An Analytical Framework for Assessing the Energy Consumption of Artificial Intelligence

1. Introduction

Artificial intelligence (AI) has become one of the most transformative technologies of the 21st century, with applications in healthcare, transportation, and other sectors. Alongside these advances, concerns have emerged regarding its environmental impacts, particularly energy consumption. Large language models (LLMs), such as GPT-3 and GPT-4, demand substantial computational resources, with energy requirements comparable to those of small countries (Agoston, 2024). While the economic and functional benefits of AI are well-documented, its environmental costs remain less explored.

Recent studies have quantified aspects of AI's energy footprint. Training GPT-3, for example, consumed approximately 1,287,000 kWh of electricity and generated 552 tons of CO_2 (Chow et al., 2024). However, existing analyses often focus on isolated stages of AI development, and technology companies rarely disclose detailed energy or emissions data (Fujianti et al., 2024). This fragmented understanding limits systematic evaluation and mitigation of AI's environmental impacts.

This study addresses this gap by adopting the three-phase framework proposed by Luccioni et al. (2022), segmenting AI systems into hardware manufacturing, model training, and inference. This preliminary section seeks to investigate and access data related to each of these phases of the generative AI lifecycle — hardware production, model training, and inference — since they represent the most energy-intensive stages of development. The analysis focuses primarily on electricity consumption, the dominant energy source across these phases, and also considers associated carbon dioxide emissions, which vary depending on the energy mix used: the more carbon-intensive the source, the higher the amount of CO_2 emitted per unit of energy produced. In doing so, the study contributes to Green AI debates emphasizing energy efficiency (Salehi & Schmeink, 2024) and promotes the adoption of sustainable AI practices.

The remainder of this article is organized as follows. Section 2 describes the theoretical foundations and methodology. Section 3 presents the framework and energy assessment results for the three phases. Section 4 discusses implications for efficiency and sustainability. Section 5 concludes and suggests directions for future work.

2. Materials and Methods

This section outlines the methodological procedures, theoretical foundations, and application examples used to develop a framework for assessing AI energy consumption. The investigation began by reviewing studies analyzing the full AI life cycle and associated environmental risks. This review revealed which stages were most frequently studied and supported by sufficient quantitative data. Consequently, the analysis focused on three critical phases, hardware manufacturing, model training, and operation (inference), which account for the majority of energy demand and environmental impact.

2.1 Scientific Methodology

A literature review was conducted to map empirical evidence on AI's energy consumption. Following Lima and Mioto (2007), the process employed structured, transparent, and replicable steps. Primary databases included Google Scholar and SciELO, complemented by reports from international agencies such as UNESCO, the European Union, and the World Economic Forum. This combination ensured broad coverage, regional representativeness, and integration of policy and regulatory perspectives.

Search terms were defined to capture the AI sustainability domain: *AI life cycle*, *AI environmental impact*, *Green AI*, and *AI for sustainability*. A two-stage selection process was applied: abstracts were first screened for relevance, followed by full-text review of promising studies. Inclusion criteria prioritized peer-reviewed studies published from 2020 onward with quantitative data.

2.2 Theoretical Framework

While the full AI life cycle includes stages such as raw material extraction, logistics, and end-of-life management, this study focuses on three core phases—hardware manufacturing, model training, and inference. This focus is both methodological and empirical: these phases are the most energy-intensive and best documented in the literature, offering consistent quantitative data on electricity use and emissions. In contrast, ancillary processes (e.g., supply chains or electronic waste) remain poorly quantified. Concentrating on these phases enables a transparent, comparable analysis that directly links design choices, computational demands, and operational practices to measurable energy outcomes. Together, they represent the central operational cycle of generative AI, where design, computation, and service delivery converge to shape its overall energy profile.

1. Hardware Manufacturing: Production of specialized computer chips, such as GPUs (Graphics Processing Units) and TPUs (Tensor Processing Units). These chips are designed for AI tasks and require a lot of electricity to make. Schneider et al. (2025) used a Life Cycle Assessment (LCA) — a method to measure environmental impacts across a product's life — and the Compute Carbon Intensity (CCI) metric, which shows emissions

per unit of computing power. They found that newer TPUs are about three times more energy-efficient than older versions.

2. Model Training: Training AI models means teaching them to recognize patterns in data. Luccioni et al. (2022) reported that training BLOOM (176B parameters) consumed over 433,000 kWh.

3. Deployment/Operation (Inference): Running trained models continuously consumes energy even during idle periods, known as readiness costs. Wu et al. (2021) and Schneider et al. (2025) highlight that constant operation accelerates hardware degradation and increases cumulative energy use.

2.3 Related Work

Research on AI's environmental impacts employs a variety of complementary methodologies, each addressing specific aspects of energy consumption, carbon emissions, and sustainability. LCA studies such as Luccioni et al. (2022) and Schneider et al. (2025) quantify energy use, carbon emissions, and hardware impacts, providing benchmarks for sustainable design and highlighting the energy-intensive nature of training and manufacturing AI systems.

Holistic system analyses examine AI's effects across entire infrastructures, considering both direct and indirect outcomes. Wu et al. (2021) and Gailhofer et al. (2021) demonstrate that optimizations spanning data processing, model training, and inference can significantly reduce energy consumption, while systemic or rebound effects may increase net environmental impacts.

Econometric and longitudinal analyses employ statistical models to assess AI's environmental effects over time and across regions. Wang et al. (2024) use SYS-GMM and Dynamic Panel Threshold Models to show that AI's carbon impact is non-linear, depending on development level, technology adoption, and regional energy infrastructure, highlighting the need for context-specific assessments.

Geographic and infrastructure modeling evaluates how the spatial location of AI resources affects environmental and social costs. Li et al. (2024) implemented Geographical Load Balancing to redistribute data center workloads based on energy efficiency and social equity, showing that strategic placement and load management can mitigate environmental footprints.

Finally, theoretical and regulatory analyses provide ethical and governance frameworks for sustainable AI development. UNESCO (2021) established global ethical principles incorporating environmental goals, while the EU Artificial Intelligence Act (2024) mandates life cycle assessments, energy efficiency standards, and transparency.

3. Discussion

This section presents the main findings from the literature review, organized according to the three-phase framework introduced in Section 2.

3.1 Energy Consumption in Hardware Manufacturing

The manufacturing of AI-specific chips, such as TPUs, is highly energy-intensive. Approximately 50% of total emissions during production stem from electricity use, while the manufacturing stage accounts for about 25% of the hardware's overall lifecycle emissions. Regional energy sourcing plays a crucial role: producing chips in areas where 90% of the electricity comes from clean sources can reduce emissions by up to 47%. These findings highlight that both hardware design and manufacturing location are key levers for lowering the carbon footprint of AI infrastructure.

3.2 Energy Consumption in Model Training

Training large-scale AI models demands substantial, concentrated energy resources. A well-documented example is BLOOM, an open-access language model with 176 billion parameters, developed under the BigScience collaborative project. Its final training consumed approximately 433,196 kWh over 118 days on a supercomputer in France.

By contrast, GPT-3, a similarly sized model with 175 billion parameters developed by OpenAI, has no officially published energy data. However, independent estimates suggest its training required around 1,287,000 kWh, resulting in a significantly higher carbon footprint — approximately 502 metric tons of CO_2 equivalent — mainly due to the higher carbon intensity of the energy grid used during training.

3.3 Energy Consumption in Operation (Inference)

During deployment, large models like BLOOM remain continuously active to respond to real-time user queries. Although hourly energy use during inference is lower than during training, continuous operation results in substantial cumulative consumption. In a monitored inference test using 16 GPUs over 18 days, BLOOM consumed 914 kWh, even with periods of low demand. The findings emphasize that idle energy use constitutes a non-negligible share of total operational energy consumption, underscoring the importance of optimizing inference workloads and scheduling.

4. Conclusion

This study systematically assessed the energy consumption of AI using a three-phase framework, hardware manufacturing, model training, and operational deployment, based on the approach proposed by Luccioni et al. (2022). The analysis reveals that hardware production contributes substantially to AI's lifecycle energy use, with emissions strongly influenced by the carbon intensity of the manufacturing region (Schneider et al., 2025). Model training emerges as highly energy-intensive, as exemplified by large-scale models such as BLOOM, which consumed hundreds of thousands of kWh during training (Luccioni et al., 2022), demonstrating the significant environmental impact of a single model. Operational deployment also incurs persistent energy costs, including readiness and idle consumption, which can be mitigated through strategic placement of data centers in regions with cleaner energy grids (Gailhofer et al., 2021).

These findings emphasize that AI's environmental footprint cannot be addressed by focusing on any single phase in isolation, since decisions in one area, such as model architecture, have cascading effects on hardware production and long-term operational energy demands. Two overarching factors — regional energy grid carbon intensity and the lack of corporate transparency regarding energy consumption — emerge as decisive for sustainability outcomes.

As a preliminary mapping, this study provides an integrated overview of AI's energy footprint and identifies the phases and variables most relevant for future assessment. Subsequent work will expand this framework through a targeted case study of a specific AI system, enabling more detailed and context-based evaluation of energy use and emissions, and supporting the gradual development of evidence-based strategies for sustainable AI.

References

AGOSTON, D. V. Of artificial intelligence, machine learning, and the human brain: celebrating Miklos Palkovits' 90th birthday. *Frontiers in Neuroanatomy*, v. 18, p. 1374864, 2024. Available at: https://doi.org/10.3389/fnana.2024.1374864. Accessed: Aug. 15, 2025.

CHOW, K.; TANG, Y.; LYU, Z.; RAJPUT, A.; BAN, K. Performance optimization in the LLM world 2024. In: *Companion of the 15th ACM/SPEC International Conference on Performance Engineering*, London, United Kingdom, 2024. p. 156–157. New York, NY, USA: Association for Computing Machinery, 2024. Available at: https://doi.org/10.1145/3629527.3651436. Accessed: Aug. 15, 2025.

CUNHA, F. A. N.; NASCIMENTO, M. J. L. do. Legal aspects of artificial intelligence in Brazil: a comparison between bill 2.338/23 and the European Union regulation. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, v. 11, n. 6, p. 3183–3192, 2025. DOI: 10.51891/rease.v11i6.19928. Available at: https://periodicorease.pro.br/rease/article/view/19928. Accessed: Aug. 15, 2025.

DESROCHES, C. et al. Exploring the sustainable scaling of AI dilemma: a projective study of corporations' AI environmental impacts. arXiv preprint arXiv:2501.14334, 2025. Available at: https://arxiv.org/pdf/2501.14334. Accessed: Aug. 15, 2025.

EUROPEAN UNION. *Artificial Intelligence Act*. Brussels: European Commission, 2024. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206. Accessed: Aug. 28, 2025.

FUJIANTI, L.; AZIZAH, W.; RIZAL, N. The double-edged sword of transparency: ESG disclosure's effect on competitive dynamics in manufacturing. *Edelweiss Applied Science and Technology*, v. 8, n. 6, p. 1735–1750, 2024. Available at: https://doi.org/10.55214/25768484.v8i6.2335. Accessed: Aug. 15, 2025.

GAILHOFER, Peter; HEROLD, Anke; SCHEMMEL, Jan Peter; SCHERF, Cara-Sophie; URRUTIA, Cristina; KÖHLER, Andreas R.; BRAUNGARDT, Sibylle. The role of Artificial Intelligence in the European Green Deal. European Parliament, 2021. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662906/IPOL_STU(2021)662906_EN.pdf. Accessed: Aug. 27, 2025.

LIMA, Telma Cristiane Sasso de; MIOTO, Regina Célia Tamaso. Procedimentos metodológicos na construção do conhecimento científico: a pesquisa bibliográfica. *Revista Katálysis*, v. 10, n. esp., p. 37–45, 2007. DOI: https://doi.org/10.1590/S1414-49802007000300004. Accessed: Aug. 28, 2025.

LI, P.; LIU, Y.; YANG, J.; REN, S. Towards socially and environmentally responsible AI. arXiv preprint arXiv:2407.05176, Apr. 23, 2024. Available at: https://arxiv.org/pdf/2407.05176. Accessed: Aug. 15, 2025.

LIGOZAT, Anne-Laure; LEFÈVRE, Julien; BUGEAU, Aurélie; COMBAZ, Jacques. Unraveling the hidden environmental impacts of AI solutions for environment life cycle assessment of AI solutions. *Sustainability*, v. 14, n. 9, p. 5172, 2022. DOI: https://doi.org/10.3390/su14095172. Accessed: Aug. 27, 2025.

LUCCIONI, Alexandra Sasha; VIGUIER, Sylvain; LIGOZAT, Anne-Laure. Estimating the carbon footprint of BLOOM, a 176B parameter language model. arXiv, 2022. Available at: https://arxiv.org/pdf/2211.02001. Accessed: Aug. 24, 2025.

SCHNEIDER, Ian; XU, Hui; BENECKE, Stephan; PATTERSON, David; HUANG, Keguo; RANGANATHAN, Parthasarathy; ELSWORTH, Cooper. Life-cycle emissions of AI hardware: a cradle-to-grave approach and generational trends. arXiv, 2025. Available at: https://arxiv.org/pdf/2502.01671v1. Accessed: Aug. 25, 2025.

SALEHI, S.; SCHMEINK, A. Data-centric green artificial intelligence: a survey. *IEEE Transactions on Artificial Intelligence*, v. 5, n. 5, p. 1973–1989, 2024. Available at: https://doi.org/10.1109/TAI.2023.3315272. Accessed: Aug. 15, 2025.

UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO, 2021. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000381137. Accessed: Aug. 28, 2025.

WANG, Qiang; LI, Yuanfan; LI, Rongrong. Ecological footprints, carbon emissions, and energy transitions: the impact of artificial intelligence (AI). *Humanities and Social Sciences Communications*, v. 11, p. 1043, 2024. DOI: https://doi.org/10.1057/s41599-024-03520-5. Accessed: Aug. 26, 2025.

WORLD ECONOMIC FORUM. *Global Risks Report 2024*. Geneva: WEF, 2024. Available at: https://www.weforum.org/reports/global-risks-report-2024. Accessed: Aug. 28, 2025.

WU, C.-J. et al. Sustainable AI: Environmental implications, challenges and opportunities. arXiv preprint arXiv:2111.00364, 2021. Available at: https://arxiv.org/pdf/2111.00364. Accessed: Aug. 15, 2025.