Energy Proceedings
Vol 55, 2025
ISSN 2004-2965

Carbon Footprint of AI Data Centers: A Life Cycle Approach[#]

Alexandre d'ORGEVAL[1, 2, 3*], Edi ASSOUMOU[2], Valentina SESSA[2], Ilknur COLAK[3], Stuart SHEEHAN[3], Quentin AVENAS[1]

1 IAC Partners, Paris, France.
2 Mines Paris - PSL, Centre de Mathématiques Appliquées, Sophia Antipolis, France.
3 Schneider Electric, France
*Corresponding author. Email: alexandre.dorgeval@minesparis.psl.eu

ABSTRACT

Data centers are energy-intensive infrastructures that generate, manage, and store information for our interconnected society. Models based on Artificial Intelligence (AI) such as ChatGPT are increasingly accessible, leading to significant energy consumption and associated carbon emissions.

Assessing the carbon footprint of AI data centers is essential for evaluating their environmental impact and, consequently, promoting responsible AI development and encouraging sustainable practices. In this work, we evaluate an AI data center's carbon footprint using a life cycle assessment approach. Unlike existing literature, we analyze the entire data center architecture rather than solely focusing on the servers' footprint. Additionally, we assess the impact of varying the electricity mix and extending the lifetime of servers, providing potentials for emission reductions.

Keywords: Data center, Life Cycle Assessment, Carbon emission, Sustainable AI

1. INTRODUCTION

Data centers are critical infrastructures supporting the exponential growth in data generation, particularly in Artificial Intelligence (AI) and High-Performance Computing (HPC). AI data centers are designed to handle high computation demands and feature advanced hardware like GPUs or TPUs, with high rack densities. These facilities are essential for a wide range of applications, from data storage and processing to the complex computations required by AI models training and scientific research. Global data creation is projected to rise from 1.2 trillion gigabytes in 2010 to 175 trillion gigabytes by 2025 [1], highlighting the need for robust data center operations. Additionally, reports by the IEA showed that the global energy consumption for data centers could more than double from 460 TWh in 2022 to 1000 TWh by 2026, with countries like Denmark potentially experiencing increases up to 15% of their total electricity use [2,3].

The importance of assessing the environmental impacts of data centers is underscored by their significant energy consumption and carbon emissions. Evaluating these impacts is critical not only for reducing carbon footprints but also for achieving sustainability goals set by major internet giants and data center operators. Companies like Google and Microsoft have pledged to match 100% of their hourly electricity consumption with zero-carbon energy purchases [4,5], and Amazon aims to be carbon neutral by 2040 [6]. Additionally, to get on track with the Net Zero Scenario defined by the IEA, emissions of data centers must be cut in half by 2030 [7]. These commitments reflect a broader industry trend towards sustainable practices, emphasizing the urgent need for comprehensive environmental assessments to guide these efforts and promote green technologies and practices.

The literature provides comprehensive methodologies for assessing the environmental impact of data center architectures, emphasizing both operational and embedded emissions.

Embedded emissions include the environmental footprint of manufacturing data center hardware. In Life Cycle Assessment (LCA) methodologies they are commonly used to evaluate these impacts. For example, the ACT framework proposed in [8] based on the work done in [9] for the case of processors provides a detailed model for estimating the embodied carbon footprint of processors and other key server components based on workload characteristics, hardware specifications, and semiconductor fab characteristics. This model has been the basis for calculating the embedded emissions for CPUs, GPUs, DRAM and storage in various studies [10,11]. However, current assessments often overlook other hardware components such as cooling systems,

[#] This is a paper for the 16th International Conference on Applied Energy (ICAE2024), Sep. 1-5, 2024, Niigata, Japan.

which can be significant contributors to energy consumption and emissions.

The operational carbon footprint focuses on the energy consumed during the use phase. Tools like Carbontracker [12] enable real-time monitoring of energy consumption and carbon emissions for training Deep Learning models. Studies highlight the importance of considering the carbon intensity of the energy source, with renewable energy sources significantly reducing operational emissions. For instance, [13] emphasizes detailed reporting of energy consumption and suggests strategies for reducing emissions, such as optimizing server utilization and improving cooling efficiency. However, these studies often focus on the carbon footprint of servers, excluding other significant contributors like cooling and power systems.

The literature advocates for a holistic approach, integrating both operational and embedded emissions. Studies such as [14] and [15] argue that achieving sustainability requires considering the entire lifecycle of data center components, including emissions from manufacturing, transportation, usage, and disposal. Innovative strategies such as carbon-intensity-aware job scheduling are also explored to reduce the overall carbon footprint [16,17].

In this paper, we focus on LCA for AI data centers. Evaluating their environmental impact is crucial to promoting responsible AI development and encouraging sustainable practices. As an example, we mention the examples of two Large Language Models (LLMs): GPT-3 and BLOOM. These AI models require vast computational resources, leading to substantial energy consumption and associated carbon emissions. In [18], it is shown that the carbon footprint of LLMs is heavily influenced by the energy source's carbon intensity. For instance, training GPT-3 resulted in emissions of approximately 552 tons of CO2eq, mainly due to the high carbon intensity of the energy grid used. In contrast, BLOOM's training emissions were significantly lower at 30 tons, benefiting from the lower carbon intensity of the French energy grid. These comparisons illustrate the potential for significant emission reductions by selecting energy-efficient infrastructures and cleaner energy sources.

In this paper, we propose using LCA to take a comprehensive approach to analyzing the carbon footprint of AI activities. The detailed nature of LCA facilitates a holistic understanding of AI-related carbon footprint assessments. In our study, we consider the broader implications of carbon footprint exercises, examining the impact from the perspective of entire data center architecture rather than solely focusing on the servers' footprint.

2. DEFINITION & SCOPE

The assessment was done based on a reference design, published by Schneider Electric, dedicated to AI applications [19]. The architecture is a 3.6MW data center, comprised of 2 IT rooms – one AI cluster, and one retrofitted room with an AI cluster installed with IT room, and equipped with Nvidia's H100 GPU. The methodology for this study is structured according to the phases of an LCA ensuring a comprehensive evaluation of the carbon emissions associated with an AI data center.

The system boundaries are defined as follows: the assessment encompasses the entire lifecycle of the data center, including manufacturing, operational, and end-of-life phases. The components considered within the boundaries include IT equipment (servers, storage, networking), cooling systems, power infrastructure, and building infrastructure. Components that comprise a data center are complex and usually the bill of material are not publicly shared, making LCA analysis a tedious process for researchers. However, companies have adopted various strategies to assess the carbon footprints of their products, by using methodologies to assess the environmental footprint of their components. Two main methodologies exist: 1) developed by the MIT (PAIA method) the Product Carbon Footprint (PCF) [20] which is used by companies such as HP, Apple or Dell, 2) developed by the PEP Ecopassport institution, the Product Environmental Profiles (PEP) [21] are used by companies such as Schneider, Legrand or ABB. In this work, the analysis integrates detailed emissions data for major components based on the PEP and PCF sheets available. For components with no PEP or PCF evaluation, proxies based on technological representativeness are utilized, such as using similar components from a competitor e.g., PDU from APC [22] replaced by this product from Legrand [23]. Additionally, the servers' values were built from data collected from the literature, as no PEP or PCF sheets for servers integrating GPUs have been found. The study assumes a 20-year lifespan for the data center with fixed replacement rate values for components, as provided by manufacturers.

For the geographical scope, our results were computed for operation in France. However some product sheets used a European mix for the use phase, and thus were adjusted to match France's electricity mix. We also computed the values for two other regions, using the mix of Europe and Germany.

3. LIFE CYCLE INVENTORY ANALYSIS

The inventory phase involves the collection and quantification of data on all material and energy inputs and outputs throughout the lifecycle of the data center components (manufacturing, distribution, installation, use, End-of-Life (EoL)). These values were collected from the collected PEP or PCF sheet, at the exception of the servers. For the servers, the values where built from the CPU and GPU results found in [8,9], and supported by vendor specific values for storage components, and DRAM values extracted from [8].

Manufacturing emissions

The reference design has two type of servers, one focused towards AI, based on NVIDIA's DGX pod configuration with the H100 GPU, and another more adapted to regular IT loads. For the AI optimized servers, the CO2 footprint for manufacturing is computed as in [8], that is:

$$E_{proc} = \frac{(FPA + GPA + MPA) \cdot A_{die}}{Yield} \quad (1)$$

with A_{die} the die area, FPA the carbon emission per unit area related to fab location and lithography, GPA emissions from chemicals and gases per unit area, MPA emissions from raw materials, and $Yield$ the fab yield.

For regular IT servers, which are assumed to be air-cooled and have no GPUs, the data is normalized to MW based on available data from Lenovo, HP, and Dell [24–27].

Operational emissions

To compute the servers' operational footprint, it is assumed that two states can be taken by a component: it is either at its TDP, or at its idle point, which gives the following formula based on [28]:

$$E_{Op} = CF * \sum_{i=1}^{n} TDP_i * \eta_i + (1 - \eta_i) * P_{idle} \quad (2)$$

with n the number of components, TDP the Thermal Design Power, η_i the utilization rate of the component when active – assumed at 60% at 100% load, P_{idle} the power consumed at idle, and CF the emission factor of the country's electricity mix. The energy mix considered in the first case is that of France. For the PEP sheets, use phases were adjusted to match France's emissions factor.

Furthermore, certain equipment was not considered in this initial assessment. This includes pumps, chemical dosing unit for cooling, storage tank, air/waste separator, cables, for cooling which were excluded due to data unavailability. For the servers, switches and connectors were excluded also, because of a lack of available data. Future iterations of this analysis will aim to incorporate these components to provide a more holistic view of the embodied emissions associated with data centers.

4. LIFE CYCLE ASSESSMENT & INTERPRETATION

The focus of the current analysis was limited to the CO2 footprint. The main reason behind these assumptions is that for GPUs, no emission factors other than CO2 were found at this stage in the literature. In contrast, for CPUs, studies such as the LCA done by Dell on a server or this study by the German Environmental Agency provides data for up to 5 additional impacts. PEP sheets data for cooling and power components can provide up to 8 additional impact categories. Furthermore, ongoing work at Boavizta aims to expand the assessment to include other emission factors in future analysis [29] potentially enabling multi-criteria assessments for entire AI data centers. Finally, the XRAF chiller from Schneider was replaced with that of BCW family because of data consistency.

Figure 1 shows the overall adjusted results, considering a 20-year data center lifetime, with emissions detailed by component category. The lifecycle phases are dominated by the use phase (29%) and manufacturing phase (70%).

To better understand what settings could impact the total carbon footprint, two cases are analyzed: varying the electricity mix and increasing the lifetime of the servers.

4.1. Electricity mix variation

The first use case examines the impact of varying the electricity mix on the use phase emissions of data

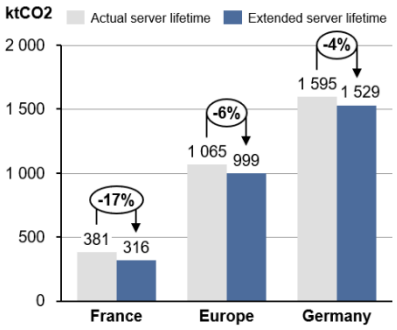

Figure 2 Carbon footprint of data centers for different countries when increasing server lifetime.

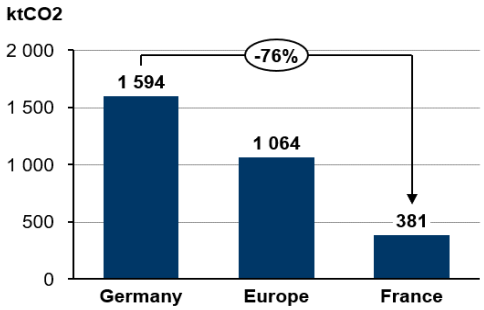

Figure 3 Carbon footprint for different electricity mix scenario

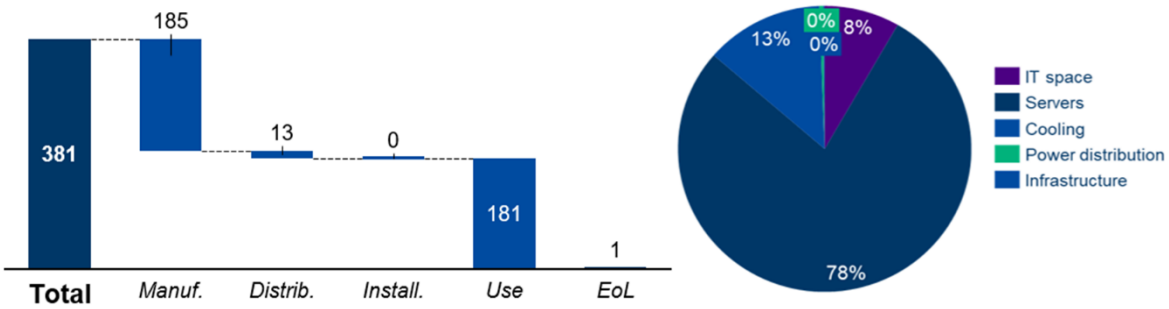

Figure 1: Carbon footprint per lifecycle state (left) and system category (right).

centers. The emission factors were adjusted to match those of France, Germany, and the average of the European Union, based on 2023 data from the Electricity Map website [30]. The results are illustrated in Figure 3. As anticipated, a lower energy carbon intensity leads to a lower overall carbon footprint. A data center located in France could potentially achieve a 3.7-fold reduction – or 76%, in carbon emissions compared to one in Germany, primarily due to France's electricity mix, which relies heavily on nuclear energy. Given the substantial investments by Internet giants in Power Purchase Agreements (PPAs) and Guarantees of Origin (GOs) – with Amazon and Meta being the top purchasers in 2023, accounting for 26% of all PPAs – and the increasing regulatory constraints on data centers in Europe, a viable strategy from a CO2 viewpoint might be to establish data centers in low-carbon regions such as France or the Nordic countries. While the electricity mix can influence the decision-making process, it is not the sole or decisive factor when selecting data center locations. Other critical factors such as the reliability of electricity supply, land acquisition costs, political stability, and regulatory environments also play significant roles in these decisions.

4.2. Extending the lifetime of components

The second use case aims to assess the impact of extending the lifetime of servers, with results illustrated in Figure 2. Here, the indicative lifetime of 5 years is extended by 50% to 7.5 years. Increasing the lifespan of components reduces the frequency of replacements, thereby decreasing embedded emissions. However, this comes with a trade-off: future generations of servers are likely to be more energy-efficient, potentially lowering the carbon footprint of the use phase. Consequently, hardware upgrades might therefore be more advantageous in regions with higher carbon intensity energy sources. However, this does not take into account potential additional carbon intensity of new processors. Extending the lifespan of servers results in significant emissions savings for data centers, with the benefits varying by location due to differences in electricity mix. For a data center in France, extending the server lifespan can save up to 19% of total emissions over a 20-year period. For an average European data center, the savings amount to 8%, while a data center in Germany sees a 5% reduction in total emissions.

By comparing the reduction in manufacturing and end-of-life (EoL) emissions to the use phase, it becomes evident that for a data center in Germany, extending the server lifetime is beneficial only if the next-generation GPU (assuming the same carbon footprint for the embedded emissions) is less than 6.1% more energy-efficient (respectively less 9.7% for a data center using the average European mix, and less than 45.7% for one in France). However, this analysis does not account for potential technological adaptations required or the effects on other environmental impacts (not yet computed for GPUs).

5. *CONCLUSIONS*

This comprehensive Life Cycle Assessment (LCA) of an AI data center, based on a Schneider Electric's reference design, is the first LCA done on an AI data center architecture. The study also highlights the critical influence of the electricity mix on carbon emissions, showing a potential 3.7-fold reduction through deployment in France compared to Germany due to France's reliance on nuclear energy. Moreover, extending server lifespans from 5 to 7 years can save up to 19% of emissions in France, 8% in Europe, and 5% in Germany over the entire lifecycle, yet this must be weighed against potential efficiency gains of newer hardware that could offset this lifetime prolongation. Future analyses should include all relevant components and expand beyond CO2 emissions towards other environmental impacts.

REFERENCE

[1] Daigle BR. Data Centers Around the World: A Quick Look 2021.

[2] Data centres & networks. IEA n.d. https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks.

[3] Electricity 2024 - Analysis and forecast to 2026 2024.

[4] Our Commitment to Sustainability. Google Sustainability n.d. https://sustainability.google/commitments/.

[5] Azure Sustainability—Sustainable Technologies | Microsoft Azure n.d. https://azure.microsoft.com/en-us/explore/global-infrastructure/sustainability.

[6] Cloud computing durable | Amazon Web Services. Amazon Web Services, Inc n.d. https://aws.amazon.com/fr/sustainability/.

[7] Net Zero Emissions by 2050 Scenario (NZE) – Global Energy and Climate Model – Analysis. IEA n.d. https://www.iea.org/reports/global-energy-and-climate-model/net-zero-emissions-by-2050-scenario-nze.

[8] Gupta U, Elgamal M, Hills G, Wei G-Y, Lee H-HS, Brooks D, et al. ACT: designing sustainable computer systems with an architectural carbon modeling tool. Proceedings of the 49th Annual International Symposium on Computer Architecture, New York, NY, USA: Association for Computing Machinery; 2022, p. 784–99.

[9] DTCO including Sustainability: Power-Performance-Area-Cost-Environmental score (PPACE) Analysis for Logic Technologies | IEEE Conference Publication | IEEE Xplore n.d.

[10] Faiz A, Kaneda S, Wang R, Osi R, Sharma P, Chen F, et al. LLMCarbon: Modeling the end-to-end Carbon Footprint of Large Language Models 2024.

[11] Li B, Roy RB, Wang D, Samsi S, Gadepally V, Tiwari D. Toward Sustainable HPC: Carbon Footprint Estimation and Environmental Implications of HPC Systems. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2023, p. 1–15.

[12] Anthony LFW, Kanding B, Selvan R. Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models 2020.

[13] Henderson P, Hu J, Romoff J, Brunskill E, Jurafsky D, Pineau J. Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning 2022.

[14] Wu C-J, Raghavendra R, Gupta U, Acun B, Ardalani N, Maeng K, et al. Sustainable AI: Environmental Implications, Challenges and Opportunities 2022.

[15] Strubell E, Ganesh A, McCallum A. Energy and Policy Considerations for Deep Learning in NLP 2019.

[16] Zhang G, Zhang S, Zhang W, Shen Z, Wang L. Distributed Energy Management for Multiple Data Centers With Renewable Resources and Energy Storages. IEEE Transactions on Cloud Computing 2022;10:2469–80.

[17] Zhao D, Zhou J. An energy and carbon-aware algorithm for renewable energy usage maximization in distributed cloud data centers. Journal of Parallel and Distributed Computing.

[18] Luccioni AS, Viguier S, Ligozat A-L. Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model 2022.

[19] EcoStruxure Data Center Reference Designs n.d. https://www.se.com/ww/en/work/solutions/for-business/data-centers-and-networks/reference-designs/.

[20] PAIA - Information & Communication Technology. Quantis n.d. https://quantis.com/who-we-guide/our-impact/sustainability-initiatives/paia/.

[21] Create a PEP n.d. http://www.pep-ecopassport.org/create-a-pep/.

[22] APC NetShelter Rack PDU Advanced, Metered, 3Phase, 22.1kW 400V 32A or 17.3kW 415V 30A, 48 Outlets, IEC309 - APDU10350ME | APC Francophone Africa n.d. https://www.apc.com/africa/fr/product/APDU10350ME/apc-netshelter-rack-pdu-advanced-metered-3phase-22-1kw-400v-32a-or-17-3kw-415v-30a-48-outlets-iec309.

[23] Technology S. PRO2™ HDOT® 11.0kW - 22.0kW (48-54) outlets Rack PDU. Server Technology n.d. https://www.servertech.com/power-distribution-unit-pdu/switched-pops-pdu/0u-vertical-2N34.

[24] pcf-thinksystem-sr250.pdf n.d.

[25] lca-poweredge-r6515-r7515-r6525-r7525.pdf n.d.

[26] Full_LCA_Dell_R740.pdf n.d.

[27] VGL0LLMZ.pdf n.d.

[28] Ji S, Yang Z, Chen X, Hu J, Shi Y, Jones AK, et al. Towards Data-center Level Carbon Modeling and Optimization for Deep Learning Inference 2024.

[29] Boavizta. Missions | Boavizta n.d. https://boavizta.org/missions.

[30] Live 24/7 CO_2 emissions of electricity consumption n.d. http://electricitymap.tmrow.co.