**Journal of Information Systems Engineering and Management**
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

# Toward Federally Aligned Green AI: Reducing Data Processing Time and Carbon Emissions with Sustainable Vendor-Marketplace Architectures

Mohammed Omer Shakeel Ahmed
*University of Texas at Arlington*
*Orcid ID: 0009-0001-3330-1892*
*Fnu.mohammedomersha@mavs.uta.edu*

**ARTICLE INFO**

Received: 24 Mar 2025

Revised: 29 Apr 2025

Accepted: 20 May 2025

**ABSTRACT**

Data Vendors and data marketplaces are seeing increasing adoption in research and marketing analytics. The carbon footprint of AI has increased rapidly as its adoption has increased, with data processing and data storage contributing to the majority of the impact of AI on the environment. In this paper, we can explore how data vendors and data marketplaces can play a key role in helping reduce the carbon footprint of AI by reducing the amount of data processing by using existing datasets or pre-processed data and can make the whole process not just leaner but also good for the environment. We also discuss how data vendors and data marketplaces can use sustainable practices to further help reduce the impact of AI on the environment.

**Keywords**: Data Vendors, Data Marketplaces, AI, Sustainability.

## 1. Introduction

One of the major contributors to carbon emissions from AI, particularly deep learning networks, is they require massive datasets and computational resources. These large data sets need large amounts of energy for data processing. GPT-3, which is classified as a large-scale model, consumes a significant amount of energy, with estimates indicating that the training of the GPT-3 model generated approximately 502 metric tons of CO2 [1]. Data storage and retrieval operations are some other additional energy-intensive operations that also have a huge environmental impact.

The US government has recently been focusing on sustainability in AI, amid growing concerns about AI's power consumption. They want a sustainable approach without limiting the capabilities of AI to stay competitive in the world.

Agencies like the DOE and NIST have developed frameworks to reduce AI's environmental impact, with federal initiatives encouraging carbon-efficient cloud computing and AI-driven climate solutions. The Biden administration's AI policies emphasize sustainability, ensuring innovation aligns with national environmental goals. By collaborating with the private sector, the US aims to balance advancement in AI with eco-friendly practices by reinforcing AI's role in a sustainable digital economy.

Large-scale computational infrastructure is powered by electricity, which relies on energy derived from fossil fuels. The increasing need for high-quality datasets drives energy consumption in data centers [2].

This paper investigates how data vendors and data marketplaces can optimize data usage, storage, and transactions to minimize AI's environmental impact.

## 2. AI's Electricity consumption and impact on the environment

Training deep learning models is a very energy-intensive process, with a single AI model consuming as much electricity as 126 Danish households in a year [3]. The exponential growth of AI has increased the demand for large datasets, which require substantial storage and are major contributors to high energy consumption and usage at data

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

centers. These facilities not only consume vast amounts of electricity but also require extensive water resources for cooling computational infrastructure, further impacting the environment[2]. Studies show that training models like GPT-3 require 1,287 MWh of electricity, producing 502 metric tons of CO2 emissions[1]. Additionally, as organizations prioritize high-quality data for AI applications, redundant data storage and inefficient data management practices further escalate energy use.

The environmental footprint of AI extends beyond just carbon emissions, making it imperative for companies and researchers to adopt more sustainable AI practices, including optimizing model efficiency using renewable energy sources and implementing better data storage strategies.

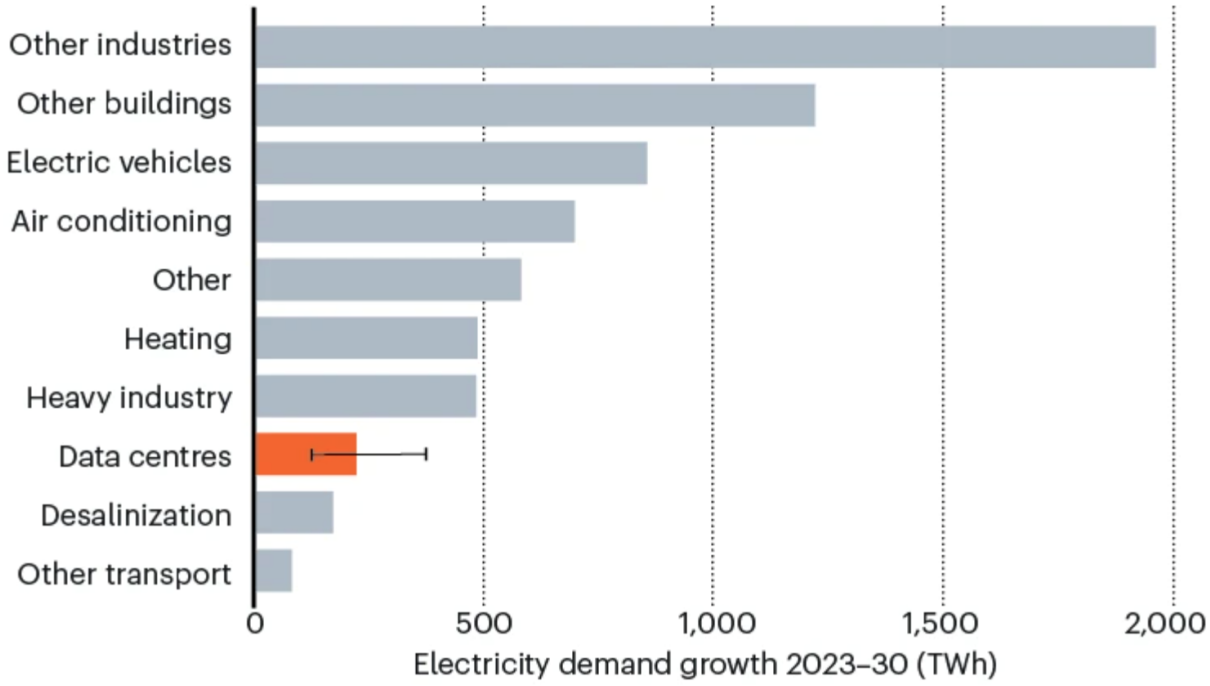


Fig 1 : World electricity demand growth by 2030, Nature[7]

## 3. The Role of Data Vendors in AI Sustainability

Data vendors gather data and provide these datasets to anyone who may need to use them for research purposes or training AI models. They offer these services on an on-demand basis by collecting required data or providing access to pre-curated datasets.

By providing access to pre-curated datasets, they minimize the need for energy-intensive data collection by eliminating the duplication of efforts across organizations. They play a crucial role in promoting AI sustainability by reducing redundancy in data collection, optimizing storage efficiency, and curating high-quality datasets to lower computational demands.

Google Dataset Search is a dataset discovery tool that provides search capabilities over potentially all datasets published on the Web [4]. There are many platforms like Google Dataset Search that facilitate the discovery of optimized datasets for AI applications, thereby reducing unnecessary data accumulation and preventing the excessive use of computational resources. By using this targeted approach, we can ensure that AI systems access only the most relevant and high-quality data and help to significantly lower energy consumption in the data preparation phase.

**Energy-Efficient Data Storage:**

Energy efficient data storage is defined as using green energy powered grids, using processes that require less energy, etc.Data vendors can contribute towards sustainability by using energy efficient data storage. It is a critical factor in AI sustainability and data vendors can contribute by leveraging cloud storage solutions powered by renewable energy.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Companies such as Amazon Web Services (AWS) and Google Cloud operate their data centers by using carbon-neutral methods that help to significantly cut emissions associated with traditional data storage methods[1]. By adopting these sustainable practices, the data vendors can help reduce the carbon footprint of AI-driven processes and also ensure uninterrupted data availability.

**Curated Data for Efficiency:**

When organizations need access to external data to gather insights or train their AI models using pre-curated datasets can be a very sustainable option. This is where data vendors come into play. These offerings that are tailored to specific domains help to enhance AI efficiency by reducing computational load and unnecessary data processing. They help by cutting down the need to use vast amounts of raw data that can be used to train AI models by using these refined, high-quality pre-curated datasets that are being offered by these data vendors. They can help in optimizing performance while also minimizing energy usage. This targeted approach helps to reduce the need for excessive computing power which has been identified as a major contributor to AI-related carbon emissions. Training a single AI model can emit over 626,000 pounds of $CO_2$, equivalent to the emissions of five cars over their lifetimes [5]. Data vendors act as key enablers of sustainable AI, and they help by refining data access, promoting efficient storage solutions, and reducing unnecessary computational overhead. As AI continues to evolve the goal to integrate sustainability into data management practices will be essential for ensuring that technological advancements align with global environmental goals.

## 4. The Role of Data Marketplaces in Reducing the Carbon Footprint of AI

**Centralized Data Access:**

Data marketplaces offer the capability for organizations to make their data available to other organizations at a premium price by using the marketplace feature. Many database tools such as snowflake and AWS are currently offering them. The major difference between Data marketplaces and data vendors is that data vendors are solely responsible for collecting raw data and creating datasets, whereas data marketplaces allow any organization to sell their data on the data marketplace platform. For example an insurance company can sell a dataset of what the auto insurance claims looked like after a hurricane in a specific area. Data marketplaces play a crucial role in reducing the carbon footprint of AI by optimizing data access, transactions, and storage efficiency. They offer a centralized data access platform that helps to minimize redundant storage by hosting datasets in a single location and, in turn, significantly reducing energy waste associated with duplicative data copies. By eliminating the need for multiple organizations to store the same datasets, these platforms can lower the demand for storage infrastructure and its associated carbon emissions.

**Optimizing Data Transactions:**

Data marketplaces have advanced search and filtering mechanisms that ensures that only relevant data is retrieved which int turn helps in optimizing data transactions and reducing unnecessary computational processing where large amounts of energy consumption required for excessive data handling. This streamlined approach not only enhances efficiency but also supports sustainability by conserving valuable computational resources.

**Carbon-Aware Transactions:**

Organizations are moving towards a green approach in all aspects of their operations. One such way companies with large servers handle this is by using carbon-aware transactions. Carbon-aware transactions represent a transformative approach to eco-friendly data management. It allows for the implementation of mechanisms that prioritize transactions based on carbon intensity that can help to significantly incentivize environmentally responsible data usage. For instance, Microsoft's carbon-aware computing initiative dynamically shifts workloads to regions with lower carbon intensity, thereby reducing overall energy consumption and emissions [3]. Such an approach can be applied to data marketplaces. This approach when taken up by data marketplaces where transactions could be scheduled based on real-time carbon intensity metrics helps to ensure that data transfers and processing occur in the most sustainable manner possible. By integrating these principles, data marketplaces can actively contribute to making AI more sustainable, significantly reducing its environmental impact while enabling advanced

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

data-driven innovations. As the demand for AI and big data continues to grow, incorporating sustainability measures within data marketplaces will be essential for balancing technological progress with environmental responsibility.

## 5. Collaborative Approaches to Sustainable AI

### Federated Learning and Decentralized Data Use:

Collaborative approaches to sustainable AI emphasize on decentralized data processing techniques that minimize energy consumption and help to reduce environmental impact. One such approach is federated learning, which enables AI models to be trained across multiple decentralized devices without transferring raw data to a central server. By keeping data local, federated learning significantly reduces the need for large-scale data transfers and cloud storage, which can help cut down on storage-related emissions created by AI. Google's federated learning framework for mobile devices is a very good example as it allows machine learning models to improve using on-device data without excessive cloud dependence [2]. This approach not only enhances privacy and efficiency but also contributes to AI sustainability by limiting the carbon footprint associated with data transmission

### Distributed Data Processing:

Another key strategy in sustainable AI is distributed data processing through edge computing, where AI models perform computations locally rather than relying on remote cloud infrastructure. NVIDIA's edge computing solutions is a prime example how this approach enables AI processing at the device level by reducing the energy required for cloud-based data transmission and storage [1].

## 6. Sustainable Practices by Data Vendors and Marketplaces

### Ethical Sourcing of Data:

Ethical sourcing of data ensures that datasets are collected responsibly, reducing redundancy and minimizing the need for excessive energy-intensive data gathering. By curating these high quality datasets that are relevant, vendors not only help to enhance AI model efficiency but also help in decreasing the carbon footprint associated with unnecessary data storage and processing.

### Carbon Offsets and Green Certifications:

Data vendors are largely investing in carbon offset programs to mitigate their environmental impact. These initiatives by data vendors involve funding reforestation projects, promoting renewable energy development and utilizing carbon capture technologies to counterbalance emissions generated by data storage and processing activities. These companies that prioritize sustainability often seek green certifications, which demonstrates their commitment to reducing the carbon footprint of AI-related operations.

### Integrating Renewable Energy into Infrastructure:

Another key sustainability practice that can be used by data vendors and data marketplaces is integrating renewable energy sources into their data infrastructure. Worlds leading cloud providers, such as AWS, Google Cloud, and Microsoft Azure are increasingly relying on wind, solar, and hydroelectric power to run their data centers which contribute to significantly cutting down emissions. Google utilizes renewable energy credits (RECs) to offset its energy usage which allows it to claim that its data centers are powered 100% by renewable sources, even if the direct supply is not entirely renewable [6]. By transitioning to greener energy solutions, data vendors and marketplaces contribute to the broader goal of sustainable AI development while ensuring that AI-driven innovations align with global climate objectives.

## 7. Collaborative Benefits Between Data Vendors, Marketplaces, and AI Companies

### Shared Responsibility for Sustainability:

By fostering a shared responsibility model, Data vendors, Data Marketplaces, and AI Companies can align their goals to reduce environmental impact while ensuring ethical AI development. Marketplaces play a key role in curating datasets from vendors that adhere to green data practices, such as energy-efficient data processing and responsible

data sourcing. AI companies, in turn, can prioritize partnerships with sustainable vendors, reinforcing the importance of environmentally conscious data usage.

**Incentivizing Green Data Practices:**

Incentive programs should be established to reward vendors who adopt eco-friendly practices, such as reducing carbon footprints and optimizing storage efficiency. These rewards could take the form of financial benefits, priority listing on marketplaces, or exclusive partnerships with AI firms. By embedding sustainability into the data supply chain these organizations can mitigate the environmental costs associated with AI advancements while maintaining high data quality standards. Ultimately this collaborative effort helps to foster a more responsible AI landscape where ethical and sustainable data sourcing becomes the industry norm.

## 8. Conclusion

In conclusion, data vendors and marketplaces can play a pivotal role in minimizing AI's carbon footprint by optimizing data usage, storage, and transactions. As AI models grow in complexity, the demand for vast amounts of data and computational power increases, leading to greater energy consumption and higher emissions. By implementing energy-efficient data management strategies, such as deduplication, intelligent caching, and streamlined data pipelines, vendors and marketplaces can significantly reduce AI's environmental impact. Additionally, adopting cloud-based storage solutions powered by renewable energy can further enhance sustainability efforts.

Looking ahead, future research should explore innovative approaches such as carbon-aware data processing that focuses on prioritizing energy efficiency in data operations and federated learning, which helps to minimize the need for centralized data transfers by enabling localized model training. These advancements could revolutionize the AI ecosystem by reducing both computational waste and the reliance on energy-intensive data centers. Understanding how these techniques can be scaled across industries can help us in shaping a more sustainable AI future.

To achieve meaningful progress, AI companies, data vendors, and marketplaces must take immediate action to integrate sustainable data practices. This includes prioritizing partnerships with eco-conscious data providers, adopting green computing infrastructures, and actively supporting research in sustainable AI methodologies. Governments and regulatory bodies should also encourage the adoption of environmentally responsible AI development through policies and incentives. The future of AI depends on the industry's collective ability to balance innovation with sustainability. By embracing responsible data stewardship today, all stakeholders can ensure that AI continues to evolve in an ethical and environmentally friendly manner, benefiting both technological progress and the planet.

## References:

[1] R. Huang, M. Shuai, "Carbon Footprint Management in Global Supply Chains: A Data-Driven Approach Utilizing Artificial Intelligence Algorithms," IEEE Xplore, 2025, https://ieeexplore.ieee.org/document/10547105

[2] L. Pengfei, Y. Jianyi, I. Mohammed and R. Shaolei, "AI's Water and Carbon Footprint: A Sustainability Crisis," 2024, https://arxiv.org/pdf/2304.03271.pdf

[3] C. Chinmayee, G. Rishab, G. Khushi and H. Nitasha, "Platforms to Calculate Carbon Footprints: A Step Towards Environment Sustainability," IEEE Xplore, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/10141821

[4] B. Dan, B. Matthew, N. Natasha, "Google Dataset Search: Building a search engine for datasets in an open Web ecosystem", Association for Computing Machinery, https://dl.acm.org/doi/abs/10.1145/3308558.3313685

[5] D. Payal, "The carbon impact of artificial intelligence", (2020). https://doi.org/10.1038/s42256-020-0219-9

[6] A. Martineau, "Amazon, Google, and Microsoft want to make cloud computing greener," Wired, Feb. 14, 2023. [Online]. Available: https://www.wired.com/story/amazon-google-microsoft-green-clouds-and-hyperscale-data-centers.

**Journal of Information Systems Engineering and Management**
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

[7] S. Chen, “How much energy will AI really consume? The good, the bad and the unknown”, Nature, Mar. 05, 2025.[Online]. Available : https://www.nature.com/articles/d41586-025-00616-z