Carbon Footprint of AI Agents: Review of Current Research

Francesco De Nicola, Lamyaa Rida
University of Padua
{francesco.denicola,lamyaa.rida}@studenti.unipd.it

Abstract

This paper examines the environmental impact of Artificial Intelligence (AI) by reviewing a range of recent studies and quantifying key metrics of energy consumption, greenhouse gas emissions, and resource depletion. We first describe how AI spending reached $154 billion in 2023 (Alzoubi and Mishra 2024), with the banking sector alone accounting for $20.6 billion and retail $19.7 billion. We then outline a lifecycle framework based on Life Cycle Assessment (LCA) to quantify first, second, and third-order impacts, and compare common electricity-consumption estimation methods (TDP-based, software power meters, and power-meter replication). We present detailed numerical examples—such as GPT-3's training cost of 1,287 MWh and 552 tons CO_2 e and a case study on Stable Diffusion (Berthelot et al. 2025) showing 14,400 kWh for v1-1 (256×256) training, among others. Next, we survey 55 Green AI initiatives grouped into six themes: cloud optimization, model efficiency, carbon footprinting, sustainability-based development, open source initiatives, and green AI communities highlighting potential reductions of up to 75% in energy use. Finally, we discuss methodological limitations, trade-offs between accuracy and efficiency, hardware constraints, and regulatory gaps. We conclude by proposing future research directions such as empirical validation of early-stage tools, fine-tuning cost integration, and strategies for circular hardware lifespans. Through this synthesis, we aim to guide AI practitioners, policymakers, and researchers toward sustainable AI practices that minimize negative environmental consequences without compromising innovation.

Introduction

The rapid deployment of Artificial Intelligence (AI) has brought significant innovations across various sectors. However, it has also introduced serious environmental concerns. The training and usage of AI systems, especially large-scale models like generative AI, involve intensive computational processes that contribute substantially to CO_2 emissions (Berthelot et al. 2025). These emissions are often overlooked in discussions of AI advancement but are now receiving urgent attention from researchers and policymakers.

Zewe (Zewe 2025) reports that the energy cost of training a single large AI model can be equivalent to the lifetime emissions of multiple cars. The deployment phase also carries environmental weight, particularly when models are served at scale (Zewe 2025). This necessitates a comprehensive understanding of the full lifecycle of AI systems and their cumulative carbon footprint.

This paper aims to uncover the environmental damage driven by AI agents and explore actionable mitigation strategies. By grounding our discussion in current literature and case studies, we propose a path toward more sustainable AI practices (Wu et al. 2022). Over the past decade, Artificial Intelligence (AI) has experienced dramatic progress, catalyzing a wave of technological and industrial transformation. In 2023 alone, organizations worldwide invested approximately $154 billion in AI technologies, with the banking industry leading at $20.6 billion and retail following at $19.7 billion (Alzoubi and Mishra 2024). While AI holds promise as a tool to address environmental challenges—particularly greenhouse gas (GHG) emissions—its own carbon footprint cannot be overlooked. This dual role, often described as a "double-edged relationship," underscores the necessity of both leveraging AI to combat climate change and mitigating AI's own environmental burden. The AI community has begun to confront these issues by quantifying impacts, proposing methodologies for assessment, and developing strategies for sustainable AI.

Lifecycle Framework for AI's Environmental Impacts

AI's environmental footprint can be classified into three tiers of impact (Alzoubi and Mishra 2024):

- **First-order (direct) impacts:** These impacts arise from the life cycle phases of equipment used for AI services—raw material extraction, manufacturing, transportation, usage, and end-of-life disposal. They affect multiple environmental indicators such as Global Warming Potential (GWP), water footprint, human toxicity, and Abiotic Depletion Potential (ADP). For instance, manufacturing accounts for roughly 75% of an iPhone 5's total emissions and about 40% of GHG emissions for a CPU-only data center in France.
- **Second-order impacts:** These represent the differential environmental outcomes between a reference application (without AI) and the AI-enhanced application. In other

words, the net gain or loss in environmental impact due to substituting or augmenting a process with an AI solution.

- **Third-order impacts:** These capture broader systemic or societal changes induced by AI adoption—such as shifts in business models, user behavior, or regulatory landscapes that indirectly alter environmental outcomes.

To quantify these impacts comprehensively, we adopt Life Cycle Assessment (LCA) as the primary methodological foundation. An LCA approach, aligned with ISO 14040 and 14044 standards (and ITU L1410 for ICT systems), enables quantification across multiple categories—ADP, GWP, and Primary Energy (PE)—and all life cycle phases, from raw material acquisition through end-of-life disposal. The LCA process entails four main steps: goal and scope definition, life cycle inventory (LCI) data collection, impact assessment based on LCI data, and interpretation of results. By selecting minimum recommended impact categories (ADP, GWP, PE), one can achieve a balanced and data-driven assessment even when certain data are scarce.

Estimating Electricity Consumption for AI Workloads

Accurate estimation of electricity consumption is essential for reliable GWP quantification. Traditional methods multiply Thermal Design Power (TDP) of hardware components by the training duration, but this TDP-based approach can overestimate consumption by as much as 6%. For example, during Stable Diffusion v1-1 training (256×256), a TDP-based estimate of 16,000 kWh exceeded the actual power-meter (PM) measurement of 15,100 kWh by approximately 6%. Software-based power meters—such as Intel's RAPL or NVIDIA's NVML—offer component-level granularity but can underestimate total energy by roughly 20% compared to external PM readings. To reconcile these discrepancies, a practical replication approach involves running a fraction of the training epochs on the target hardware while measuring actual consumption with an accurate external power meter. Because per-epoch energy consumption remains relatively constant, one can extrapolate the fractional measurement to the full training run, yielding reliable estimates without requiring continuous full-scale measurement.

Energy and Carbon Footprint of Large AI Models

Computational Costs and Energy Intensiveness

Deep Learning (DL) model training has exhibited exponential growth in computational requirements: computational power for advanced AI—including DeepMind's AlphaZero—has doubled approximately every 3.4 months in recent years, ultimately tripling between 2012 and 2018 as described by Loveleen Gaur (Gaur et al. 2023). As a result, training a single state-of-the-art model can be extremely energy-intensive, similar to the annual emissions of 123 midsize gasoline-powered passenger cars. Similarly, a single query to ChatGPT (inference) can consume about five times more electricity than a typical web search (Zewe 2025).

Data Center Impacts

Bashir (Bashir et al. 2024) reported that DL workloads necessitate large-scale data centers that can impose significant environmental burdens. In North America alone, data center capacity under construction rose from 2,688 MW at the end of 2022 to 5,341 MW by the end of 2023, with an additional 12,000 MW of co-location capacity anticipated. Globally, data center electricity consumption reached approximately 460 TWh in 2022—placing data centers between Saudi Arabia (371 TWh) and France (463 TWh) as a single country's consumption. Projections suggest data center consumption will approach 1,050 TWh by 2026, ranking fifth globally by energy use. Cooling demands exacerbate environmental stress: on average, each kilowatt-hour (kWh) consumed by a data center requires nearly two liters of water for cooling, much of which evaporates and becomes unavailable for other uses.

Embodied Carbon Footprint

Beyond operational electricity, embodied carbon in hardware manufacturing constitutes a growing share of AI's total footprint. For a typical CPU-only data center in France, manufacturing can account for around 40% of its life cycle GHG emissions. Notably, over 50% of Meta's total emissions derive from its value chain (Wu et al. 2022). As newer generations of GPUs and specialized accelerators proliferate, manufacturing impacts are expected to dominate unless mitigated through material efficiency, recycling, or hardware reuse.

Specific Model and Task Examples

Transformer-based natural language processing (NLP) models can be particularly carbon-intensive; for example, training a large transformer model on GPUs may emit on the order of 284,019 kg of CO_2 (Delanoë, Tchuente, and Colin 2023). Deep learning workloads, including optical neural network experiments, also consume substantial energy, and even inference phases contribute meaningfully to operational energy use. More broadly, AI's cumulative GHG emissions, resource depletion, and associated environmental pollutants contribute to climate change, water scarcity, biodiversity loss, and human health risks.

Projected Industry Trends

The European Union has warned that GHG emissions from the AI industry could increase sevenfold—potentially reaching 14% of global emissions—over the next two decades if unchecked. This projection underscores the urgency of both measuring and mitigating AI's environmental footprint.

Case Study: Stable Diffusion Service Lifecycle Assessment

A recent LCA-based case study done by Berthelot (Berthelot et al. 2025) on the Stable Diffusion service evaluated both training and inference phases over a one-year period (August 2022–August 2023). The model was originally trained on 32 nodes, each equipped with Nvidia GPUs. Regression analyses relating training steps (N) to energy consumption (kWh) yielded:

$$256\times256 \text{ images:} \quad E(N) = 5.26 \times 10^{-4}\, N + 2.01 \times 10^{-2},$$

$$512\times512 \text{ images:} \quad E(N) = 1.78 \times 10^{-3}\, N + 1.64 \times 10^{-2}.$$

Multiplying by the 32-node setup, estimated energy consumptions for different versions were:

- v1-1 (256×256, 237,000 steps): $E \approx 1.50 \times 10^4$ kWh.
- v1-1 (512×512, 1,940 steps): $E \approx 1.94 \times 10^3$ kWh.
- v1-4 (512×512, 225,000 steps): $E \approx 1.28 \times 10^4$ kWh.
- v1-5 (512×512, 595,000 steps): $E \approx 3.39 \times 10^4$ kWh.

For a one-year service period with an estimated 75 million site visits generating 150 million images (standard request: four 512×512 outputs per query), the total impacts were:

- **GWP:** 4.63×10^5 kg CO_2 eq (463 tons CO_2 eq).
- **ADP:** 4.77×10^0 kg Sb eq (kilograms of antimony equivalent).
- **PE:** 1.00×10^7 MJ.

A comparison of electricity estimation methods for v1-1 training revealed:

- **Software Power Meter:** 1.13×10^4 kWh.
- **External Power Meter (PM):** 1.51×10^4 kWh.
- **TDP-Based Estimate:** 1.60×10^4 kWh.

The TDP-based approach overshot the PM measurement by nearly 6%, while the software meter underestimated by about 25% when extrapolated to the full training run.

AI as a Double-Edged Sword

AI's potential for environmental benefit coexists with significant risks. On one hand, studies by PwC and Microsoft (2020) forecast that AI could boost global GDP by 3.1%–4.4% by 2030 and reduce GHG emissions by 1.5% – 4.0% through efficiency gains, predictive analytics, and optimized resource management. However, the share of digital activities in global CO_2 emissions increased from 2. 5% to 3. 7% between 2013 and 2018, growing approximately 4% per year. Without mitigation, DL-based AI could trigger cascading increases in energy demand, a 'butterfly effect' in emissions. For example, training a large transformer model has been estimated to produce 284,019 kg of CO_2. Thus, AI's net effect depends on implementing 'Green AI' practices to ensure that AI-driven gains outweigh AI-induced emissions.

Green AI Initiatives and Tools

We identified 55 initiatives aimed at reducing AI's environmental footprint. These initiatives fall into six principal themes:

Cloud Optimization

Cloud-based solutions that minimize power consumption include:

- **NVIDIA Green GPU Platform:** Can reduce power consumption by up to 50%.
- **Google AI Platform:** Claims up to 40% reduction in carbon footprint.
- **Microsoft Azure ML:** Reports up to 30% carbon footprint reduction.
- **Amazon SageMaker Green Inference:** Achieves up to 75% energy use reduction for inference tasks.
- **IBM Maximo Asset Efficiency Suite:** Delivers up to 20% energy savings in AI workloads.
- **Tencent Green AI Infrastructure:** Advertises up to 30% energy savings.

Model Efficiency

Techniques and toolkits that reduce computational requirements include:

- **NVIDIA DeepStream SDK:** Offers up to 40× faster inference and 30% energy savings.
- **Hugging Face Transformers Efficient Inference Mode:** Provides up to 4× faster inference with 4–8× smaller models, typically incurring just 1–2% accuracy loss.
- **OpenVINO Toolkit:** Enables up to 8× faster inference on Intel hardware.
- **Apple Core ML Model Zoo:** Facilitates up to 3× faster inference on Apple devices.
- **Intel AI Toolkit for Edge:** Achieves up to 5× faster inference on edge devices.
- **TensorFlow Lite:** Powers models on over 200 billion devices, lowering inference overhead.
- **Energy-Aware Deep Learning:** Demonstrated a 30% reduction in energy consumption for a large-scale language model.
- **AI Benchmark:** Over 1 million downloads and 100,000 active users, aiding researchers in efficiency testing.

Carbon Footprinting

Tools and calculators that estimate and optimize carbon footprints include:

- **Climate Impact Lab:** Reports 20% average carbon reduction for participated projects.
- **Google Cloud Sustainability Calculator:** Helped a large retail chain cut CO_2 by 20%.
- **Microsoft Azure Sustainable AI Calculator:** Aided a customer in reducing training time by 30%.
- **Carbon Tracker:** Used by over 100 organizations to monitor AI energy consumption, with participants noting 20% energy savings.
- **CodeCarbon:** Analyzes over 600 million lines of code to estimate emissions.
- **ECO$_2$AI:** A 3-bit GELU implementation that reduced CO_2 emissions by 17%.

Sustainability-Focused AI Development

Projects leveraging AI to advance environmental sustainability include:

- **EarthAI:** Aims to identify new lithium deposits with 50% less environmental impact than conventional methods.
- **Google AI for Social Good:** Mapped over 100,000 km² of coral reefs, aiding conservation efforts.
- **Climate TRACE:** Utilized by over 100,000 researchers, policymakers, and journalists for real-time GHG tracking.
- **Rainforest Connection:** Reduced deforestation by up to 90%, protecting over 20 million hectares of forest.
- **WWF Smart Fishing Initiative:** Achieved up to 70% reduction in illegal fishing, safeguarding billions of trees' worth of marine ecosystems.
- **Green Algorithms:** Introduced a "green" NLP algorithm that cut energy consumption by 70%, adopted by researchers in over 20 countries.
- **Aeva:** Reduced infrastructure inspection times by 80% using AI, lowering associated emissions.
- **DeepGreen:** Collaborated with over 50 entities to advance sustainable mining through AI.
- **Green.Dat.AI:** Funded by Horizon Europe to support AI for environmental monitoring.

Open-Source Initiatives

Community-driven tools and libraries include:

- **Green Software Foundation:** Over 100,000 downloads; participants report up to 20% reduction in data center carbon footprint.
- **Green AI Working Group:** Resource downloads exceed 100,000; fosters best practices.
- **Allen Institute for AI Climate Change Initiative:** Established in 2020 with a USD 100 million grant; related papers have over 1,000 citations.
- **Sustainable AI Toolkit:** Surpassed 100,000 downloads with 50 contributors.
- **ECO$_2$AI:** See above (carbon-footprinting tool).

Green AI Communities and Research

Academic and industry consortia promoting Green AI include:

- **EcoAI:** Published over 20 research papers on AI sustainability.
- **Green AI Foundation:** Website attracts more than 10,000 monthly visitors.
- **GreenAI UPPA Team:** Offers GreenAI Data Hub, GreenAI Toolkit, and GreenAI Notebooks.
- **EcoMENA:** Reached over 10,000 individuals through workshops and training.

Challenges to Sustainable AI

Despite the availability of tools and initiatives, several challenges persist:

- **Methodological Limitations:** LCA accuracy depends on data quality; manufacturing impact data for GPUs/TPUs are often sparse, and end-of-life assessments are rare. Reliable water-consumption data for cooling are also lacking.
- **Accuracy versus Efficiency Trade-off:** Highly efficient, smaller models typically concede 1 %–2% accuracy compared to larger counterparts, leading to dilemmas in critical applications.
- **Platform and Hardware Constraints:** Many optimization tools are tied to specific cloud or hardware vendors, limiting flexibility. In cloud environments, users may not control datacenter locations or hardware generations.
- **Transparency Deficits:** Proprietary algorithms and closed-source optimizations obscure true efficiency gains, and some service providers do not reveal how carbon calculations are performed.
- **Technical Expertise Requirements:** Deploying advanced profiling, power-meter replication, or custom LCA requires multidisciplinary expertise, creating barriers for smaller teams.
- **Scalability and Long-Term Impact:** While many tools show short-term gains, their long-term effects—especially as GPU demand surges—remain uncertain. Forecasting the environmental cost of rapid data center growth is an open problem.
- **Resource Constraints for Open-Source Efforts:** Volunteer-driven projects often depend on donations, leading to funding and sustainability challenges.
- **Allocation Issues:** Apportioning embodied and operational carbon across shared models—particularly when fine-tuning is widespread—remains methodologically contentious.
- **Regulatory Gaps:** No regulations yet explicitly govern AI energy usage. The EU AI Act mandates transparent accounting but stops short of setting energy or carbon thresholds. The GHG Protocol lacks specific provisions for ICT, and water footprinting remains largely unregulated.

Future Research Directions

We identify several fertile areas for further investigation:

- **Evaluating AI's Role in Green Technology Development:** Quantify how AI accelerates renewable energy integration, smart grids, and circular economy solutions.
- **Incorporating Economic Rationales:** Develop frameworks to assess AI's cost-benefit balance in environmental contexts, combining GWP, PE, ADP, and economic returns.
- **Design Thinking for Zero-Emission AI:** Extend ecodesign principles to AI workflows, from model architecture to end-of-life hardware disposal.

- **Participatory Research Approaches:** Engage stakeholders—engineers, policymakers, communities—in co-designing sustainable AI solutions.
- **Fine-Tuning Cost Integration:** Create standardized models to allocate carbon and energy costs for widely shared, fine-tuned base models.
- **Data Center Transformation Analysis:** Empirically assess environmental costs of the GPU-driven data center boom and potential mitigation (e.g., liquid cooling, server recycling).
- **Service-Level LCA Tools:** Develop user-friendly, cloud-native LCA calculators for deployed AI services, enabling real-time sustainability monitoring.
- **Empirical Validation of Emerging Tools:** As many Green AI tools are in early development, rigorous validation through empirical measurements is needed.
- **Energy System Modeling Improvements:** Address shortcomings in current energy models that under-represent AI workloads' demands, improving grid resilience planning.
- **R&D Funding for Sustainable Gen-AI:** Promote targeted research grants to optimize next-generation LLMs and multimodal models for environmental efficiency.

Conclusion

Tackling the environmental impact of Artificial Intelligence requires more than just technical innovation: It demands a change in how we design, deploy, and govern AI systems. Although the scientific community has made significant progress in understanding and measuring AI's carbon footprint, measurement alone is not enough. We must act on these insights with intention and coordination.

The path toward a more sustainable AI is not one that any single stakeholder can walk alone. Researchers must continue to explore efficiency-driven techniques, industries must adopt greener infrastructure and operational practices, and policymakers must establish clear guidelines and incentives for sustainable development.

Collaboration is the key to success. Only by fostering communication and cooperation across sectors—academia, industry, and government—can we ensure that AI evolves in harmony with the planet's ecological limits. A sustainable AI future is within reach, but only if we build it together.

References

Alzoubi, Y. I.; and Mishra, A. 2024. Green artificial intelligence initiatives: Potentials and challenges. *Journal of Cleaner Production*, 468: 143090.

Bashir, N.; Donti, P.; Cuff, J.; Sroka, S.; Ilic, M.; Sze, V.; Delimitrou, C.; and Olivetti, E. 2024. The Climate and Sustainability Implications of Generative AI. *MIT Climate and Sustainability Consortium*.

Berthelot, A.; Caron, E.; Jay, M.; and Lefevre, L. 2025. Understanding the environmental impact of generative AI services. *Communications of the ACM*. In press.

Delanoë, P.; Tchuente, D.; and Colin, G. 2023. Method and evaluations of the effective gain of artificial intelligence models for reducing CO2 emissions. *Journal of Cleaner Production*, 421: 139823.

Gaur, L.; Afaq, A.; Arora, G. K.; and Khan, N. 2023. Artificial intelligence for carbon emissions using system of systems theory. *Ecological Informatics*, 76: 102165.

Wu, C.-J.; Raghavendra, R.; Gupta, U.; et al. 2022. Sustainable AI: Environmental Implications, Challenges and Opportunities. In *Proceedings of MLSys*.

Zewe, A. 2025. Explained: Generative AI's environmental impact. *MIT News*.