MLCA: a tool for Machine Learning Life Cycle Assessment

Clément Morand, Aurélie Névéol, Anne-Laure Ligozat

MLCA: a tool for Machine Learning Life Cycle Assessment

Clément Morand
Université Paris-Saclay, CNRS, LISN
Orsay, France
clement.morand@lisn.upsaclay.fr

Anne-Laure Ligozat
Université Paris-Saclay, CNRS, LISN, ENSIIE
Orsay, France
anne-laure.ligozat@lisn.upsaclay.fr

Aurélie Névéol
Université Paris-Saclay, CNRS, LISN
Orsay, France
aurelie.neveol@lisn.upsaclay.fr

***Abstract*—The on-going environmental changes challenge the ever-increasing use of digital technologies. Tools such as Green Algorithms or Carbontracker provide support for estimating the environmental impact of calculations (e.g., training a machine learning model). However, these tools only account for the dynamic consumption induced by calculations and only document carbon footprint while other types of impacts, such as resource depletion, are not evaluated. To provide a more comprehensive assessment of machine learning impact, we propose a modeling of graphics cards manufacturing impacts and a multi-criteria estimation tool called MLCA that accounts for the production impacts of hardware used to perform calculations. We evaluate MLCA through three reproduction studies thereby showing the validity of the assessments as well as the contribution of evaluating diverse impact categories over different life cycle phases. We hope this tool will help better understand the environmental impacts of Machine learning as a whole.**

***Index Terms*—Carbon Footprint, Climate change, Information Technology, Product life cycle, Machine learning, Green computing**

I. Introduction

Machine learning is a growing field in Information and Communication Technologies (ICT), and model training alone can have a high carbon footprint [1], [2]. Facing the ever-increasing computation demand of Artificial Intelligence (AI) [3], [4], researchers advocate for "Green AI" [5] to characterize the impact of AI, and promote more frugal AI research.

Life Cycle Assessment (LCA) is a method accounting for the impacts of products over their entire life cycle, from production to use through end of life. LCA can be used for evaluating impacts of AI solutions [6]. Following the pioneering work of Strubell *et al.* [1], several tools were developed to evaluate the environmental impacts of Machine Learning (ML) [7]–[9]. These tools focus on evaluating the carbon footprint of the energy consumption of the training phase of ML models.

Impact calculation only considers the energy consumption induced by training the models. The impacts of producing the hardware are not addressed in spite of their significance [10]. Furthermore, the share of embodied impacts is bound to grow with the shift towards less carbon-intensive electricity.

Limitations of LCA include the complexity of applying it to ICT when data availability is scarce [6], [11]. In addition, LCA does not account for structural or societal impacts such as rebound effect [12], [13] or ethical aspects [14]–[16]. Nonetheless, LCA offers a broad view of impacts going beyond use phase impacts and greenhouse gas emissions.

In this work, we tackle the question of the evaluation of the environmental impacts of ML methods. How can we accurately evaluate the environmental impacts of a series of ML experiments? How can we incorporate LCA considerations in a tool conducting such evaluations?

We present a tool named Machine Learning life Cycle Assessment (MLCA)[1] aimed at providing researchers with LCA estimates of computation impact that can be obtained independently from running calculations.

The main contributions of this work are as follows:

1) A modeling of graphics cards manufacturing impacts;
2) An estimation tool named MLCA that combines our modeling of graphics cards manufacturing and existing tools and methodology for the evaluation of servers manufacturing and energy use impacts. This tool evaluates multiple impact categories based on multiple phases of the equipment life cycle;
3) An evaluation of the usability and quality of MLCA.

First, Section II presents the state of the art and related work, then Section III details the methodology used for creating MLCA. Section IV evaluates MLCA on a series of case studies. Lastly, Section V discusses and Section VI concludes.

II. State of the Art

This Section presents the main concepts and existing work relevant to our study. Section II-A introduces efforts towards carbon footprint estimation for machine learning and highlights the potential contributions of LCA. Section II-B describes how LCA can be adapted to evaluate the environmental impacts of computer programs.

A. Carbon footprint of Machine Learning

After the high level of carbon emissions associated with training Natural Language processing models was reported [1], the community became aware of the need to report the costs associated with training models and striving for "Green AI" [5], which is becoming a research field [17]. In addition,

CM was supported by a doctoral grant from the École Normale Supérieure de Rennes. This work was also supported by the DIM RFSI CoCa4AI project.

[1] available under AGPL 3.0 license at https://github.com/blubrom/MLCA

it has been suggested that evaluation should encompass efficiency as well as raw performance and propose methods to evaluate the increase in carbon footprint per percentage of gain in precision [18]. For instance, [19] reports that half the costs of a state-of-the-art speech system are used to gain .3% Word Error Rate in one particular experiment.

Tools have been developed to evaluate the carbon footprint of computation. Surveys review the strengths and weaknesses of each tool [7]–[9]. There are two categories of tools. Measurement tools such as *CarbonTracker* [20], *CodeCarbon* [21] or *Experiment-Impact-Tracker* [22], use Running Average Power Limit (RAPL) tool to obtain live values for CPU and DRAM energy consumption [23] and the NVIDIA Management Library (NVML) tool [24] to get live consumption values for (NVidia) GPU. Estimation tools such as *Green Algorithms* [25] or *ML CO_2 Impact* [26] model the energy consumption of the processing units based on their Thermal Design Power (TDP). The energy consumption of the memory allocated to running computer programs is estimated by multiplying the quantity of memory by a consumption/GB factor [25].

The carbon footprint is computed by all tools as follows:

$$\text{GWP} = \text{CI} * \text{PUE} * (p_c + p_g + p_m) * t$$

where p_c, p_g, and p_m respectively refer to the power consumption of CPUs, graphics cards, and memory. The energy consumption of the hardware (either estimated or measured) is multiplied by a conversion factor accounting for the consumption of the rest of the datacenter, usually the Power Usage Efficiency (PUE). Finally, the total energy consumption is multiplied by the Carbon Intensity (CI) of the electricity powering computation, which leads to greenhouse gas emissions. The CI of a country depends on the share of low carbon energy sources so that the average CI in countries such as Iceland is close to 0 gCO_2 eq/kWh while it is approximately 400 gCO_2 eq/kWh in the USA and can be as high as 800 gCO_2 eq/kWh in the case of South Africa [25].

All extant tools focus on the carbon footprint incurred by the energy needed to run a computer program. However, this does not account for the production impacts of the hardware used to run the program. Given that the production of hardware can account for 40% of the total carbon footprint of a server over its entire life cycle [2], [27], it is reasonable to assume that they account for a similar share of the carbon footprint attributable to training an AI model. In fact, for Facebook, hardware production represent 30% of the total carbon footprint of "large scale ML tasks" [10].

Furthermore, producing partial environmental assessments may incentivize users to decrease greenhouse gas emissions by shifting impacts, whether in terms of life cycle phases or impact categories (reducing impacts from one category of impacts, typically carbon footprint, by increasing the impacts in one or multiple other categories). Impact shifting could, for instance, happen when replacing working but older hardware with new and more energy-efficient ones. The energy efficiency gains could reduce the carbon footprint, but discarding functional hardware increases the production of e-waste, and manufacturing new hardware increases the consumption of resources. In order to avoid impacts shifting, it is essential to use a methodology accounting for diverse environmental impacts over the whole life cycle (from production to use and end-of-life) of products, such as LCA.

B. Life Cycle Assessment

The LCA methodology is widely recognized with ISO standards (ISO 14040 and 14044). This methodology evaluates diverse environmental impacts over the whole life cycle of products. Figure 1 presents different phases of the life cycle of an item and different environmental impacts each phase can have. LCA is a multi-criteria evaluation of the environmental impacts (common metrics or criteria include Global Warming Potential, Human Toxicity, Resource depletion, Water use, Land use, Marine eutrophication, ...).

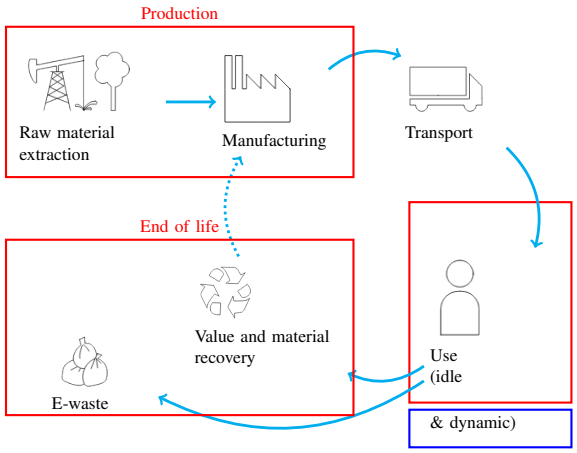
There are two types of LCA: *attributional* LCA and *consequential* LCA. Attributional LCA seeks to explain potential impacts attributed to a product or system. This approach assumes a static environment, and could try to answer the question: "What are the impacts of transporting 10,000 people a day by bus over 15 km?". Conversely, consequential LCA seeks to assess the impacts of change in a dynamic environment, with possible macro-economic responses to change. A consequential LCA could try to answer the question: "Given the current bus network, what would be the impacts of adding 1,000 new passengers a day?"

[6] proposes a framework for adapting the attributional LCA approach to evaluate AI tools. To evaluate an AI tools, one must consider not only the energy consumption induced by the training phase of the model but also the hardware manufacturing needed to produce the server on which this training phase takes place. Ideally, this analysis would also include information about the end-of-life of the hardware used, but this is a complex task because of the need for more available data on ICT end-of-life.

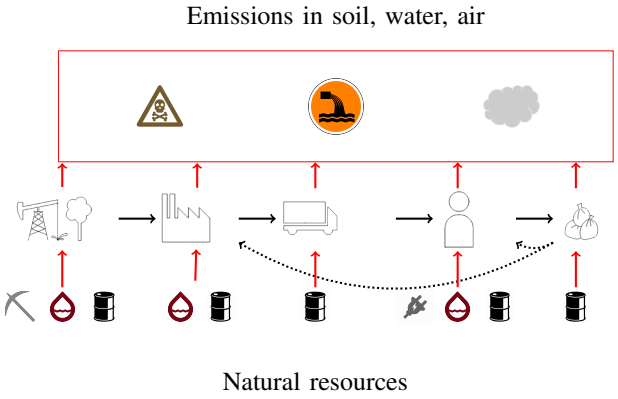
Furthermore, one should not only evaluate the impacts of producing an AI tool (i.e., training the model) but also consider the other life cycle phases of the model, such as the data collection required, the architecture search, and the inference phase. With its *Pragmating Scaling Factor*, Green Algorithms encourages its users to reflect on the production process of the computer program being evaluated and the potential multiple experiments needed to tune hyper-parameters of a model or to find the neural architecture of said model.

[28] proposes the boaviztapi tool for simplified LCA of servers[2]. It uses a bottom-up approach, estimating impacts for each component and aggregating them to obtain the total production impacts for the server. The study by Groger *et al.*, on which the boaviztapi tool is based, aims at creating a methodology for evaluating the environmental performance of Cloud services based on LCA methodology [29]. One important missing component of this tool and methodology

[2]accessible at https://github.com/Boavizta/boaviztapi



(a) different phases of the life cycle of hardware equipment. Use Dynamic is highlighted because most existing tools are focused on the dynamic use of the hardware induced by running a program.



(b) different types of impacts each phase of the life cycle can have. The barrel represents fossil fuels, the droplet represents water, the pickaxe represents metals and the plug represents electricity. The top row pictograms represent, from left to right emissions in soil, emissions in water and emissions in the air.

Fig. 1: Presentation of the possible scope for a LCA

is that it does not account for graphics cards being present in servers. However, most computationally heavy tasks, such as training machine learning models, use servers equipped with graphics cards or specialized servers such as Google's TPU.

Indeed, one crucial difficulty when applying an LCA approach in the ICT is the need for more available quality data [11]. This difficulty especially manifests when looking at graphics cards or TPUs, where no manufacturing firm provides insights into the production impacts of these devices. [30] conducted an LCA for comparing desktop computers with Raspberry PI devices with centralized servers in the context of a higher education class. To our knowledge, this LCA is the only available LCA considering a graphics card production.

In summary, several tools exist to evaluate part of the environmental impacts of computation. These tools are focused on the carbon footprint of the changed energy consumption induced by running a computation, which can be seen as a short-term consequential analysis. However, such an analysis does not account for a number of factors, such as the production of the hardware needed to run the computation. Oppositely, a tool such as proposed by [28] assesses the life cycle impacts of servers on multiple environmental criteria. However, these assessments do not include graphics cards and are unrelated to specific tasks. By combining both approaches, we can produce attributional LCA estimates for numerical computation.

III. Defining and Implementing an Estimation Tool for the Environmental Impacts of Computation

This section describes the methodology and implementation used to create *MLCA*, a tool aimed at providing researchers with attributional LCA estimates for numerical computation.

A. Goals and Scope

The objective of MLCA is to allow ML practitioners to estimate the environmental impacts/ benefits balance of their experiments and decide if it is worth pursuing them. The question MLCA tries to answer (e.g., the functional unit) is as follows: "What are the impacts of running program X on the hardware Y during Z hours?" where X, Y, and Z are parameters provided by the user. Program X can be training a Natural Language Processing (NLP) model.

The environmental impacts of running program X are considered to be those due to the hardware that it is run on, i.e., the impacts associated with the energy consumption of the hardware during the Z hours of the task, but also the impacts associated with the production of the hardware that can be attributed to using it for Z hours. An analysis of different phases of the life cycle of hardware Y is required to evaluate the different impacts each considered part of the life cycle has. Figure 1 presents different phases in the life cycle of hardware equipment and different types of impacts each phase can have.

Table I shows existing tools and the desired features for MLCA they take into account: Life cycle phases considered (manufacturing, distribution,...); diversity of impacts (not focused only on carbon footprint); graphics card support (since computing intensive programs such as training an AI uses servers with specialized hardware such as graphics cards or TPU). The choice for creating an estimation tool is driven by the fact that estimation tools can be used before running an experiment but also by the fact that estimates of the consumption based on the TDP of the processing units used (such as in Green Algorithms) might provide better quality estimates of the actual consumption than software measures (such as in CarbonTracker) [8].

Table I shows that combining Boavizta's tool [28] with Green Algorithm's methodology [25] is the best match since it allows to get both usage (Dynamic and Infrastructure) and production impacts, multiple impact indicators, and GPU support. Compared to Green Algorithms, ML CO_2 Impacts [26] only accounts for the energy consumption of one type

Outil	Life cycle phase considered						Multiple impacts considered	Estimates consumption	GPU support
	Ext.	Man.	Tra.	Uti.		EoL.			
				Infra.	Dyn.				
Green Algorithms	✗	✗	✗	✓	✓	✗	✗	✓	✓
ML CO_2 Impact	✗	✗	✗	✗	✓	✗	✗	✓	✓
CarbonTracker	✗	✗	✗	✓	✓	✗	✗	✗	✓
CodeCarbon	✗	✗	✗	✓	✓	✗	✗	✗	✓
Boavizta	✓	✓	✗	✗	✗	✗	✓	-	✗

TABLE I: Feature comparison of different existing tools to study environmental impacts of running computations

of hardware, GPUs. Our tool, MLCA, is therefore based on Boavizta's code[3] and methodology for evaluating hardware production impacts of a server [28] and models dynamic consumption in a similar way to [25].

In the end, the scope of MLCA spans the production and usage of the hardware used during the execution of program X. It does not include the distribution nor the end of life of the hardware as production and usage have the highest contribution to impact categories addressed in MLCA. End-of life is difficult to account for due to lack of data on the majority of e-waste fluxes [31]. The scope of our analysis also leaves out network usage for cloud-based server, data acquisition and storage as well as the storage of potential outputs of running program X (a trained model, for instance) and anything related to the data center building production and maintenance.

The environmental impacts are computed according to three different metrics. First, Global Warming Potential (GWP), measured in $kgCO_2$ eq for the emissions of greenhouse gas such as CO_2 [32]. Second, Abiotic resources Depletion Potential (ADP) measured in kgSb eq [33], [34], represents the use of mineral resources. This category of impacts is especially pertinent when considering ICT equipment since they use an important quantity of different (rare) metals to be manufactured. Third, for the total energy consumption, Cumulative Energy demand or Primary Energy (PE), measured in MJ [35]. PE can be interesting to show that some tasks, even though they have a low carbon footprint, can necessitate an important quantity of energy to be executed. Table II shows the different phases of the life cycle and the different impact categories considered in MLCA.

	GWP	ADP	PE	Human toxicity	Water Consumption	...
Production	✓	✓	✓	✗	✗	✗
Transport	✗	✗	✗	✗	✗	✗
Usage	✓	✓	✓	✗	✗	✗
End of Life	✗	✗	✗	✗	✗	✗

TABLE II: Summary of the scope of MLCA

B. Modeling of the production phase

The production phase considers the raw material extraction and manufacturing for the servers and the graphics cards used to run the computation. The modeling is inspired by [29] and follows the implementation of Boavizta for the servers [28].

a) Modeling of the servers: The modeling of the servers follows a bottom-up approach as described in [29]. A server is modeled as the sum of its components, namely CPU(s), RAM, SSD/HDD, Power supply, casing, and motherboard, assembled into a server in an assembly phase. The motherboards, HDD components, and assembly phase are supposed to have constant impacts. In contrast, power supplies impacts scale with weight; casing impacts depend on the server type (rack or blade), and each component that includes Integrated Circuit (IC) (CPU, RAM, and SSD) are modeled as having impacts varying with the area of the IC (die_{area}) they contain. For the RAM and SSD, the area of IC is estimated using a density factor, giving a die area per GB.

b) Modeling of the graphics cards: The modeling of the graphics cards adapts the modeling of CPUs from [29]. As for CPUs, a graphics card is modeled as a GPU (modeled by its die area), a quantity of memory and components present on all graphics cards such as the printed circuit board (PCB), gold for the connections, inductors, resistor, and capacitors present on the board. This modeling is translated in the impacts evaluation by adding together the impacts computed for the GPU die, the impacts computed for the memory, and a base impacts to account for all the other components as follows[4]:

$$\begin{aligned}\text{graphics card}_{impact} = {} & GPU_{die_{size}} * die_{impact_{per-cm^2}} \\ & + memory_{size} * memory_{impact_{perGB}} \\ & + base_{impact}\end{aligned}$$

The impacts of the GPU die are computed using the same impact factors per centimeter square of die as for CPUs. The impacts of the memory are computed as the impacts of the memory in the server, and the base impacts are obtained using the results of [30], which is, to our knowledge, the only LCA that comprises a graphics card.

In [30], results for scenario 2 are given for six servers. Each server contains two graphics cards, each with a die of .81cm^2. Dividing the total results for graphics cards in this scenario by twelve thus gives results for a single graphics card. To those results are removed, the estimated impacts of the dies of the GPUs using the $die_{impact_{per-cm^2}}$ factors. Since the graphics cards assessed in [30] do not comprise any memory, this is sufficient to obtain base impacts. The impacts in terms of ADP are obtained by converting Copper equivalent to Antimony equivalent using results from [33] where it is shown that one

[3]accessible at https://github.com/Boavizta/boaviztapi

[4]where impact $\in$ {ADP, PE, GWP}.

kg of Copper is approximately equivalent in terms of ADP to 0.02 kgSb eq. Finally, since results presented in [30] do not consider PE, the base impacts for CPU in terms of PE are used.

Once the production impacts of the hardware used are evaluated, they need to be allocated to the task under consideration (e.g., the training of the model).

C. Attribution of hardware production impacts to a specific computation

The results of the attribution of the total production impacts for a specific task are called *embodied impacts*. It is supposed that entire server modules are used for the task, meaning that the allocation only depends on use time and not on requested resources (number of CPU cores, for instance). The embodied impacts are computed by uniformly distributing impacts over the lifespan of the hardware, meaning that each hour of use is allocated the same share of the production impacts. The lifespan of the hardware consists of the total number of hours it can be used before it needs replacement. Base values from the Jean Zay cluster are used [2]. With a replacement rate of 6 years and 85% average usage, the total number of available hours are computed as follows: $h_t = 365 * 24 * \texttt{replacement rate} * \texttt{average usage}$. For a task spanning h_u hours, embodied impacts are then obtained with the formula

$$\texttt{Embodied}_{impact} = \texttt{Production}_{impact} \frac{h_u}{h_t}$$

where impact $\in$ {ADP, PE, GWP}.

Once the embodied impacts are evaluated, the use phase needs to be assessed.

D. Modeling of the use phase

The use phase consists of the energy consumption of the server for the task called *dynamic energy consumption*, to which is added the consumption of the data center used to render the server operational for the task. The dynamic energy consumption is assessed as in Green Algorithms [25] with modeling based on the TDP of the processing units. For a task spanning *hours usage* with n_p processing units and average usage of processing units u_p, dynamic energy consumption is assessed as follows:

$$E_{dynamic} = \texttt{hours usage} * \sum_{p \in \{\text{CPU,GPU}\}} (n_p * u_p * \texttt{TDP}_p) + \texttt{memory}_{size} * Power_{perGB}$$

In the same spirit as using a PUE to account for the server's energy efficiency (accounting for the infrastructure that allows the server to run our specific task), the dynamic energy consumption is multiplied by a *dynamic ratio* to obtain the total energy consumption *E*.

$$E = E_{dynamic} * \texttt{dynamic ratio} * 10^{-3}$$

The multiplication by 10^{-3} converts from Wh to kWh. This dynamic ratio is set to a default value calculated from the data gathered on the Jean Zay supercomputer in [2]. When running a series of experiments, they observed that the energy consumption was distributed as follows: 27kWh in "Infrastructure" mode (computing node off but the rest of the infrastructure running), 64 kWh in "Idle" mode (computing nodes and the rest on but no jobs running) and 109 kWh in "Production" mode (jobs running) for a total consumption of 200kWh. The dynamic ratio corresponds to the total consumption divided by the consumption in Production mode.

$$\texttt{dynamic ratio} = \frac{\text{TOTAL}}{Production} \simeq \frac{\text{TOTAL}}{\sum_{j \in \texttt{Jobs}} (E_{dynamic})_j} \simeq 1.834$$

This dynamic ratio corresponds to the average energy overheard for running the computing node. Its definition is really close to the definition of the PUE [36], [37], but it accounts for the fact that all of the work performed by the data center is not productive work (e.g., some of the work by the datacenter is only to keep the devices on).

In the end, energy-related impacts are computed as follows:

$$\texttt{Energy}_{impact} = E * impact_{perkWh}$$

where impact $\in$ {ADP, PE, GWP}. where, for instance, the $impact_{perkWh}$ corresponds to the CI if *impact* corresponds to GWP.

The total impacts evaluated by MLCA are then embodied impacts + energy impacts. In order to render the results of the evaluation understandable, they need to be put in perspective.

E. Putting impacts in perspective

While results are often put in perspective with the impacts of car or plane travel, Rasoldier *et al.* and Hauschild have highlighted the importance of putting environmental impact measurements in perspective with global sustainability objectives [13], [38]. This perspective allows to engage in a discussion on the "absolute" environmental sustainability of the solution under assessment [39] and emphasizes questioning whether an optimization effort is sufficient or not, or, as Hauschild puts it in [38], a new solution might be "Better, but is it good enough?"

Global sustainability objectives can come from the *Planetary Boundaries* (PB) framework [40], [41] or from international targets such as the Paris Agreements. There are multiple ways to divide a global objective [39]; among them, a uniform distribution, which supposes that every human being is allocated the same share of the global objective, was used as it is the simplest to compute and understand even if it supposes a uniform responsibility between countries. The results of the evaluation are therefore put in perspective with the an objective of reducing the average annual gross carbon footprint per capita to 2tCO_2 eq by 2050 to limit global warming to 1.5°C with no overshoot in 2100 (2t) [42], and with PB as devised in [41]. Results are presented in annual person consumption

in these two scenarios, meaning that if the evaluation leads to an estimate of 59tCO_2 eq for GWP and 1.2kgSb eq, it is also indicated as the annual emissions of 29 people (2t), or the annual emissions of 59 people (PB_{GWP}) and the annual resource extraction of 38 people (PB_{ADP})

F. Database

Since the estimates are based on the hardware configuration inputted by the user and in order to provide sensible estimates, one needs to be able to input the exact hardware configuration used by their experiments. To this end, a database was assembled, containing the specifications of over 150 CPUs and around 30 graphics cards. This database is based on the databases from Green Algorithms and Boaviztapi. Considering that Green Algorithms' database is mainly based on TechPowerUP databases[5] and presents CPUs (and) graphics cards by their names and TDP, and considering that the database from Boaviztapi is mainly based on the Wikichips website[6] and presents CPUs by their die size and architecture, these two sources of data were used to create a new and unified database for CPUs. Graphics card information is entirely based on the TechPowerUP database for graphics cards.

We have presented MLCA, a tool providing ML practitioners LCA estimates for their experiments. In the next section, we evaluate the quality of the estimates this tool produces, its sensitivity to parameter changes, and the pertinence of the new information MLCA brings over previous analyses.

IV. Evaluation

In order to evaluate the estimations produced by MLCA, its production impacts assessments are first compared with LCA of Dell servers. Then, results from the whole tool are compared with results from the assessment of the BLOOM model. Finally, experiments from the work of Strubell and colleagues on the impacts of NLP are re-explored to evaluate the pertinence of the new information provided by MLCA.

A. Evaluating estimations of the production impacts

In order to validate the embodied impact estimations, MLCA produces, its assessments of the production impact of servers in terms of GWP are compared with LCA results presented by Sphera for Dell on the R6515, R7515, R6525, and R7525 servers [43] [7]. Since there are few configuration differences between the R6515 and R7515 and between the R6525 and R7525, only the R6515 and R6525 servers are evaluated. For the manufacturing of the R6515, an estimate of 1200 kgCO_2 eq is obtained when the expected results stand at 1343 kgCO_2 eq. For the R6525, an estimate of 1600 kgCO_2 eq is obtained when the expected result stands at 1709 kgCO_2 eq.

Figure 2 provides a component-wise comparison of the estimation produced by MLCA with the expected results from the Dell LCA. It can be seen that MLCA underestimates the SSD impacts and produces a close estimate of the total manufacturing GWP impact. It can also be noted that a lower estimate that counterbalances the overestimate for the other components is obtained for the mainboard. This experiment confirms the adequacy of the results MLCA produces with expected results about the production impacts of a server in terms of GWP.

B. BLOOM carbon footprint estimates

After validating the assessment of production impacts produced by MLCA, we compare MLCA results on the evaluation of the training impacts of the BLOOM language model with the results from Luccioni *et al.* [2].

1) Gathering information about the setup: To replicate the experiments, we collect information on the duration of the training and hardware setup for the training phase from the paper [2]. According to Table 1, the training phase lasted for 118 days, 5 hours, and 41 minutes for a total of 1,082,990 GPU hours. Subsection 4.1 states that training used, on average, 48 computing nodes with eight graphics cards each. Combining this information with the real training duration, an estimate of the GPU time (1,089,670.4 hours) can be obtained, which is very close to the actual measured GPU time.

It is written in the paper that training took place on the Jean Zay supercomputer, using an HPE Apollo 6500 Gen10 Plus server[8]. The website of HPE indicates that this server uses AMD EPYC 7000 Series CPUs. Combining this information with information about the Jean Zay supercomputer obtained on the website of the IDRIS[9], it can be seen that only the **gpu_{p5}** partition uses such CPUs. For each of the 48 used nodes, the server configuration is thus:

- 2 CPUs : *AMD Milan EPYC 7543*
- 512 GB of Memory
- 8 *NVIDIA A100 SXM4 80Go*

2) assessment of the production impacts :

a) Comparing the estimated server footprint with the used value: It is indicated in [2] that a GWP of 2500 kgCO_2 eq is used in the BLOOM analysis, coming from the Product Carbon Footprint sheet of the closest found server, the HPE ProLiant DL345 Gen10 Plus server [10]. Table III presents the GWP production impact for the HPE's Apollo 6500 Gen10 Plus server estimated by MLCA. It can be seen that the value of 2,300 kgCO_2 eq is close to the value used in [2].

Indicator	GWP	PE	ADP
Unit	kgCO_2 eq	MJ	kgSb eq
Production	2,300	29,000	0.17

TABLE III: Estimated production impacts for the HPE's Apollo 6500 Gen10 Plus, computed by MLCA

[5]https://www.techpowerup.com/

[6]https://en.wikichip.org/wiki/WikiChip

[7]All Product Carbon Footprint and LCA produced on the different Dell products can be found at https://www.dell.com/fr-fr/dt/corporate/social-impact/advancing-sustainability/sustainable-products-and-services/product-carbon-footprints.htm#tab0=3

[8]https://buy.hpe.com/fr/fr/compute/apollo-systems/apollo-6500-system/apollo-6500-system/hpe-apollo-6500-gen10-plus-system/p/1013092236

[9]http://www.idris.fr/jean-zay/cpu/jean-zay-cpu-hw.html#gpu_p13

[10]https://www.hpe.com/psnow/doc/a50005151enw

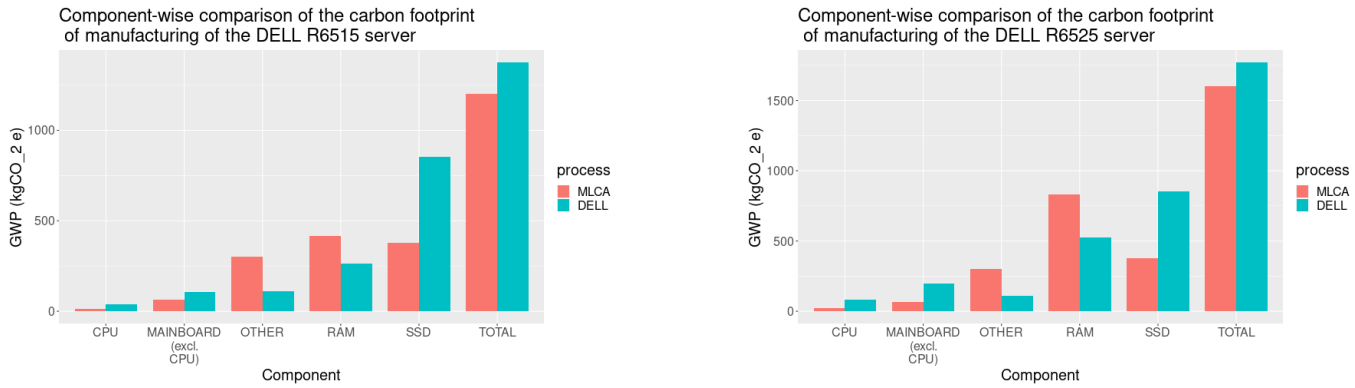


Fig. 2: Component-wise comparison of the GWP of manufacturing for the Dell R6515 (left) and R6525 servers (right)

b) Comparing the graphics card footprint with the chosen value: In subsection 4.1 of [2], it is stated that a value of 150 $kgCO_2$ eq for producing one graphics card is arbitrarily chosen. Given that in [30], a small graphics card production is estimated at emitting around 30 kg CO_2 eq, one could hypothesize that larger graphics card production impacts would be in the order of 50 to 150 kg CO_2 eq.

Indicator	GWP	PE	ADP
Unit	$kgCO_2$ eq	MJ	kgSb eq
Production	330	3,900	0.027

TABLE IV: Estimated production impacts for the NVIDIA A100 SMX4 80GB graphics card, computed by MLCA

Table IV presents the estimated production impact for the specific model used, the "NVIDIA A100 SMX4 80GB". MLCA estimates 330 $kgCO_2$ eq for the GWP of this graphics card's production. This estimate is mainly influenced by the quantity of memory on the graphics card with a carbon footprint of 290 $kgCO_2$ eq, leaving 40 $kgCO_2$ eq for the rest of the graphics card. This estimate of 40$kgCO_2$ eq for the graphics card without any memory is consistent with the values provided in [30]. The importance of the memory present on the GPU in its production impacts shows the need for an LCA of a modern graphics card used for High Performance Computing (HPC) to obtain good quality estimates.

3) Estimating the total impacts: Table V presents the estimated impacts of training the BLOOM model. In total, training the BLOOM model once is estimated to as much GWP as the annual emissions of 29 people (2t), or the annual emissions of 59 people (PB_{GWP}) and the annual resource extraction of 38 people (PB_{ADP}). Comparing the evaluation of embodied impacts in terms of GWP with the results from [2] (7.6tCO_2 eq for the servers and 3.6 tCO_2 eq for the graphics cards), it can be concluded that the main difference comes from the difference in graphics card production impact assessment. Indeed, MLCA evaluates the production of one NVIDIA A100 SMX4 80GB graphics card to 330 $kgCO_2$ eq in terms of GWP while a value of 150 $kgCO_2$ eq was chosen in [2]

For the dynamic consumption, an estimate of 23.7tCO_2 eq in terms of GWP is obtained, mainly due to the graphics cards (accountable for 22.4t; the only difference with the figure obtained in the paper being the slightly off conversion from real time to GPU hours) while the memory, not accounted for in the paper brings another 1.35tCO_2 eq.

	Indicator	GWP (tCO_2 eq)	PE (MJ)	ADP (kgSb eq)
Embodied	Servers	7	90,000	0.52
	Graphics cards	8.1	96,000	0.65
	Total	15	190,000	1.2
Dynamic	Servers	1.35	297,000	0.00128
	Graphics cards	22.4	4,920,000	0.0212
	Total	23.7	5,220,000	0.0225
Infra	Total	19.8	4,350,000	0.0187
	Total	59	9,800,000	1.2

TABLE V: Estimated production impacts for training the BLOOM model, computed by MLCA. Rows Dynamic presents the dynamic energy consumption related impacts, while row Infra presents the estimated impacts from the energy consumption of the infrastructure

Figure 3 compares the results MLCA produces with the results from [2]. As we can see, results for each stage are pretty similar, even if a higher estimate is obtained for embodied emissions due to a higher estimate of the production impacts of a graphics card. More surprisingly, one can note a significant difference in the infrastructure consumption-related impacts while the same figures should have been used to compute the infrastructure energy consumption in function of the dynamic energy consumption. This difference can be explained by the fact that the manuscript presents results on only a part of the infrastructure consumption (only the "Idle" mode is presented, and the "Infrastructure" mode, earlier mentioned in the manuscript, is omitted).

C. Impacts of Natural Language Processing

This section studies the training impacts of the models presented by Strubell *et al.* in [1]. As it is complex to find

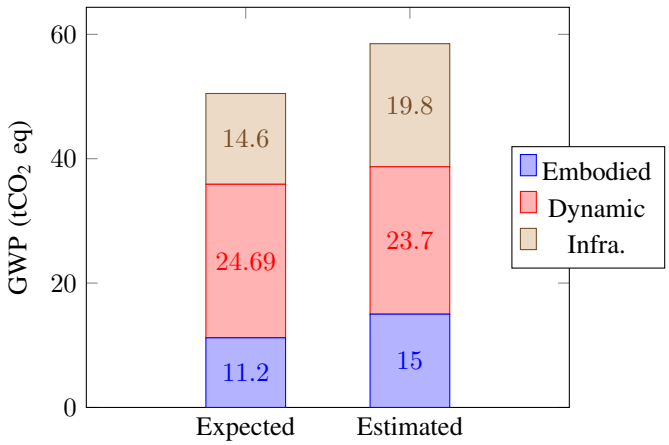


Fig. 3: Comparison of GWP estimations produced by MCLA (Estimated) with the estimated impacts presented in table 3 of [2] (Expected) over the different sources of emissions

data regarding the hardware used in TPUs, the experiments done with Google's TPU will not be included. After presenting information about the setup of the experiments, the first section will compare the results obtained on the energy consumption and related global warming potential only by MLCA with the results presented in [1]. A second section will integrate life cycle considerations and detail the full results obtained by MLCA.

1) Experiments setup and description: In [1], a PUE of 1.58 and a Carbon Intensity of 0.954 pounds CO_2 eq/kWh for American electricity production, which is equivalent to 432.72 gCO_2 eq/kWh, are used. Only the graphics card used is detailed for each model.

a) Hardware description: The papers do not detail the hardware for the ELMo and Transformer case. Only the used graphics cards are indicated for the transformer models. The following hypotheses on the hardware used for training these models are made:

- 32GB memory are required to train ELmo[11].
- 32GB memory should also suffice to train a Transformer with 65M parameters, and 64GB memory should be enough to train a Transformer with 213M. parameters[12]
- for the CPU, a CPU used in servers from the same period will be used. In particular, the CPU used in the Nvidia DGX-2H server that was used to train BERT in the experiments described in [1]: two *Intel Xeon Platinum 8174*.
- For the remaining hardware, the default values of MLCA are used.

For the BERT model, it is stated in [1] that it was trained using four Nvidia DGX-2H servers[13], with each server comprising two *Intel Xeon Platinum 8174* CPU and 1.5TB memory

[11]https://docs.deeppavlov.ai/en/0.9.0/apiref/models/elmo.html

[12]https://www.trentonbricken.com/TransformerMemoryRequirements/

[13]the specifications are available https://www.nvidia.com/conteat:nt/dam/en-zz/es_em/Solutions/Data-Center/dgx-2/dgx-2h-datasheet-us-nvidia-841283-r6-web.pdf

b) Approximating the GPU usage factor: In order not to overestimate the energy consumption, GPU usage factors are deduced from the power consumption indications provided in [1]. Using the ratio of average measured power consumption to total TDP of the used graphics cards, an approximation of the GPU usage factors can be deduced. These usage factors are presented in table VI (assuming that the vast majority of power draw comes from the GPUs)

model	estimated GPU usage
$\text{Transformer}_{\text{base}}$	0.70
$\text{Transformer}_{\text{big}}$	0.76
ELMo	0.69
$\text{BERT}_{\text{base}}$	0.75

TABLE VI: Estimated GPU usage factor when training the different models under consideration

2) Results:

a) Comparison with expected results: Table VII compares the results of the estimates produced by MLCA on two different scenarios with the results from [1]. The first scenario (match) uses the same PUE and CI as presented in [1] while the second (base) uses the base values of MLCA for the dynamic ratio and CI of the USA. Using the GPU usage ratio estimated in the previous section, the estimated energy consumptions (and thus carbon footprints of energy-related emissions) are very close to the expected values in the match scenario. Using the base values for MLCA yields a higher estimated energy consumption because it uses a dynamic ratio of 1.83 instead of a value of 1.58 when using the PUE. However, as the used CI is lower in the MLCA database than the value used in [1], this difference in estimated energy consumption does not lead to a higher estimated carbon footprint. It can be noted that adding the embodied impacts leads to a significant increase in the total estimated carbon footprint, especially on larger models that require more hardware to be run.

b) Integrating Life cycle considerations to previous analyses: To test the sensitivity of the results to changes in different parameters, variations of servers' lifespan and usage ratio, memory density, and location are explored. An interval of the possible output values is produced for the lifespan and usage ratio, as these parameters are easily bounded. It is assumed that no servers have a mean lifespan of less than one year and no more than eight years, and suppose that the servers never have a usage ratio of less than 10% and never have a higher ratio than 95%. Results will, therefore, be compared when using the default value and the values producing the highest (lowest lifespan and usage) and lowest (highest lifespan and usage) impacts. A scenario using the lifespan (3 years) and usage ratio (50%) of machine learning servers at Facebook described in [10] will also be explored.

For the sensitivity to changes in the estimated memory density, one scenario using the memory density used [29] will be tested. Memory density estimates can be an important factor when estimating the production impacts of servers using a considerable amount of memory. Memory density

model	expected energy (kWh)	estimated energy match (kWh)	estimated energy base (kWh)	expected GWP ($kgCO_2$ eq)	estimated GWP match ($kgCO_2$ eq)	estimated GWP base ($kgCO_2$ eq)	estimated GWP total ($kgCO_2$ eq)
$\text{Transformer}_{\text{base}}$	27	27	31	11.79	11	11	12
$\text{Transformer}_{\text{big}}$	201	203	235	87.09	87	87	90
$\text{BERT}_{\text{base}}$	1507	1500	1750	652.17	651	646	830
ELMo	275	281	326	118.84	122	121	130

TABLE VII: Comparison of the measures presented in [1] with estimates produced by MLCA. The *base* scenario uses the default dynamic ratio and CI for the USA in MLCA. The *match* scenario uses the PUE and CI presented in [1]. The *GWP total* presents results including embodied emissions in the base scenario while the other columns only include energyemissions.

is used to estimate the needed surface of IC to produce a fixed amount of memory. The default value of MLCA uses a high estimate, while the value used in [29] results in lower estimated memory production impacts. Sensitivity to changes in location is explored trough two different scenarios, one in France and one in the USA.

Figure 4 presents the results of the previously described experiments. As expected, changing the location to a country with a lesser CI can lead to significant reductions in terms of GWP. However, this does not hold for ADP, where the embodied impacts represent the vast majority of impacts. Neither does this hold for PE as changing location does not change the energy draw. These observations highlight the importance of using a multi-criteria approach to prevent impact shifting. The vast difference between the evaluation with the base parameters and the evaluation maximizing the embodied impacts allocation (top of error bars) shows the importance of keeping hardware for a long time and increasing server usage over buying new servers when trying to lower environmental impacts, especially in terms of resource depletion (ADP).

V. Discussion

A. About the validity of the tool

Our evaluation of MLCA showed the validity of the estimates of the production impacts of servers in terms of GWP (section IV-A), of estimates of the energy consumption in the use phase (section IV-C) and of the overall assessment (section IV-B). The sensitivity of MLCA to diverse parameter changes has been explored in section IV-C. These diverse experiments demonstrate the validity and usability of MLCA in diverse scenarios.

The results produced by MLCA are put in perspective with global sustainability scenarios to place computing activities within limits on their environmental impacts. This perspective highlights that continued slow growth, stagnation, or even a slight decrease of the impacts over time are not sustainable trajectories for ML.

Furthermore, proposed methods for stabilizing the carbon footprint of ML rely on frequent hardware updates to increase energy efficiency [44]. These methods will inevitably generate impacts shifting from the use phase to the production and end-of-life phases with more hardware being produced and decommissioned, but also from the global warming impact category to impacts categories like resource depletion or toxicity. A multi-criteria, multi-life-cycle phases analysis as implemented in MLCA highlights such impacts shifting.

Still, our tool and methodology have some limitations and uncertainties. First, the evaluation methodology was unable to validate results on indicators other than GWP as no known previous work explored the impacts of AI systems in terms of ADP or PE. Second, there are some limitations and uncertainties in the assessments produced by MLCA, ranging from uncertainties and data quality to limitations of scope and methodological limitations.

B. Limitations and uncertainties

a) Uncertainties and data quality: Production impacts of IC have been shown to vary with the technological node, with smaller nodes (≤ 14nm) having increasing impacts [45]. Since GPUs tend to be produced with smaller nodes, their impacts might be higher than the ones currently evaluated in MLCA. Still, the order of magnitude of the impacts remains consistent with current data.

The sensitivity analysis shows that the used memory density factor can significantly impact the results. Adapting density factors based on memory technology (the memory embedded in the graphics card does not use the same technology as DRAM) would help improve the precision of production impacts estimates. Estimates of graphics card production impacts could also be much more precise if LCAs for modern graphics cards were available. Then, it might also be possible to differentiate between graphics cards using PCIe connectors and those using SXM modules.

b) Limitations of the scope: The analysis conducted in MLCA does not include the distribution or the end-of-life of the hardware. Distribution is often not considered because it has only minor impacts relative to the other phases. For the end-of-life, this phase is frequently omitted as it is supposed that it does not emit many greenhouse gases. However, the end of life can also have major impacts in terms of toxicity, for instance [46]. The significant challenge to considering the end of life is the lack of available information. The lack of available data also prevented using other impact metrics, such as water consumption.

Data storage is not included, but as it is mostly independent of computation, combining MLCA with another tool specialized for this task seems to be the best possible solution. Data transfers are not currently included as their impacts are expected to be small compared with the impacts of the computation themselves. Datacenter building, maintenance,

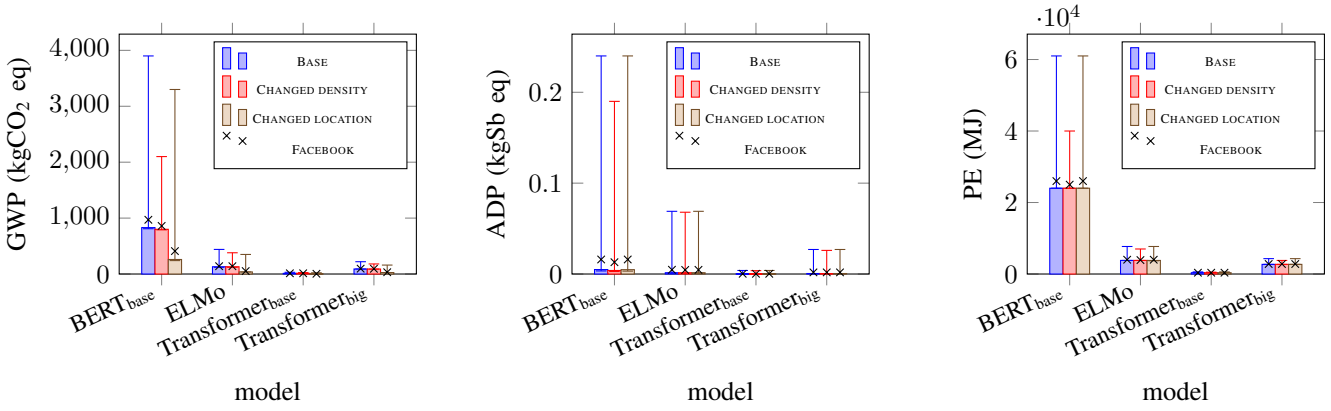


Fig. 4: Evaluation of the impacts of training NLP models on GWP, PE, and ADP. Scenario 'Base' uses MLCA default parameters with a US location. Scenario 'Changed density' uses memory density from [29] instead of MLCA's default value. Scenario 'Changed location' uses the default parameters with a location in France. The value interval represents the variation due to the possible range of embodied impacts, with the mark Facebook corresponding to a scenario in Facebook's data centers.

and cooling equipment production are also not included, but these are shared with many servers, leading to a minimal contribution per task.

c) Limits of Life Cycle Assessment: Conducting an attributional LCA does not allow to explore all of the possible impacts of the considered solution. Indeed, it cannot explore the social consequences and ethical challenges a solution poses. Such social consequences and ethical challenges are explored in [14]–[16], [47]. LCA also cannot explore the changes induced by introducing the new solution, such as the rebound effect that are frequent in ICT [48], [49].

It has been proposed that the impacts of data centers are already negligible since big companies buy and produce 'green' energy to power their data centers and offset their carbon emissions [44]. While LCA is great at highlighting the impacts of other phases of the hardware life cycle than the use phase, its results can greatly vary depending on the chosen impact factors. For instance, choosing the impact factor of the bought 'green' energy over the impact factor of the local electricity mix puts forth the assumption that renewable energy can be primarily used by digital companies over other activities. Furthermore, including carbon offsetting in the analysis dramatically lowers the assessed carbon footprint. However, the true potential of carbon offsetting[14] and the relevance of removing carbon offset from carbon emissions accountability have been criticized [49].

The results of LCA depend on a multitude of hypotheses and might greatly differ depending on the scope of the analysis, rendering these results difficult to exploit in comparison studies [50]. Furthermore, while feedback from impacts assessment studies may induce small changes to the system, there is a need to associate these assessments with proactive alternatives to the growth paradigm.

[14]https://www.theguardian.com/environment/2023/jan/18/revealed-forest-carbon-offsets-biggest-provider-worthless-verra-aoe

VI. Conclusion and Future Work

This paper addressed the evaluation of the environmental impacts of ML applications. We introduced a tool named MLCA to support researchers in estimating the impacts of computations. MLCA can contribute to cost/benefit analysis. This tool leveraged existing methodologies and tools to integrate LCA considerations for a more comprehensive estimate of carbon footprint, as well as other environmental impacts such as resource depletion.

A series of case studies assessed the quality of MLCA estimates, including independent reproduction of prior experiments. The tool and code for all the experiments presented in this paper are available at https://github.com/blubrom/MLCA.

Our experiments suggest that the bigger the trained models, the bigger the required quantity of hardware to train the model, leading to higher shares of embodied impacts. This observation, combined with the growing size of models [4] and shift towards less carbon-intensive energy sources for data centers, indicate embodied emissions constitute an increasingly significant portion of the environmental impacts of ML. Results also suggest that multi-criteria impact evaluation can highlight impact shifting as common strategies that reduce carbon emissions also increase metal depletion.

Future work should strive to include additional indicators, such as water consumption or human toxicity, as well as an assessment of hardware end-of-life impacts. This inclusion will require efforts towards collecting dedicated data, which is currently unavailable for specialized hardware such as graphic cards.

Finally, the availability of a tool for impact assessment for a specific application paves the way for an assessment at the scale of a field in order to align with global sustainability objectives, such as the planetary boundaries. Such an assessment must be combined with a broader reflection on the role of ICT in a sustainable society.

References

[1] E. Strubell, A. Ganesh, and A. McCallum, "Energy and Policy Considerations for Deep Learning in NLP," in *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, A. Korhonen, D. R. Traum, and L. Màrquez, Eds., Association for Computational Linguistics, 2019, pp. 3645–3650.

[2] A. S. Luccioni, S. Viguier, and A.-L. Ligozat, "Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model," *Journal of Machine Learning Research*, vol. 24, no. 253, pp. 1–15, 2023.

[3] N. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "The Computational Limits of Deep Learning," in *Ninth Computing within Limits 2023*, LIMITS, Jun. 2023.

[4] J. Sevilla, L. Heim, A. Ho, T. Besiroglu, M. Hobbhahn, and P. Villalobos, "Compute Trends Across Three Eras of Machine Learning," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1–8.

[5] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green AI," *Commun. ACM*, vol. 63, no. 12, pp. 54–63, Nov. 2020.

[6] A.-L. Ligozat, J. Lefevre, A. Bugeau, and J. Combaz, "Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessment of AI Solutions," *Sustainability*, vol. 14, no. 9, 2022.

[7] L. Bouza, A. Bugeau, and L. Lannelongue, "How to estimate carbon footprint when training deep learning models? A guide and review," *Environmental Research Communications*, vol. 5, no. 11, p. 115 014, Nov. 2023.

[8] M. Jay, V. Ostapenco, L. Lefevre, D. Trystram, A.-C. Orgerie, and B. Fichel, "An experimental comparison of software-based power meters: focus on CPU and GPU," in *2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, 2023, pp. 106–118.

[9] N. Bannour, S. Ghannay, A. Névéol, and A.-L. Ligozat, "Evaluating the carbon footprint of NLP methods: A survey and analysis of existing tools," in *Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing*, Virtual: Association for Computational Linguistics, Nov. 2021, pp. 11–21.

[10] C. Wu, R. Raghavendra, U. Gupta, *et al.*, "Sustainable AI: environmental implications, challenges and opportunities," in *Proceedings of Machine Learning and Systems (MLSys) 2022, Santa Clara, CA, USA, August 29 - September 1, 2022*, mlsys.org, 2022.

[11] L.-P. P.-V. Clément, Q. E. Jacquemotte, and L. M. Hilty, "Sources of variation in life cycle assessments of smartphones and tablet computers," *Environmental Impact Assessment Review*, vol. 84, p. 106 416, 2020.

[12] L. H. Kaack, P. L. Donti, E. Strubell, G. Kamiya, F. Creutzig, and D. Rolnick, "Aligning artificial intelligence with climate change mitigation," *Nature Climate Change*, vol. 12, pp. 518–527, 6 Jun. 2022.

[13] A. Rasoldier, J. Combaz, A. Girault, K. Marquet, and S. Quinton, "How realistic are claims about the benefits of using digital technologies for GHG emissions mitigation?" In *Eighth Workshop on Computing within Limits 2022*, LIMITS, Jun. 2022.

[14] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?" In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, ser. FAccT '21, Virtual Event, Canada: Association for Computing Machinery, 2021, pp. 610–623.

[15] P. Dauvergne, "The globalization of artificial intelligence: Consequences for the politics of environmentalism," *Globalizations*, vol. 18, no. 2, pp. 285–299, 2021.

[16] H. H. Jiang, L. Brown, J. Cheng, *et al.*, "AI art and its impact on artists," in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, ser. AIES '23, Montréal, QC, Canada: Association for Computing Machinery, 2023, pp. 363–374.

[17] R. Verdecchia, J. Sallou, and L. Cruz, "A systematic review of green AI," *WIREs Data Mining and Knowledge Discovery*, vol. 13, no. 4, e1507, 2023.

[18] M. Dinarelli, M. Naguib, and F. Portet, "Toward low-cost end-to-end spoken language understanding," in *23rd Annual Conference of the International Speech Communication Association (Interspeech), Incheon, Korea, 18-22 September 2022*, H. Ko and J. H. L. Hansen, Eds., ISCA, 2022, pp. 2728–2732.

[19] T. Parcollet and M. Ravanelli, "The Energy and Carbon Footprint of Training End-to-End Speech Recognizers," working paper or preprint, Apr. 2021.

[20] L. F. W. Anthony, B. Kanding, and R. Selvan, *Carbontracker: Tracking and predicting the carbon footprint of training deep learning models*, 2020.

[21] V. Schmidt, Goyal-Kamal, B. Courty, *et al.*, *Mlco2/codecarbon: V2.1.4*, version v2.1.4, Sep. 2022.

[22] P. Henderson, J. Hu, J. Romoff, E. Brunskill, D. Jurafsky, and J. Pineau, "Towards the systematic reporting of the energy and carbon footprints of machine learning," *J. Mach. Learn. Res.*, vol. 21, no. 1, Jan. 2020.

[23] H. David, E. Gorbatov, U. R. Hanebutte, R. Khanna, and C. Le, "RAPL: Memory power estimation and capping," in *Proceedings of the 16th ACM/IEEE International Symposium on Low Power Electronics and Design*, ser. ISLPED '10, Austin, Texas, USA: Association for Computing Machinery, 2010, pp. 189–194.

[24] NVIDIA Corporation, *Nvidia Management Library (NVML)*, accessed may 2023 at https://developer.nvidia.com/nvidia-management-library-nvml, Jan. 2021.

[25] L. Lannelongue, J. Grealey, and M. Inouye, "Green algorithms: Quantifying the carbon footprint of compu-

tation," *Advanced Science*, vol. 8, no. 12, p. 2 100 707, 2021.

[26] A. Lacoste, A. Luccioni, V. Schmidt, and T. Dandres, *Quantifying the carbon emissions of machine learning*, 2019.

[27] U. Gupta, Y. G. Kim, S. Lee, *et al.*, "Chasing carbon: The elusive environmental footprint of computing," *IEEE Micro*, vol. 42, no. 4, pp. 37–47, 2022.

[28] Boavizta, *Numérique et environnement : Comment évaluer l'empreinte de la fabrication d'un serveur, au-delà des émissions de gaz à effet de serre?* https://www.boavizta.org/blog/empreinte-de-la-fabrication-d-un-serveur, Accessed on May 15, 2023, 2021.

[29] J. Gröger, R. Liu, L. Stobbe, J. Druschke, and N. Richter, *Green Cloud Computing: Lebenszyklusbasierte Datenerhebung zu Umweltwirkungen des Cloud Computing: Abschlussbericht*. Umweltbundesamt, 2021.

[30] P. Loubet, A. Vincent, A. Collin, C. Dejous, A. Ghiotto, and C. Jego, "Life cycle assessment of ICT in higher education: A comparison between desktop and single-board computers," *The International Journal of Life Cycle Assessment*, pp. 1–19, 2023.

[31] M. Ficher, T. Bauer, and A.-L. Ligozat, "A comprehensive review of the end-of-life modeling in LCAs of digital equipment," working paper or preprint, Mar. 2024.

[32] P. Forster, T. Storelvmo, K. Armour, *et al.*, "The earth's energy budget, climate feedbacks and climate sensitivity," in *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte, P. Zhai, A. Pirani, *et al.*, Eds., Cambridge University Press, 2023, pp. 923–1054.

[33] L. van Oers, J. B. Guinée, and R. Heijungs, "Abiotic resource depletion potentials (ADPs) for elements revisited—updating ultimate reserve estimates and introducing time series for production data," *The International Journal of Life Cycle Assessment*, vol. 25, pp. 294–308, 2020.

[34] H. Bruijn, R. Duin, M. A. J. Huijbregts, *et al.*, *Handbook on Life Cycle Assessment - Operational Guide to the ISO Standards*. Springer Dordrecht, 2002.

[35] R. Frischknecht, F. Wyss, S. B. Knöpfel, T. Lützkendorf, and M. Balouktsi, "Cumulative energy demand in LCA: The energy harvested approach," *International Journal of Life Cycle Assessment 20*, pp. 957–969, 2015.

[36] V. Avelar, D. Azevedo, A. French, and E. N. Power, "PUE: A comprehensive examination of the metric," *White paper*, vol. 49, 2012.

[37] G. A. Brady, N. Kapur, J. L. Summers, and H. M. Thompson, "A case study and critical assessment in calculating power usage effectiveness for a data centre," *Energy Conversion and Management*, vol. 76, pp. 155–161, 2013.

[38] M. Z. Hauschild, "Better – but is it good enough? on the need to consider both eco-efficiency and eco-effectiveness to gauge industrial sustainability," *Procedia CIRP*, vol. 29, pp. 1–7, 2015, The 22nd CIRP Conference on Life Cycle Engineering.

[39] A. W. Hjalsted, A. Laurent, M. M. Andersen, K. H. Olsen, M. Ryberg, and M. Hauschild, "Sharing the safe operating space: Exploring ethical allocation principles to operationalize the planetary boundaries and assess absolute sustainability at individual and industrial sector levels," *Journal of Industrial Ecology*, vol. 25, no. 1, pp. 6–19, 2021.

[40] W. Steffen, K. Richardson, J. Rockström, *et al.*, "Planetary boundaries: Guiding human development on a changing planet," *Science*, vol. 347, no. 6223, p. 1 259 855, 2015.

[41] S. Sala, E. Crenna, M. Secchi, and E. Sanyé-Mengual, "Environmental sustainability of european production and consumption assessed against planetary boundaries," *Journal of Environmental Management*, vol. 269, p. 110 686, 2020.

[42] M. Talalkhokh and F. Laugier, "Mise à plat méthodologique de la révision de l'objectif d'émissions moyennes par personne à l'échelle mondiale en 2050," Tech. Rep., 2024, note réalisée par la 2tonnes Compagnie.

[43] Sphera, *Life Cycle Assessment Dell Servers R6515, R7515, R6525, R7525*, 2021.

[44] D. Patterson, J. Gonzalez, U. Hölzle, *et al.*, "The carbon footprint of machine learning training will plateau, then shrink," *Computer*, vol. 55, no. 7, pp. 18–28, 2022.

[45] T. Pirson, L. Golard, and D. Bol, "Evaluating the (ir)relevance of IoT solutions with respect to environmental limits based on LCA and backcasting studies," in *Ninth Computing within Limits 2023*, LIMITS, Jun. 2023.

[46] W. H. Organization, *Children and digital dumpsites: e-waste exposure and child health*. World Health Organization, 2021, xix, 86 p.

[47] O. Keyes, "The misgendering machines: Trans/HCI implications of automatic gender recognition," *Proc. ACM Hum. Comput. Interact.*, vol. 2, no. CSCW, 88:1–88:22, 2018.

[48] C. Gossart, "Rebound effects and ICT: A review of the literature," in *ICT Innovations for Sustainability*, L. M. Hilty and B. Aebischer, Eds., Cham: Springer International Publishing, 2015, pp. 435–448.

[49] D. Bol, T. Pirson, and R. Dekimpe, "Moore's law and ICT innovation in the anthropocene," in *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2021, pp. 19–24.

[50] K. Ellsworth-Krebs, M. Niero, and T. Jack, "Feminist LCAs: Finding leverage points for wellbeing within planetary boundaries," *Sustainable Production and Consumption*, vol. 39, pp. 546–555, 2023.