Research in Health & Medical Sciences *Review Article*
2024 March; 3(1)

Assessment of Carbon Emissions Associated with Artificial Intelligence: A Narrative Review of Data Center Environmental Impacts and Green AI Strategies

Abbas Abbasnia[1,2], Morteza Bahlgerdi[3], Hossein Farash Khayalo[3], Edris Bazrafshan[1,2], Zohre Moeini[1,2]*

[1] Health Sciences Research Center, Torbat Heydariyeh University of Medical Sciences, Torbat Heydariyeh, Iran
[2] Department of Environmental Health Engineering, School of Health, Torbat Heydariyeh University of Medical Sciences, Torbat Heydariyeh, Iran
[3] Student Research Committee, School of Health, Iran University of Medical Sciences, Tehran, Iran

Corresponding Author Email: moeini.zohre@yahoo.com
Received: 2025/8; Revised: 2025/12; Accepted: 2025/12

Abstract

The rapid growth of Artificial Intelligence (AI), especially Large Language Models (LLMs), fuels digital transformation but raises computational demand, making data centers major energy users and emission sources. The Information and Communication Technology (ICT) sector contributes 2–4% of global emissions. Assessing AI's carbon footprint is vital for sustainability and policy planning. This narrative review systematically searched Scopus, Web of Science, PubMed, ScienceDirect, and Google Scholar from January 2019 to October 2025. Keywords related to AI, data centers, carbon, greenhouse gas emissions, and green AI were combined using Boolean operators. Included studies covered original research, reviews, and technical reports on measuring or mitigating AI's carbon footprint. Studies focused only on AI's environmental applications or hardware design were excluded. Data were qualitatively categorized and analyzed. AI's carbon footprint arises from the full model lifecycle—including embodied carbon, training, inference, and end-of-life—along with growing computational demand, hardware efficiency, and geographic carbon intensity variations. Currently, 369 generative models emit 10–18 million tons of CO_2 annually, projected to reach 245 million tons by 2035. Efficient architectures like Mixture-of-Experts (MoE) can reduce energy use tenfold; Tensor Processing Units (TPUs) are about 50% more efficient than GPUs; and data centers with a Power Usage Effectiveness (PUE) of 1.1–1.4 outperform those above 1.6. Geographic location can cause 5- to 10-fold differences in carbon intensity. Green AI techniques—such as knowledge distillation, quantization, data optimization, renewable energy, and tools like Code Carbon—can cut emissions by up to three orders of magnitude.

AI's growing carbon footprint challenges the shift to a low-carbon economy. Mitigation requires Green AI, transparency, standardized metrics, and efficient data centers. Sustainable AI depends on collaboration among researchers, industry, and policymakers, with sustainability as a key principle.

Keywords: Artificial intelligence, Carbon footprint, Data centers, Greenhouse gas emissions, Green AI.

How to Cite this Article: Abbasnia A, Bahlgerdi M, Farash Khayalo H, Bazrafshan E, Moeini Z. Assessing the carbon emissions of artificial intelligence: A narrative review of the environmental impacts of data centers and green AI strategies. *Journal of Research in Health and Medical Sciences*. 2024;3(1)

Introduction

Recently, Artificial Intelligence (AI) has evolved from an academic research domain into a transformative technology impacting nearly all aspects of human activity (1). This rapid advancement, however, has been accompanied by significant environmental costs (2). As AI models, particularly large language models (LLMs), become more complex and computationally intensive, the demand for large-scale processing power has grown exponentially (3). These substantial computational workloads are executed in energy-intensive data centers, which have emerged as major consumers of electricity worldwide (4, 5).

The energy required to operate such facilities directly translates into Greenhouse Gas (GHG) emissions, intensifying the global climate crisis (6-8). The Information and Communication Technology (ICT) sector, with AI as a rapidly expanding component, accounts for approximately 2–4% of global GHG emissions (9, 10). This share is comparable to that of the aviation industry and is projected to rise further as AI adoption continues to accelerate (11). Training a single large Natural Language Processing (NLP) model has been estimated to generate carbon emissions equivalent to the lifetime emissions of five automobiles. This amounts to roughly 300,000 kilograms of CO_2, equivalent to about 125 round-trip flights between New York and Beijing (12). These alarming statistics highlighted scientific and public attention to the carbon footprint of AI systems (13, 14).

This recognition has created a compelling duality. On one hand, AI holds considerable potential to support climate mitigation efforts (15), for instance, through optimizing power grids, improving climate modeling, and enabling the design of low-carbon infrastructure (16, 17). On the other hand, the technology itself has become an emerging source of carbon emissions (18, 19). The trend in which increasingly large and resource-intensive models are developed to achieve marginal improvements. Some researchers have called this approach Red AI, where accuracy gains come at the cost of massive computational consumption (20, 21). This term describes approaches in which enhanced model performance is disproportionately associated with higher energy use and elevated carbon emissions; for example, training state-of-the-art NLP architectures can emit up to 626,000 pounds of CO_2, comparable to the annual emissions of 125 passenger vehicles (22). In response, a counter-movement termed Green AI has emerged, emphasizing the development of models that balance innovation with computational efficiency and environmental sustainability (23-25). Advocates of Green AI argue for the incorporation of energy efficiency and carbon accountability as primary evaluation metrics, alongside accuracy. They also recommend mandatory reporting of energy consumption and carbon footprint to curb the unsustainable scaling of model sizes (24).

The carbon footprint of AI is a major challenge in the digital era. This review aims to synthesize existing evidence on greenhouse gas emissions from data centers throughout the AI lifecycle, considering technical, environmental, economic, and societal aspects.

This study aims to identify key factors influencing AI-related emissions and to evaluate practical strategies to reduce the environmental impact of AI technologies. Focusing on moving from high-consumption to sustainable architectures, this review presents a framework to expose hidden computational costs in AI systems. Unlike prior reviews on the ICT sector or hardware efficiency, this study synthesizes the full lifecycle carbon footprint of Generative AI and Large Language Models (LLMs), including embodied carbon, training, and inference. It also evaluates the environmental trade-offs of emerging architectures like Mixture-of-Experts (MoE), linking technical optimization with environmental policy.

Materials and Methods

This systematic narrative review was conducted using a structured search strategy. Scopus, Web of Science, PubMed, ScienceDirect, and Google Scholar were searched for articles published between January 2019 and October 2025. The initial search identified 450 records. After removal of 120 duplicates and screening of titles and abstracts for relevance, 55 studies were included in the final review.

Studies were prioritized if they reported quantitative data on carbon dioxide (CO_2) emissions or specific energy metrics, including Power Usage Effectiveness (PUE) and floating-point operations per second (FLOPS). Editorial commentaries without quantitative data were excluded.

The literature search used English keywords combined with Boolean operators (AND, OR), including artificial intelligence, carbon footprint, data center, greenhouse gas emissions, sustainability, Green AI, energy consumption, and machine learning.

Eligible studies included review articles, original research papers, and technical reports that focused on evaluating, quantifying, or mitigating the carbon footprint of artificial intelligence systems. Studies addressing only environmental applications of AI or limited solely to hardware-level assessments were excluded.

Results and Discussion

To ensure consistency, the following definitions are used throughout this review. Carbon Footprint refers to the total greenhouse gas emissions expressed as CO_2-equivalent (CO_2e). Embodied Carbon represents emissions from hardware manufacturing and infrastructure construction. Training refers to the model development phase, while Inference denotes the operational phase where the model generates predictions.

A review of the existing literature shows that the carbon footprint of AI is a multidimensional issue, stemming from various sources and shaped by multiple influencing factors. The main insights can be organized into several key dimensions:

1. Energy Use and Carbon Emissions in AI

The computational demand for advanced AI models has grown at an exceptionally rapid pace (26). According to a comprehensive analysis, the compute required to train large models doubles every three to four months, which is faster than Moore's Law (27). This exponential increase is primarily driven by the growth of generative AI models, which require billions of parameters and extensive training datasets to yield even marginal performance improvements. For instance, training a state-of-the-art NLP model can generate as much as 626,000 pounds (approximately 300,000 kilograms) of CO_2. It is equivalent to the annual emissions of 125 passenger vehicles or roughly 125 round-trip flights between New York and Beijing (22). A comprehensive analysis of 369 generative AI models found that they collectively consumed between 25 and 41 terawatt-hours (TWh) of energy and emitted 10 to 18 million metric tons of carbon dioxide (28). Meanwhile, the United States and China, both leading actors in this domain account for more than 99 percent of these emissions (Figure 1) (29). The ICT sector, including AI, currently accounts for approximately 2–4% of global greenhouse gas emissions, a share comparable to that of the aviation industry (31). This energy demand accumulates not only during the training phase but also often more significantly during inference. In fact, billions of daily requests can drive the model's lifetime energy consumption beyond that of training (32). Figure 2 illustrates the primary sources of energy consumption and carbon emissions. It shows that the adoption and energy intensity of specialized models are driven primarily by intrinsic industry characteristics rather than economic scale alone. Critical factors include data abundance and openness, privacy requirements, and task–model alignment. Healthcare, for example, despite its smaller market size compared to finance, supports highly energy-intensive models due to rich data availability. Likewise, the legal sector consumes approximately seven times more energy than education, despite similar market scales. This is because legal tasks such as case analysis and document review are particularly well-suited to large language models, while educational data tend to be sparse and heterogeneous.

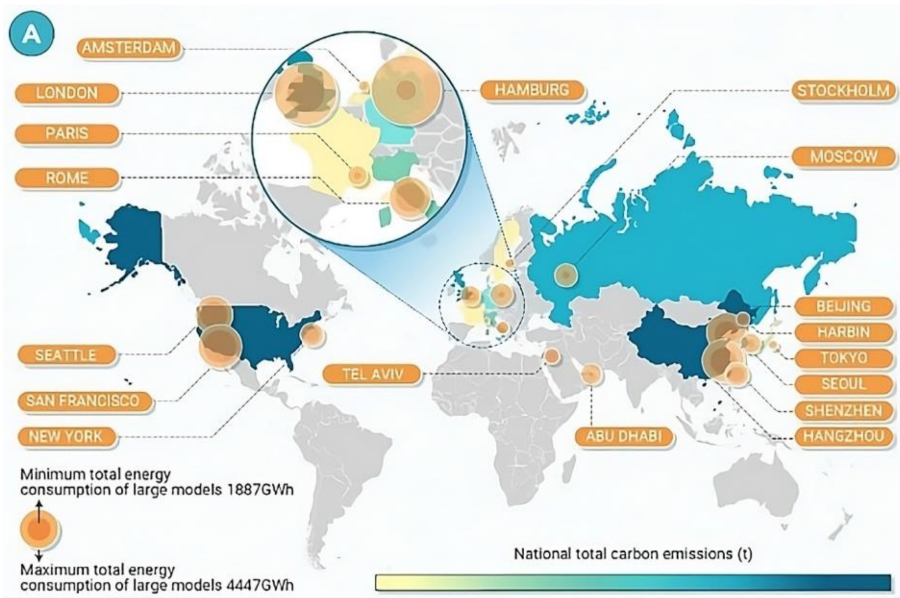

Figure 1. Regional energy consumption and carbon emissions from generative AI models. Bubble size represents energy use, and color intensity indicates emission levels. Adapted from Ding et al. (2025) (30)

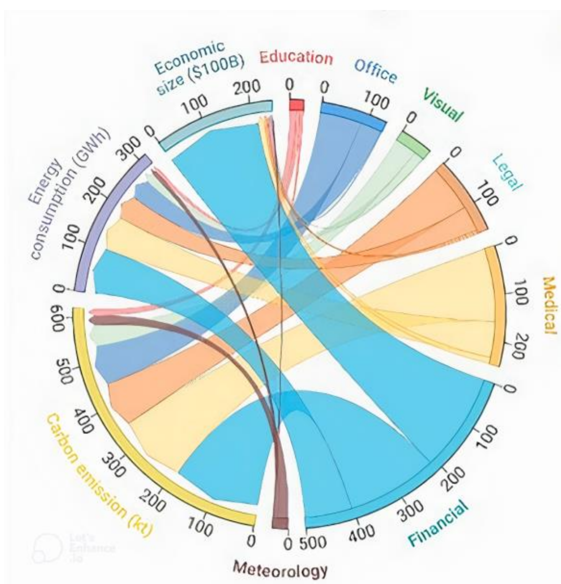

Figure 2. Energy consumption, carbon emissions, and economies of scale (left) alongside specialized model characteristics (right). Adapted from Ding et al. (2025) (30)

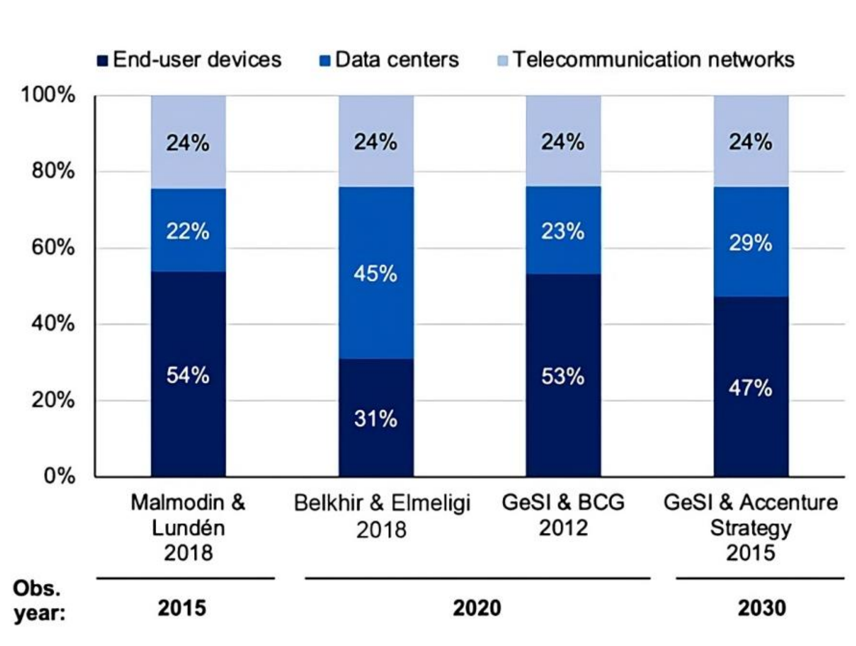

Figure 3. Contribution of ICT subsectors to the sector's total greenhouse gas footprint based on four independent assessments. Reprinted from Bieser et al. (2023) (31).

Conversely, as shown in Figure 3, Bieser et al. compared four studies regarding the proportion of total ICT-sector GHG emissions attributable to end-user devices, data centers, and telecommunication networks (31).

According to Figure 3, end-user devices are the dominant source of greenhouse gas emissions in the ICT sector, contributing 47–54% of the total, with substantial additional contributions from data centers and communication networks. Projections by Schneider Electric (2024) show that, under business-as-usual trends, generative AI could generate up to 245 million tons of CO_2-equivalent emissions annually by 2035, posing a serious threat to the achievement of global decarbonization targets (33-35). The rapid expansion of AI has rendered the transition from high-resource to efficient and sustainable paradigms more critical than ever. While giving rise to the "Green AI" movement, an emerging framework that treats environmental sustainability as a first-class objective on par with model accuracy. The computational demands of training advanced AI models have grown at an extraordinary pace (24, 36). According to one widely cited analysis, the compute required to train state-of-the-art models has been doubling roughly every 3–4 months, far outpacing the 18-month doubling cycle of Moore's Law (37, 38). An extensive study of 369 generative AI models estimated total energy consumption during training and inference at 25–41 TWh, with associated CO_2 emissions of 10–18 million metric tons (39). Over 99% of these emissions originate from facilities in the United States and China, the two leading countries in GAI development (40, 41). Continued growth at current rates would drive generative AI's annual CO_2 emissions to 245 million tons by 2035, creating a significant challenge for worldwide decarbonization goals (42-44).

2. Environmental Life-Cycle Assessment of Carbon Emissions from AI Systems

The carbon footprint of an AI system extends far beyond electricity consumption during inference; it encompasses the entire lifecycle, spanning four main phases: embodied (or latent) carbon, training, inference, and end-of-life.

Embodied carbon encompasses the greenhouse gas emissions associated with raw material extraction, hardware manufacturing (e.g., GPUs, TPUs, and servers), and data center construction prior to system operation (33, 45). Studies show that manufacturing a single high-end GPU can emit up to 100 kg of CO_2e, with emissions reaching millions of tonnes at the scale of data centers containing thousands of units (35). Although often overlooked, this production phase can account for 20–30% of the total lifecycle carbon footprint in large models (32).

The training phase is highly energy-intensive and has been a primary focus of Green AI research. Training a large language model such as GPT-3, for example, can consume more than 1,287 MWh of electricity, equivalent to the annual consumption of approximately 120 U.S. households, and produce over 550 tons of CO_2 emissions (34, 35).

However, the inference phase typically dominates total energy consumption, owing to the prolonged and continuous deployment of models in production environments. Although individual inference requests require minimal energy, their aggregation, often in the billions per day, means that inference can account for 80–90% of a model’s lifetime energy use (46). For instance, a widely used conversational system like ChatGPT may consume as much electricity in a single day as 17,000 average households (29).

Finally, the end-of-life phase presents significant challenges for recycling and disposing of AI hardware, with recycling rates remaining low owing to the incorporation of rare materials and intricate design features (47, 48). Globally, less than 20% of electronic equipment is properly recycled, leaving the majority to become e-waste that releases methane and heavy-metal contaminants (49). This lifecycle perspective highlights AI’s carbon footprint as a systemic issue needing intervention from hardware design to end-of-life management and recycling. (24).

3. Main Factors Influencing the Carbon Footprint of AI Systems

Research has shown that the carbon footprint of AI systems is highly variable and depends heavily on technical and infrastructure decisions. Four primary factors have been identified in this context:

• *Algorithm and Model*

Sparsely activated models, such as Mixture-of-Experts (MoE) architectures, contain billions of parameters. However, they activate only a small fraction of the network during each inference, achieving up to 10× lower energy consumption than dense models of comparable accuracy (50, 51). However, implementing MoE architectures presents specific trade-offs. While they significantly reduce computational costs, they require high memory bandwidth to load the large number of parameters, which can complicate deployment on standard hardware. Therefore, the energy savings are most effective when paired with specialized hardware optimized for sparse operations. This efficiency arises from a dynamic routing mechanism, where a lightweight gating network selects which expert subnetworks to activate for a given input. For instance, the Switch Transformer, with 1.6 trillion parameters, routes each token to just 2–4 out of 2048 experts, reducing inference-phase energy use to roughly one-tenth that of an equivalent dense model (52). Similarly, xAI's Grok-1 MoE model, with 314 billion parameters, activates only 25% of its weights per inference, yielding comparable energy savings and substantially lower operational carbon emissions (53).

Experimental studies show that this energy reduction is substantial in both the inference and training phases, since gradients are computed only along active paths. This is because it significantly lowers the memory required to store them (54). For instance, the Mixtral 8x7B model, which employs a Mixture-of-Experts (MoE) architecture and outperforms LLaMA-2 70B in accuracy, consumes just 1/5 of the training energy and 1/8 of the inference energy required by the dense model (55). This paradigm enables green scaling—better performance without proportional energy increase—unlike dense models, which nearly double energy use when parameters double (56).

Furthermore, combining MoE architectures with advanced optimization techniques, such as dynamic quantization and expert distillation, can improve inference energy efficiency by up to 15 times (57). When coupled with custom accelerators designed for distributed training and inference, these advances mark a fundamental shift from "red AI" to "green AI" (57). Consequently, distributed architectures have evolved from purely technical solutions into a core sustainability strategy for substantially reducing the gLocal carbon footprint of AI systems (24, 34).

• *Hardware*

While algorithmic optimization is vital, hardware largely determines AI's carbon footprint, with over 80% of inference energy and a significant portion of training energy tied to processor type and efficiency. Although high-performance GPUs such as NVIDIA's A100 and H100 deliver exceptional computational capability, they also exhibit considerable power demands (up to 700 W per unit). Their dense architectures and high clock frequencies generate substantial heat, thereby increasing reliance on energy-intensive cooling systems (46). Custom accelerators like Google's TPU v5 and Cerebras CS-2, optimized for matrix operations and low-precision quantization, can be up to 50% more energy-efficient than comparable GPUs (58, 59)

Furthermore, the shift from dense to sparse architectures has substantial implications at the hardware level. Accelerators like the Graphcore IPU and Groq LPU, supporting sparse matrix multiplication, allow MoE models to use only a fraction of computational units, cutting power consumption by up to 70%.(60). This architectural approach proves effective not only during inference but also during training, as it lowers the memory required for gradient storage and improves overall memory bandwidth utilization (61). Finally, the embodied carbon of hardware, emissions associated with raw material extraction, manufacturing, and transportation, is becoming an increasingly significant component of the AI lifecycle. Producing a single A100 GPU, for example, generates roughly 120 kg of CO_2-equivalent, which accounts for only about 10% of its total carbon footprint over an estimated 10,000 hours of operation (62). By contrast, custom accelerators manufactured using 3 nm process technologies and incorporating recycled materials can reduce embodied carbon by up to 30% (63). These advances, along with policies promoting longer hardware lifespans and improved e-waste recycling, mark a shift from a "performance at any cost" approach. This new focus on "sustainability by design" offers a practical framework to reduce AI's carbon footprint at the infrastructure level (64).

• *Data center efficiency*

Beyond the efficiency of individual chips and models, the broader infrastructure are crucial. Specifically, data center efficiency is commonly assessed using the Power Usage Effectiveness (PUE) index, defined as the ratio of a data center's total energy consumption to the energy consumed directly by its computing equipment (IT load). While a PUE of 1 represents ideal efficiency, the global average in 2023 was approximately 1.58 (65, 66). In contrast, modern cloud data centers operated by companies such as Google, Microsoft, and Meta achieve substantially higher efficiencies, operating up to 1.4 times more efficiently than conventional facilities with PUE values near 1.59; for example, Google reported an average PUE of 1.10 in 2024 (67). These improvements stem from advances like direct-to-chip liquid cooling, free-air cooling in cold climates, and AI-driven workload orchestration, which can cut cooling energy—usually 30–50% of total use—by up to 70%(68).

Lowering PUE significantly reduces carbon emissions. For example, cutting PUE from 1.8 to 1.2 in a 100-MW data center can save about 60,000 tons of CO_2 annually—equal to emissions from 13,000 cars (69). Leading companies use immersion cooling and waste-heat recovery to heat buildings or generate electricity. For example, Google's Finland facility recovers up to 90% of thermal energy, supplying heat to the local district network (70). Also, modular data-center designs and AI-based predictive load management allow operators to power down idle servers, improving peak-shaving performance by as much as 40% (71).

Finally, the standardization of PUE and its transparent reporting as part of sustainability metrics has become an industry-wide imperative. Initiatives like The Green Grid and ISO 30134 now require organizations to annually report their PUE metrics to improve data center energy efficiency and sustainability. In addition, platforms such as Google Cloud Carbon Footprint enable users to estimate carbon emissions based on actual PUE values (72). Such transparency not only drives competition among providers to lower energy consumption but also empowers customers to select cloud services with the smallest carbon footprint. Consequently, enhancing data center efficiency has evolved from a competitive advantage into both an environmental and economic necessity, playing a pivotal role in advancing the objectives of 'green AI (24, 34).

• *Geographic Location and Energy Mix*

The power grid's energy mix and data center location critically affect AI's carbon footprint. CO_2 emissions per kilowatt-hour can vary 5 to 10 times depending on fossil fuel or renewable energy dominance (73).

Training a large language model in a wind-rich region like Iowa (50 g CO_2/kWh) produces a much lower carbon footprint than in coal-dependent areas like Poland (750 g CO_2/kWh) (74, 75). This disparity accumulates during training and inference, where a popular model handling billions of daily requests can emit tens of thousands of tons less CO_2 annually when deployed in a low-carbon region. Leading technology companies have mitigated their environmental impact by adopting location-aware deployment strategies and purchasing renewable energy certificates (RECs). For example, Google has situated more than 90% of its data centers in regions with a high share of clean energy, such as Finland, Sweden, and

Quebec, and has achieved operational carbon neutrality through renewable power purchase agreements (PPAs) (76). Microsoft has similarly reduced emissions by up to 30% by time-shifting computational workloads to periods of peak solar and wind power generation. These practices show that data center location is a key sustainability strategy, capable of cutting AI's carbon footprint by up to 90% and advancing the shift toward green AI (74).

4. Operational Solutions for Sustainable AI

The Green AI movement promotes practical, multi-layered strategies to reduce AI's carbon footprint across its lifecycle, focusing on measurement, transparency, and optimization of models, data, and infrastructure (77). The foundational step in this process is the precise measurement and transparent reporting of energy consumption and associated carbon emissions. Several tools enable real-time assessment of AI's environmental impact by considering data center PUE, energy mix, and hardware. Code Carbon is an open-source Python library that estimates CO_2e emissions by tracking energy use and regional carbon intensity. ML CO_2 Impact offers a framework to estimate and compare emissions from training and running models. The Experiment Impact Tracker similarly records CO_2e emissions per experiment by monitoring hardware energy consumption in real time (78, 79).

At the model optimization level, techniques such as knowledge distillation, pruning, quantization, and sparse architectures play a pivotal role. For instance, knowledge distillation transfers knowledge from a large model (Teacher) to a smaller model (Student), potentially reducing energy consumption by up to 90% while incurring only a 1–2% loss in accuracy (80, 81). Similarly, data-centric AI approaches, which emphasize improving data quality, intelligent feature selection, and active sampling, can achieve comparable performance using approximately 50% less data, thereby lowering computational demands (82, 83). Beyond algorithmic strategies, adopting green infrastructure represents a highly effective operational measure. Deploying models in data centers powered entirely by renewable energy (e.g., Google's facilities in Finland or Microsoft's in Sweden) and utilizing advanced cooling technologies such as liquid or immersion cooling can reduce operational carbon footprints by as much as 95% (84). Integrating algorithmic optimization, transparency, and sustainable infrastructure can reduce AI's carbon footprint by up to 1000-fold, creating a strong foundation for sustainable AI development (34). These measures are not only environmentally necessary but also provide a strategic advantage for both industry and the scientific community (35, 64).

Conclusion

The results of this study demonstrate that the carbon footprint of AI is substantial and rapidly increasing. This growth results from rising computational demand, the AI lifecycle (from embodied carbon to end-of-life), and infrastructure factors like data center efficiency, energy mix, and hardware. As the global community endeavors to transition toward a low-carbon economy, the rapid expansion of such energy-intensive technologies poses a significant challenge to these efforts.

The findings distinguish "red AI," prioritizing accuracy at any cost, from "green AI," which balances efficiency and accuracy—offering both a challenge and an opportunity to reshape AI development. A key insight from this review is the critical role of choice. The carbon footprint of an AI model is not a fixed, inherent property but rather the outcome of deliberate decisions made by researchers and engineers. Evidence suggests that strategic selection of algorithms, hardware, data centers, and geolocations can reduce carbon emissions by up to three orders of magnitude .This highlights the promising potential to mitigate environmental impacts without impeding technological advancement. The Green AI movement provides a practical framework with strategies for measurement, model design, data management, and infrastructure optimization.

Despite these advances, significant challenges remain. The lack of transparency among major technology companies regarding the energy consumption of their operations complicates efforts to assess the carbon footprint of the industry as a whole. Additionally, the research community's predominant focus on the model training phase has resulted in relative neglect of the inference stage, which, at scale, can account for a substantial portion of energy use. This study highlights that sustainability should be considered a primary evaluation criterion for AI systems—alongside accuracy, speed, and cost. Collaboration among scientists, industry stakeholders, and policymakers is essential to establish standards for transparent reporting and to incentivize the development and deployment of green AI. Ultimately, the future of AI must not only be intelligent but also sustainable and responsible.

Recommendations for Future Work

Based on this review, three immediate actions are proposed. Standardization: Policymakers should establish standardized metrics for reporting AI carbon intensity (e.g., kg CO_2e per query). Transparency: Technology companies should disclose energy mixes and Power Usage Effectiveness (PUE) data for specific model training runs rather than relying solely on annual averages. Inference focus: Future research should prioritize optimization of the inference phase, as it accounts for the majority of lifecycle emissions in widely deployed models.

Conflict of Interest

No Conflict of Interest.

This manuscript was edited with the assistance of artificial intelligence tools to improve language clarity.

Reference

1. Úbeda-García M, Marco-Lajara B, Zaragoza-Sáez PC, Poveda-Pareja E. Artificial intelligence, knowledge and human resource management: A systematic literature review of theoretical tensions and strategic implications. Journal of Innovation & Knowledge. 2025;10(6):100809. https://doi.org/10.1016/j.jik.2025.100809
2. Doo FX, Kulkarni P, Siegel EL, Toland M, Yi PH, Carlos RC, et al. Economic and Environmental Costs of Cloud Technologies for Medical Imaging and Radiology Artificial Intelligence. Journal of the American College of Radiology. 2024;21(2):248-56. https://doi.org/10.1016/j.jacr.2023.11.011
3. Ji Z, Jiang M. A systematic review of electricity demand for large language models: evaluations, challenges, and solutions. Renewable and Sustainable Energy Reviews. 2025;225:116159. https://doi.org/10.1016/j.rser.2025.116159
4. Kaminski AM, Carloto FG, Jacques MM, Garcia VJ, Filho OC, Barriquello CH. Prospecting electricity theft through assignment of AI detected buildings to registered consumer units. Sustainable Energy, Grids and Networks. 2025;41:101609. https://doi.org/10.1016/j.segan.2024.101609

5. Mamkhezri J. Out of sight, out of mind? Consumer awareness and (mis)understanding of electricity bills in the United States. Energy Research & Social Science. 2025;127:104271. https://doi.org/10.1016/j.erss.2025.104271
6. Ricci M, Dominguez IP, Van Houtven S, Hristov J, Vandyck T. Pricing GHG emissions in agriculture: Accounting for trade and fairness for effective climate policy. Ecological Economics. 2025;239:108789. https://doi.org/10.1016/j.ecolecon.2025.108789
7. Wang Y, Liu H, Tong B, Zhu J, Mei T. Exploring greenhouse gas emission metrology research by bibliometric analysis. Measurement. 2025;256:118395. https://doi.org/10.1016/j.measurement.2025.118395
8. Yang Y, Tang Y, Sun S, Yang Z, Wang S, Zhang P, et al. Characterization of greenhouse gas emissions and water requirement of farmland in China's main grain-producing areas under future climate scenarios.

Agricultural Systems. 2025;225:104293. https://doi.org/10.1016/j.agsy.2025.104293
9. Shobanke M, Bhatt M, Shittu E. Advancements and future outlook of Artificial Intelligence in energy and climate change modeling. Advances in Applied Energy. 2025;17:100211. https://doi.org/10.1016/j.adapen.2025.100211
10. Yuan R, Ma T, Wang Y. Multiperspective analysis of emission responsibility in the global information and communication technology sector: New insights from value chain profits. Journal of Cleaner Production. 2025;517:145850. https://doi.org/10.1016/j.jclepro.2025.145850
11. Erdenetsogt O, Fu X, Wu H, Tsui KWH. Aviation industry growth in a landlocked developing country: The case of Mongolia. Case Studies on Transport Policy. 2025;21:101550. https://doi.org/10.1016/j.cstp.2025.101550
12. Martínez FS, Parada R, Casas-Roma J. CO2 impact on convolutional network model training for autonomous driving through behavioral cloning. Advanced Engineering Informatics. 2023;56:101968. https://doi.org/10.1016/j.aei.2023.101968
13. Amorim KA, Passos LP, Galdino MLS, Souza AHS, Ribeiro FB, de Oliveira Ferreira T, et al. Impact of carbon footprint of food production on consumer perceptions and attitudes: Insights from text highlighting. Food Quality and Preference. 2025;129:105498. https://doi.org/10.1016/j.foodqual.2025.105498
14. Zhang Y, Yang Y, Chai Y, Liang T, Li L. China's financial opening, green technical innovation and low-carbon economy. International Review of Economics & Finance. 2025;103:104546. https://doi.org/10.1016/j.iref.2025.104546
15. Ha LT. How Robotics & Artificial Intelligence development help the global mitigate energy crisis: Fresh insights from the R2 decomposed linkage method. Sustainable Futures. 2025;10:101044. https://doi.org/10.1016/j.sftr.2025.101044
16. Assareh E, Rismanchi B, Izadyer N, Bedakhanian A, Jamei E, Alagarasan JK. Artificial Intelligence-based multi-objective optimization of a solar-driven system for hydrogen production with integrated oxygen and power Co-generation across different climates. International Journal of Hydrogen Energy. 2025;184:151880. https://doi.org/10.1016/j.ijhydene.2025.151880
17. Kordi A, Mohebbi K, Bararzadeh Ledari M, Shirafkan SM. Transforming ports for a low-carbon future: Nexus modeling of hydrogen infrastructure, employment, and resource management in contrasting climates. Energy Nexus. 2025;19:100515. https://doi.org/10.1016/j.nexus.2025.100515
18. Aurangzeb I, Yoon JH. Artificial intelligence- and blockchain-enabled carbon emissions ledger system (AB-CELS) for sustainable construction processes. Automation in Construction. 2025;176:106286. https://doi.org/10.1016/j.autcon.2025.106286
19. Zhang T, Peng X, Zhong W, Shen F, Ding J. Automated machine learning-assisted enhanced product carbon footprint tracking and analysis in refinery industry: A graph-based life cycle assessment framework. Journal of Cleaner Production. 2025;514:145613. https://doi.org/10.1016/j.jclepro.2025.145613
20. M.S GP, H.N NK, Jain AK, Syed J, Baig RU. Convergence of improved particle swarm optimization based ensemble model and explainable AI for the accurate detection of food adulteration in red chilli powder. Journal of Food Composition and Analysis. 2025;143:107577. https://doi.org/10.1016/j.jfca.2025.107577
21. Niang N, Barriere T, Gabrion X, Holopainen S, Placet V. Predicting the deformability of natural short-fiber reinforced polymer composites through combined constitutive mathematical and AI-based modeling approaches. Composites Science and Technology. 2025;273:111353. https://doi.org/10.1016/j.compscitech.2025.111353
22. Strubell E, Ganesh A, McCallum A, editors. Energy and policy considerations for modern deep learning research. Proceedings of the AAAI conference on artificial intelligence; 2020. https://doi.org/10.1609/aaai.v34i09.7123
23. Borraccia S, Masciari E, Napolitano EV. Green metrics for AI: A hybrid strategy for environmental impact assessment. Array. 2025;28:100528. https://doi.org/10.1016/j.array.2025.100528
24. Verdecchia R, Sallou J, Cruz L. A systematic review of Green AI. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2023;13(4):e1507. https://doi.org/10.1002/widm.1507
25. Yokoyama AM, Ferro M, de Paula FB, Vieira VG, Schulze B. Investigating hardware and software aspects in the energy consumption of machine learning: A green AI-centric analysis. Concurrency and Computation: Practice and Experience. 2023;35(24):e7825. https://doi.org/10.1002/cpe.7825
26. Amdahl D. Navigating The Temporality Effect: NTNU; 2023.
27. Ganguli D, Hernandez D, Lovitt L, Askell A, Bai Y, Chen A, et al., editors. Predictability and surprise in large generative models. Proceedings of the 2022 ACM conference on fairness, accountability, and

transparency; 2022. https://doi.org/10.1145/3531146.3533229

28. Paccou R, Wijnhoven PF. Artificial Intelligence and Electricity: A System Dynamics Approach: Schneider Electric; 2024.
29. Wu L, Zheng Z, Qiu Z, Wang H, Gu H, Shen T, et al. A survey on large language models for recommendation. World Wide Web. 2024;27(5):60. https://doi.org/10.1007/s11280-024-01291-2
30. Ding Z, Wang J, Song Y, Zheng X, He G, Chen X, et al. Tracking the carbon footprint of global generative artificial intelligence. The Innovation. 2025;6(5):100866. https://doi.org/10.1016/j.xinn.2025.100866
31. Bieser JCT, Hintemann R, Hilty LM, Beucker S. A review of assessments of the greenhouse gas footprint and abatement potential of information and communication technology. Environmental Impact Assessment Review. 2023;99:107033. https://doi.org/10.1016/j.eiar.2022.107033
32. Luccioni AS, Viguier S, Ligozat A-L. Estimating the carbon footprint of bloom, a 176b parameter language model. Journal of machine learning research. 2023;24(253):1-15.
33. Gupta R, Nair K, Mishra M, Ibrahim B, Bhardwaj S. Adoption and impacts of generative artificial intelligence: Theoretical underpinnings and research agenda. International Journal of Information Management Data Insights. 2024;4(1):100232.https://doi.org/10.1016/j.jjimei.2024.100232
34. Luccioni S, Jernite Y, Strubell E, editors. Power hungry processing: Watts driving the cost of AI deployment? Proceedings of the 2024 ACM conference on fairness, accountability, and transparency; 2024. https://doi.org/10.1145/3630106.3658542
35. Patterson D, Gonzalez J, Le Q, Liang C, Munguia L-M, Rothchild D, et al. Carbon emissions and large neural network training. arXiv preprint arXiv:210410350. 2021.
36. Lin B, Zhou D. How does the explosive growth of AI affect China's power supply and demand: A scenario simulation based on the LEAP model. Renewable Energy. 2025;256:124485. https://doi.org/10.1016/j.renene.2025.124485
37. Ai Team NM. Tailoring AI for Turkish Law: Domain-Specific Fine-Tuning of Small Language Models for Legal Expertise. Procedia Computer Science. 2025;267:124-35. https://doi.org/10.1016/j.procs.2025.08.239
38. Liu J, Chen Y, Chen J, Li T. Research on collaborative decision-making model for site selection of wind and solar power stations based on AI large model and GAN. Microchemical Journal.2025;218:115406.https://doi.org/10.1016/j.microc.2025.115406
39. Friday B, Ali D. Balancing Innovation and Sustainability: Assessing the Impact of Generative AI on Energy Consumption. International Journal of Innovative Research and Development.2024. https://doi.org/10.24940/ijird/2024/v13/i9/sep24021
40. Lin J-W. The United States withdraws from the Paris Climate Agreement again: Influences and challenges of sustainable development in Chinese Taiwan. Journal of Environmental Management.2025;394:127365. https://doi.org/10.1016/j.jenvman.2025.127365
41. Tang D, Xi X, Li Y, Hu M. Regulatory approaches towards AI Medical Devices: A comparative study of the United States, the European Union and China. Health Policy. 2025;153:105260. https://doi.org/10.1016/j.healthpol.2025.105260
42. Wang J, Dong K, Dong X, Taghizadeh-Hesary F. Assessing the digital economy and its carbon-mitigation effects: The case of China. Energy Economics. 2022;113:106198. https://doi.org/10.2139/ssrn.4080737
43. Zhang W, Liu X, Wang D, Zhou J. Digital economy and carbon emission performance: Evidence at China's city level. Energy Policy. 2022;165:112927. https://doi.org/10.1016/j.enpol.2022.112927
44. Zhao X, Ma X, Chen B, Shang Y, Song M. Challenges toward carbon neutrality in China: Strategies and countermeasures. Resources, Conservation and Recycling. 2022;176:105959. https://doi.org/10.1016/j.resconrec.2021.105959
45. Chien F. The mediating role of energy efficiency on the relationship between sharing economy benefits and sustainable development goals (Case of China). Journal of Innovation & Knowledge.2022;7(4):100270. https://doi.org/10.1016/j.jik.2022.100270
46. Desislavov R, Martínez-Plumed F, Hernández-Orallo J. Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning. Sustainable Computing: Informatics and Systems. 2023;38:100857. https://doi.org/10.1016/j.suscom.2023.100857
47. Antony Jose S, Cook CAD, Palacios J, Seo H, Torres Ramirez CE, Wu J, et al. Recent Advancements in Artificial Intelligence in Battery Recycling. Batteries. 2024;10(12). https://doi.org/10.3390/batteries10120440
48. Lakhouit A. Revolutionizing urban solid waste management with AI and IoT: a review of smart solutions for waste collection, sorting, and recycling.

Results in Engineering. 2025:104018. https://doi.org/10.1016/j.rineng.2025.104018

49. Forti V, Baldé CP, Kuehr R, Bel G. The global e-waste monitor 2020. United Nations University (UNU), International Telecommunication Union (ITU) & International Solid Waste Association (ISWA), Bonn/Geneva/Rotterdam. 2020;120.

50. Chen Q, Zhang Z, Zhang Z, Zhang K, Li D, Wang W, et al. Distilled large language model-driven dynamic sparse expert activation mechanism. Applied Soft Computing. 2025;185:114037. https://doi.org/10.1016/j.asoc.2025.114037

51. Zhu S, Pan L, Jian D, Xiong D. Overcoming language barriers via machine translation with sparse Mixture-of-Experts fusion of large language models. Information Processing & Management.2025;62(3):104078. https://doi.org/10.1016/j.ipm.2025.104078

52. Fedus W, Zoph B, Shazeer N. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. Journal of Machine Learning Research. 2022;23(120):1-39.

53. Polignano M, Musto C, Pellungrini R, Purificato E, Semeraro G, Setzu M, editors. XAI. it 2024: An Overview on the Future of AI in the Era of Large Language Models. Proceedings of the 5th Italian Workshop on Explainable Artificial Intelligence, co-located with the 23rd International Conference of the Italian Association for Artificial Intelligence, Bolzano, Italy, November 26-27, 2024 (CEUR Workshop Proceedings; 2024. https://doi.org/10.1007/978-3-031-80607-0

54. Chen J, Zheng L, Yao Z, Wang D, Stoica I, Mahoney M, et al., editors. Actnn: Reducing training memory footprint via 2-bit activation compressed training. International Conference on Machine Learning; 2021: PMLR.

55. Jiang AQ, Sablayrolles A, Roux A, Mensch A, Savary B, Bamford C, et al. Mixtral of experts. arXiv preprint arXiv:240104088. 2024.

56. Kaplan J, McCandlish S, Henighan T, Brown TB, Chess B, Child R, et al. Scaling laws for neural language models. arXiv preprint arXiv:200108361. 2020.

57. Liu J, Tang P, Wang W, Ren Y, Hou X, Heng P-A, et al. A survey on inference optimization techniques for mixture of experts models. arXiv preprint arXiv:241214219. 2024.

58. Jouppi N, Kurian G, Li S, Ma P, Nagarajan R, Nai L, et al., editors. Tpu v4: An optically reconfigurable supercomputer for machine learning with hardware support for embeddings. Proceedings of the 50th annual international symposium on computer architecture; 2023. https://doi.org/10.1145/3579371.3589350

59. Lie K, Moyner O, Klemetsdal Ø, Skaflestad B, Moncorge A, Kippe V, editors. Enhancing performance of complex reservoir models via convergence monitors. ECMOR 2024; 2024: European Association of Geoscientists & Engineers. https://doi.org/10.3997/2214-4609.202437057

60. Sharma A. AI Accelerators for Large Language Model In-ference: Architecture Analysis and Scaling Strategies. arXiv preprint arXiv:250600008. 2025.

61. Shao H, Hu Y, Wang L, Song G, Waslander SL, Liu Y, et al., editors. Lmdrive: Closed-loop end-to-end driving with large language models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2024. https://doi.org/10.1109/cvpr52733.2024.01432

62. Pirson T, Bol D. Assessing the embodied carbon footprint of IoT edge devices with a bottom-up life-cycle approach. Journal of Cleaner Production. 2021;322:128966. https://doi.org/10.1016/j.jclepro.2021.128966

63. Lee BC, Brooks D, van Benthem A, Gupta U, Hills G, Liu V, et al. Carbon connect: An ecosystem for sustainable computing. arXiv preprint arXiv:240513858. 2024.

64. Henderson P, Li X, Jurafsky D, Hashimoto T, Lemley MA, Liang P. Foundation Models and Fair Use. Journal of Machine Learning Research. 2023;24(400):1-79. https://doi.org/10.2139/ssrn.4404340

65. Berger A. Artificial Intelligence Data Centers and United States Based Hyperscalers: Impacts and Solutions. 2025.

66. Hubert M-H, Le Texier T. Environmental Impact of Digitalization. 2025.

67. Wang F, Lv C. A data center expansion scheme considering net-zero carbon operation: Optimization of geographical location, on-site renewable utilization and green certificate purchase. Socio-Economic Planning Sciences. 2024;92:101789. https://doi.org/10.1016/j.seps.2023.101789

68. Shehabi A, Hubbard A, Newkirk A, Lei N, Siddik MAB, Holecek B, et al. 2024 united states data center energy usage report. 2024.

69. Masanet E, Shehabi A, Lei N, Smith S, Koomey J. Recalibrating global data center energy-use estimates. Science. 2020;367(6481):984-6. https://doi.org/10.1126/science.aba3758

70. Tervo S, Syri S, Hiltunen P. Reducing district heating carbon dioxide emissions with data center waste heat–Region perspective. Renewable and

Sustainable Energy Reviews. 2025;208:114992. https://doi.org/10.1016/j.rser.2024.114992

71. Tuan NA, Yonghan A. Edge AI for Smart Energy Systems: A Comprehensive Review. 2025.

72. Tyagi V, Tyagi P, Tyagi G, Goel PK. Cloud Computing for Carbon Monitoring and Analytics. Advanced Systems for Monitoring Carbon Sequestration: IGI Global Scientific Publishing; 2025. p. 281-300. https://doi.org/10.4018/979-8-3373-2091-5.ch012

73. Hussain M, Sultan M, Uzma F, Longsheng C, Malik MY, Butt AR, et al. A comparative analysis of renewable and non-renewable energy generation to relegate CO2 emissions and general costs in household systems. Environmental Science and Pollution Research. 2022;29(52):78795-808. https://doi.org/10.1007/s11356-022-21121-0

74. Chatzipanagi A, Jaeger-Waldau A, LETOUT S, MOUNTRAKI A, GEA BJ, GEORGAKAKI A, et al. Clean Energy Technology Observatory: Photovoltaics in the European Union-2024 Status Report on Technology Development, Trends, Value Chains and Markets. 2022.

75. Wojtaszek H. Energy Transition 2024–2025: New Demand Vectors, Technology Oversupply, and Shrinking Net-Zero 2050 Premium. Energies. 2025;18(16):4441. https://doi.org/10.3390/en18164441

76. Pollitt MG, Duma D, Covatariu A, Nillesen P. A global map of electricity and gas distribution network companies. Energy Policy Research Group, Working Papers. 2025(2519).

77. Ranpara R. Energy-Efficient Green AI Architectures for Circular Economies Through Multi-Layered Sustainable Resource Optimization Framework. arXiv preprint arXiv:250612262. 2025. https://doi.org/10.1007/s43621-025-01846-x

78. Lacoste A, Luccioni A, Schmidt V, Dandres T. Quantifying the carbon emissions of machine learning. arXiv preprint arXiv:191009700. 2019.

79. Lottick K, Susai S, Friedler SA, Wilson JP. Energy Usage Reports: Environmental awareness as part of algorithmic accountability. arXiv preprint arXiv:191108354. 2019.

80. Gou J, Yu B, Maybank SJ, Tao D. Knowledge distillation: A survey. International journal of computer vision. 2021;129(6):1789-819. https://doi.org/10.1007/s11263-021-01453-z

81. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv preprint arXiv:150302531. 2015.

82. Mindermann S, Brauner JM, Razzak MT, Sharma M, Kirsch A, Xu W, et al., editors. Prioritized training on points that are learnable, worth learning, and not yet learnt. International Conference on Machine Learning; 2022: PMLR.

83. Zha D, Bhat ZP, Lai K-H, Yang F, Hu X, editors. Data-centric ai: Perspectives and challenges. Proceedings of the 2023 SIAM international conference on data mining (SDM); 2023:SIAM. https://doi.org/10.1137/1.9781611977653.ch106

84. Hitesh V. Carbon Footprint of Digital Infrastructure: Data Centers, Cloud Computing, and the Path to Net Zero. 2025. https://doi.org/10.15680/ijirset.2025.1404001