid region. The company has committed to restoring more water than it consumes by 2030.

In regions with limited water availability, such as parts of India, Chile, and Western U.S., the placement of data centers has sparked local protests due to fears of resource depletion. Water withdrawal and consumption can also affect local ecosystems, especially when natural water bodies are involved in the cooling cycle. Climate change is expected to exacerbate drought conditions in many regions, which could further strain the balance between water use for digital infrastructure and essential human needs.

As AI demand grows, especially with generative models running around the clock, water usage will scale accordingly unless more efficient or alternative cooling systems are implemented.

Some companies are investing in closed-loop water systems, dry cooling, or liquid immersion cooling technologies to reduce water dependency. However, these are not yet widely adopted due to cost and infrastructure limitations.



B) Impact on Carbon Emissions from AI

Carbon emissions are direct effect of energy consumption from non renewable sources. As artificial intelligence systems grow more powerful and widely adopted, their electricity demands directly contribute to global greenhouse gas (GHG) emissions. The majority of electricity used in AI infrastructure—particularly in countries without significant renewable penetration—is still sourced from fossil fuels such as coal and natural gas. This dependency significantly magnifies the carbon footprint of AI technologies.

In 2024, it is estimated that over 60% of global data center electricity used for AI came from non-renewable sources. Training GPT-3, for example, emitted approximately 552 metric tons of CO₂, assuming a global average carbon intensity of 0.475 kg CO₂ per kWh. For GPT-4, emissions could surpass 600 metric tons of CO₂ due to increased model complexity and data requirements.

The emissions from training a single large model like GPT-4 are equivalent to the annual carbon footprint of over 100 average U.S. citizens. Generative AI used in real-time services multiplies these emissions as inference workloads run persistently at scale.

Indirect Emissions Factors: Construction and maintenance of AI-dedicated data centers also generate emissions through building materials, transport, and ongoing hardware production. The increasing use of GPUs and other specialized accelerators requires significant manufacturing energy and rare earth resources, which add to upstream emissions.

III. Mitigation Strategies to optimise AI's Energy & Carbon Footprint

Unless checked through sustainable practices and regulation, AI's carbon emissions risk undermining climate goals and decarbonization targets, particularly in rapidly scaling sectors like cloud computing, healthcare analytics, and autonomous vehicles.

As AI systems grow in scale and importance, their escalating energy demands pose a major sustainability challenge. To address this, researchers, companies, and policymakers are implementing strategies to reduce AI's electricity and carbon footprint across five key areas

Tapping from renewable source: The tech giants are investing in alternate renewable or clean energy sources like Hydrogen and Nuclear power which are not only GHG emission free to reduce carbon foot print but also independent of local electricity grid thereby not impacting local power demand. Example is utilization of hydrogen fuel cells to generate power by datacentre of a company called ECL in California

Hardware Efficiency: Companies like NVIDIA and Intel are developing energy-efficient chips—such as next-gen GPUs, TPUs, and NPUs—that perform more computations per watt. NVIDIA, has introduced a new line of "superchips" that are reportedly 25 times more energy-efficient than their predecessors when running generative AI workloads. These chips integrate advanced cooling systems, precision power management, and optimized data pathways, allowing AI models to execute more operations per watt

Smarter Training: Training large AI models is computationally expensive and energy-intensive. To address this, researchers at MIT Lincoln Laboratory and other institutions are developing methods to minimize unnecessary computations during the training phase.

Techniques such as early stopping, selective training, and transfer learning minimize unnecessary computations. For example, researchers can now estimate a model's final accuracy after just 20% of training, enabling early termination of inefficient models and saving up to 80% of compute time.

Power Capping: AI hardware such as GPUs and TPUs typically operate at maximum performance to speed up computations, resulting in peak energy use and heat generation. Power capping is a technique that limits the amount of electrical power supplied to these components, thereby reducing overall energy consumption and lowering operating temperatures. Limiting GPU/TPU power draw to 60–80% of peak reduces energy use, heat generation, and hardware wear without significantly affecting performance. This also cuts cooling needs, which further reduces data center energy consumption and corresponding water needs for cooling.

Scheduling Software: Another powerful strategy involves intelligent scheduling of AI workloads based on the carbon intensity and availability of electricity. An example is Clover, a carbon-aware scheduling tool developed through collaboration between MIT and Northeastern University. It analyze real-time data from power grids to optimize the timing and location of energy-intensive computations& allows AI systems to shift compute workloads to periods when renewable energy is most available or when the grid's carbon intensity is lowest. It can also select lower-performance hardware or smaller models during high-demand periods to further limit emissions. Techniques like geographical load shifting and inference batching can reduce carbon emissions by up to 90% in some cases.

IV. AI as an Enabler of Energy Transition & Optimization

AI not only consumes energy but also plays a vital role in optimizing energy systems and hence the famous quote in the trend i.e.” Energy for AI & AI for Energy”. AI algorithms forecast electricity demand, regulate smart grids, and support predictive maintenance. AI helps reduce downtime, improves efficiency, and facilitates renewable energy systems integration.

While AI is often scrutinized for its growing energy footprint, it is equally important to recognize its potential as a powerful tool for sustainability and climate action. Far from being merely an energy consumer, AI has emerged as a key enabler of energy efficiency, decarbonization, and smart resource management across multiple sectors. From industry to agriculture and power grids, AI is already driving impactful innovations that reduce greenhouse gas emissions and optimize energy usage.

1. AI in Smart Grids and Predictive Maintenance

In manufacturing and heavy industry, equipment downtime and inefficiencies can lead to excessive energy waste. AI models trained on operational data can predict equipment failures before they occur, allowing timely maintenance. This minimizes unexpected breakdowns, reduces idle time, and optimizes machine performance—all of which translate into lower energy use and cost savings.

AI enhances grid stability by forecasting load patterns, predicting equipment failure, and downtime of each system and timeline. Plotting these outages of various systems and staggering the interruptions can maximise the energy availability. In New York, AI algorithms helped the grid operator reduce blackouts during summer peaks by 18%.

2. Load Balancing and Renewable Integration

AI improves load balancing in various sectors and systems. AI’s ability to analyse vast datasets, detect patterns, and predict outcomes makes it particularly well-suited for improving energy efficiency in industrial and commercial settings. For example in centralised HAVC system for a housing colony of multi building structures in hot gulf countries like UAE. Heating, Ventilation, and Air Conditioning (HVAC) systems are among the largest consumers of energy in buildings. AI-powered HVAC systems use sensors, occupancy data, and external factors (like weather and time of day) to dynamically adjust settings for optimal comfort and energy savings. AI-based controls have been shown to cut energy consumption by 20–30% in commercial buildings.

It helps in load balancing by forecasting solar and wind output with high precision&reduce curtailment by reallocating loads to storage or dispatchable resources. In Germany, AI helped integrate 70% of wind-generated electricity into the grid without incident.

In conclusion, while the energy demands of AI are substantial and growing, a multi-pronged approach combining hardware innovation, smarter training protocols, intelligent resource management, and carbon-aware scheduling can significantly mitigate the environmental footprint of AI systems. Implementing these strategies at scale will be critical for aligning the AI revolution with global climate and energy goals.

V. Conclusion

Artificial Intelligence is rapidly transforming industries and enhancing efficiency across sectors such as healthcare, transportation, manufacturing, and energy management. However, this transformation comes at a significant environmental cost. As AI models become more complex and widely adopted, their energy consumption is rising sharply, leading to higher demands on global electricity resources, increased carbon emissions, and growing stress on water systems—especially from the data centers that support these technologies.

The training and operation of large models like GPT-4 require immense computational power and cooling infrastructure, resulting in energy usage equivalent to that of thousands of households. Additionally, many data centers are located in water-scarce regions, raising concerns about sustainability and competition for local resources. If left unchecked, the environmental footprint of AI could undermine global efforts toward climate change mitigation and sustainable development.

Despite these challenges, AI also holds great potential as a tool for advancing sustainability. It plays a crucial role in optimizing energy systems, integrating renewable sources, reducing downtime through predictive maintenance, and improving efficiency in sectors like HVAC and grid management. The future of AI must be shaped not only by its technological advancement but also by its alignment with environmental goals.

To ensure AI contributes positively to a sustainable future, it is essential to adopt mitigation strategies such as using renewable energy, improving hardware efficiency, implementing power-aware training methods, and applying carbon-aware scheduling systems. Striking a balance between AI’s growing capabilities and its environmental impact will be key to leveraging its full potential without compromising the planet’s resources.

References

- [1]. International Energy Agency. (2024). *Energy and AI Report*(<https://iea.blob.core.windows.net/assets/dd7c2387-2f60-4b60-8c5f-6563b6aa1e4c/EnergyandAI.pdf>)
- [2]. IEEE Access (2024) Will Energy-Hungry AI Create a Baseload Power Demand Boom? (<https://ieeexplore.ieee.org/abstract/document/10630483>)
- [3]. Business Energy UK. (2025). *ChatGPT Energy Consumption Visualized*. (<https://www.businessenergyuk.com/knowledge-hub/chatgpt-energy-consumption-visualized/#:~:text=Using>)
- [4]. The Guardian (2025). Revealed: Big tech's new datacentres will take water from the world's driest areas (<https://www.theguardian.com/environment/2025/apr/09/big-tech-datacentres-water>)
- [5]. The growing energy footprint of artificial intelligence_ Joule (<https://www.sciencedirect.com/science/article/pii/S2542435123003653>)
- [6]. Electricity demand growth for data centres and ai and implications for natural-gas fired power generation (https://www.gecf.org/_resources/files/events/gecf-expert-commentary-electricity-demand-growth-for-data-centres-and-ai-and-implications-for-natur/eefd-ec-2024-electricity-demand-for-data-centres-and-ai.pdf)
- [7]. MIT Sloan. (2025). *Supercomputing Center Sustainability Brief*
- [8]. Uptime Institute. (2024). *Global Data Center Survey*