A Practical Guide to Quantifying Carbon Emissions for Machine Learning researchers and practitioners

Anne-Laure Ligozat, Sasha Luccioni

A Practical Guide to Quantifying Carbon Emissions
for Machine Learning researchers and practitioners

Anne-Laure Ligozat & Sasha Luccioni

October 13, 2021

The goal of this short guide is to help the Machine Learning (ML) community better understand their carbon impact and to take steps to mitigate it.

1 Carbon Tracking

At the center of the climate crisis is a commonplace but very important concept: that of **carbon dioxide** (CO_2), low amounts of which occur naturally in the Earth's atmosphere, but whose concentration has been rapidly increasing in recent decades due to human activity. This increase is dangerous because of CO_2's effect as a **greenhouse gas**, which means that it contributes to global warming by keeping heat trapped within the atmosphere, which consequences on global ecosystems. In order to minimize these impacts, it is important to: 1) quantify the carbon impact of our actions; and 2) reduce, or mitigate, that impact in order to help slow down global warming and climate change more broadly.

What are the carbon impacts of Machine Learning?

The most directly visible impact of training and deploying a Machine Learning model is the emission of CO_2 and other greenhouse gases due to the increase in power consumption (i.e. dynamic consumption) incurred by the equipment at running time. While the impact of dynamic consumption is significant, we should not fail to see the forest for the trees, and consider the entirety of the ML pipeline. Notably, other dimensions of model impact that should be considered include: model preparation overhead, static consumption of the equipment, infrastructure, as well as the overall Life Cycle Analysis of the equipment, which we will describe in the sections below [C10].

The overhead of preparing a Machine Learning model can be envisioned by considering the deployment of Machine Learning systems beyond a two-step process whereby models are first trained and then used for inference. So far, the focus on carbon emission assessment has been predominantly on the training phase, since the impact can be substantial if the model is trained on large datasets with many hyperparameters. However, the exact carbon footprint of model training depends on multiple factors, including the electric grid utilized, the type of energy mix, the energy consumption of hardware, and training time.

In comparison, during the **inference stage**, each forward pass through the model incurs a lower impact relative to the entirety of the training phase. However, given that an increasing number of dynamic Machine Learning models are 'always on' and ready for live deployment (e.g. online search and retrieval systems, voice assistants, ML-powered navigation, etc.), multiple uses will accrue carbon impact across the weeks and months of deployment.

The static consumption of equipment should also be partly attributed to the ML computing process, since equipment is switched on in part to address the computing needs of the ML model. This static consumption includes the electrical costs due to the server power supply (which depends on its efficiency), the motherboard, the network card, and all the server components that have a static power consumption which does not depend on the load.

In addition to the hardware used to launch the ML program, the carbon footprint estimation should take into account the complete infrastructure that constitutes the **environment of the machine** including routers, storage servers, Air Conditioning, etc. This can be done using a metric such as *Power Usage Effectiveness (PUE)*, which estimates the energy consumption of the whole infrastructure. The average reported data center PUE is 1.58, meaning that around 37% of the energy consumed is used for things like data center cooling, lighting and distribution [B6]. Also, it can be noted that while some facilities reuse the waste heat produced by the servers and other computing equipment, this does not mean their computing is positive for the environment. Some data center metrics such as Energy Reuse Effectiveness can nonetheless take this into account.

Life Cycle Assessment is a methodology that uses allocation methods to define adequate environmental impact estimates [C7]. In fact, the carbon footprint should account for the entire life cycle of equipment, including its production and end of life: the fabrication and transportation of servers and chips for instance, their maintenance and, eventually, their disposal, since all of these steps come with costs to the environment. This is because they require materials such as rare metals and water, and emit greenhouse gases both during their physical creation and their transportation to the customer. Life Cycle Assessment usually allocates part of the CO_2 and other greenhouse gases emitted during the production of equipment to its usage, e.g. when they are used for running computer programs, since the equipment was partly produced for this purpose.

Additional activities can also be considered as contributing to the model overhead. They include upstream tasks such as data collection and processing, concurring tasks such as the engineering efforts involved in deploying models, and downstream tasks such as the presentation of the work internationally. It can be noted that ML also has indirect impacts that come with its use in products or services – in that it induces changes in other processes or even more generally in everyday behaviors, economic structures and lifestyles – e.g. increased purchases due to recommendation systems, increased mobility with autonomous vehicles, over-activity of financial markets with high-frequency trading [C8].

However, we recognize that whether all of these are to be considered as part of the carbon footprint of ML models is, however, up for debate!

2 Offsetting

The concept of **carbon offsetting** often comes up when discussing the carbon impact of goods and services. However, its purpose and function is not always clear, so we will elaborate on it below.

What is carbon offsetting?

Carbon offsetting consists of *compensating* for the greenhouse gas emissions that cannot be avoided by financing projects that store or reduce an *equivalent* amount of emissions.

Is offsetting a way to reduce my carbon footprint?

No. To limit global warming, everyone needs to drastically reduce their emissions, not via offsetting, which consists in repairing damages once they have already been made, but via a direct decrease of emissions, i.e. by emitting less (see [D12] reference for further information).

When can offsetting be useful?

Offsetting should be a last resort to counteract those emissions that you cannot avoid, for instance when you need to travel to present your work at a conference, or your commute to your office and back.

3 What are the most impactful steps I can take?

As a practitioner (ordered by impact)

- **Reduce your I/O and redundant computation/data copying/storage**: start with smaller datasets to debug your model, and use shared data storage with members of your team so you don't need to have individual copies.
- **Choose a low-carbon data center**: When running models on the cloud, consult a tool like Electricity Map to choose the least carbon-intensive data center.
- **Avoid wasted resources**: by steering clear of grid search and by reusing or fine-tuning previously trained models when possible. Also, strive towards designing your training and experimentation to minimize discarded computing time and resources in case of failure.
- **Quantify and disclose your emissions**: use packages like CodeCarbon, Carbon tracker and Experiment impact tracker, which can be included in your code at runtime or online tools like Green algorithms and ML CO_2 Impact that can allow you to estimate your emissions afterwards. In both cases, share these figures with your community to help establish benchmarks and track progress!

As an institution

- **Deploy your computation in low-carbon regions** when possible.
- **Provide institutional tools** for tracking emissions and enable them by default on your computing infrastructure
- **Cap computational usage**: for instance at maximum 72 hours per process, in order to reduce wasted resources.
- **Carry out awareness campaigns** regarding the environmental impact of ML.
- **Facilitate institutional offsets** for those emissions that cannot be avoided, such as commuting and building construction.

Contact Information

Sasha Luccioni (`sasha.luccioni@mila.quebec`) and
Anne-Laure Ligozat (`anne-laure.ligozat@lisn.upsaclay.fr`).

Acknowledgements

Thanks to all contributors of the document: Emmanuelle Frenoux, Aurélie Névéol, Anne-Cécile Orgerie, Jake Tae, Canwen Xu

References

Machine Learning Impacts

[A1] Peter Henderson, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, and Joelle Pineau. Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248):1–43, 2020.

[A2] Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*, 2019.

[A3] Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*, 2019.

Data Center Energy Use

[B4] Udit Gupta, Young Geun Kim, Sylvia Lee, Jordan Tse, Hsien-Hsin S Lee, Gu-Yeon Wei, David Brooks, and Carole-Jean Wu. Chasing carbon: The elusive environmental footprint of computing. In *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 854–867. IEEE, 2021.

[B5] George Kamiya. Data centres and data transmission networks. *Int. Energy Agency*, 2020.

[B6] Carolina Koronen, Max Åhman, and Lars J Nilsson. Data centres in future european energy systems—energy efficiency, integration and policy. *Energy Efficiency*, 13(1):129–144, 2020.

Life Cycle Assessment and ICT impacts

[C7] Michael Z Hauschild, Ralph K Rosenbaum, and Stig Irvin Olsen. *Life cycle assessment*, volume 2018. Springer, 2018.

[C8] Lorenz M Hilty and Magda David Hercheui. ICT and sustainable development. In *What kind of information society? Governance, virtuality, surveillance, sustainability, resilience*, pages 227–235. Springer, 2010.

[C9] GHG Protocol. ICT Sector Guidance built on the GHG Protocol Product Life Cycle Accounting and Reporting Standard. *Global: GHG Protocol*, 2017.

[C10] International Communications Union. Methodology for environmental life cycle assessments of information and communication technology goods, networks and services. https://www.itu.int/rec/T-REC-L.1410, 2015.

Carbon Offsetting

[D11] Derik Broekhoff, Michael Gillenwater, Tani Colbert-Sangree, and Patrick Cage. Securing climate benefit: A guide to using carbon offsets. *Stockholm Environment Institute & Greenhouse Gas Management Institute*, 60, 2019.

[D12] carbone 4. Stop saying "carbon offset": Moving from "offsetting" to "contributing". http://www.carbone4.com/stop-saying-carbon-offset-from-offsetting-to-contributing/?lang=en, 2019.

[D13] United Nations. United nations carbon offset platform. https://offset.climateneutralnow.org/.