Integrating carbon footprint information in research papers: presentation and demo of existing tools

Clément Morand, Aurélie Névéol, Anne-Laure Ligozat

November 19, 2024

Université Paris-Saclay, Laboratoire Interdisciplinaire des Sciences du Numérique, CNRS

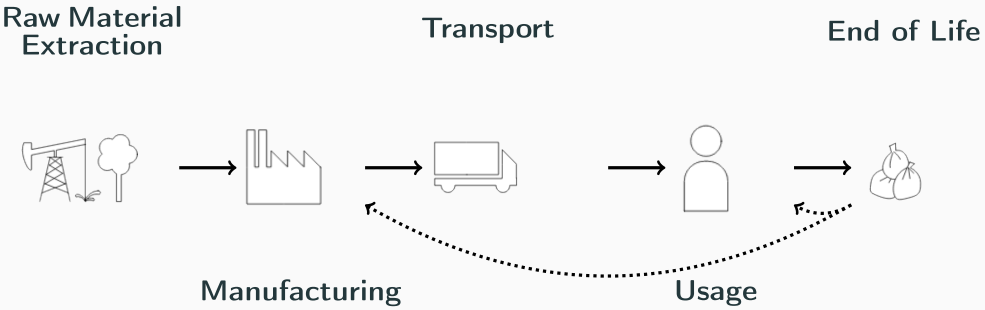
Why measure the impact of experiments?

- Need for sustainable research
- Need for a comprehensive approach to evaluation, beyond leaderboards

Sources :

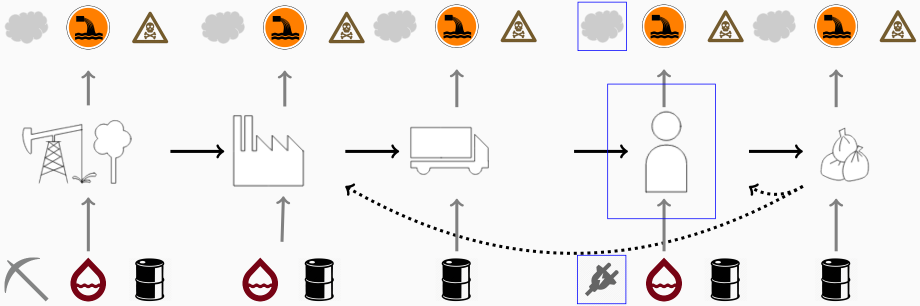
Strubell E, Ganesh A and McCallum A. Energy and Policy Considerations for Deep Learning in NLP. Proc Annual Meeting of the Association for Computational Linguistics (ACL):3645-3650 (2019).
Ethayarajh K and Jurafsky D Utility is in the Eye of the User: A Critique of NLP Leaderboards. Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP) 4846-53. (2020).

Life cycle phases of hardware



Each phase has different impacts

Pollution (Emissions in air, water, soil)



Natural resources

How to evaluate the carbon footprint of ML tasks

Using a Wattmeter

- Feasible in theory
- Impractical
- Difficulties with allocation to single process

tools for assessing energy consumption

- Need to manually convert to obtain environmental impact

Tools for assessing environmental impact

Online tools

1. Green Algorithms
2. ML CO2 Impact
3. MLCA

Python toolkits

3. Experiment Impact Tracker
4. Carbon Tracker
5. Cumulator

2'. Code Carbon

Features of measurement tools

Feature	**online** (Green Algorithms)	**toolkit** (Code Carbon)
direct measure	X	✓
estimation	✓	X
asynchronous	✓	X
comparison on same hardware	$\sim$	✓
easy to install	✓	$\sim$

How is energy consumption assessed by these tools?

$$E_{dynamic} = \texttt{running time} \times (P_{CPU} + P_{GPU} + P_{mem})$$

measures of power draw

- CPU (P_{CPU}) and memory (P_{mem}): RAPL
- GPU (P_{GPU}): NVML / nvidia-smi

or

estimation of power draw

- CPU and GPU: *Thermal Design Point* (TDP) $\times$ usage factor
- Memory: $\texttt{memory}_{size} \times P_{per\ GB}$

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

How is energy consumption assessed by these tools?

$$E_{dynamic} = \texttt{running time} \times (P_{CPU} + P_{GPU} + P_{mem})$$

measures of power draw

- CPU (P_{CPU}) and memory (P_{mem}): RAPL
- GPU (P_{GPU}): NVML / nvidia-smi

or

estimation of power draw

- CPU and GPU: *Thermal Design Point* (TDP) $\times$ usage factor
- Memory: $\texttt{memory}_{size} \times P_{per\ GB}$

$$E = E_{dynamic} \times \texttt{Efficiency of Datacenter}$$

Typically the *Power Usage Efficiency* (PUE), but incomplete.

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

How is energy consumption assessed by these tools?

$$E_{dynamic} = \texttt{running time} \times (P_{CPU} + P_{GPU} + P_{mem})$$

measures of power draw

- CPU (P_{CPU}) and memory (P_{mem}): RAPL
- GPU (P_{GPU}): NVML / nvidia-smi

or

estimation of power draw

- CPU and GPU: *Thermal Design Point* (TDP) $\times$ usage factor
- Memory: $\texttt{memory}_{size} \times P_{per\ GB}$

$$E = E_{dynamic} \times \texttt{Efficiency of Datacenter}$$

Typically the *Power Usage Efficiency* (PUE), but incomplete.

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

Tests on Jean Zay [Luccioni et al., 2023]

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

Green Algorithm online tool

http://calculator.green-algorithms.org/

Code Carbon python package

https://github.com/mlco2/codecarbon

```
import tensorflow as tf

from codecarbon import EmissionsTracker

mnist = tf.keras.datasets.mnist

(x_train, y_train), (x_test, y_test) = mnist.load_data()
x_train, x_test = x_train / 255.0, x_test / 255.0


model = tf.keras.models.Sequential([tf.keras.layers.Flatten(input_shape=(28, 28)),
        tf.keras.layers.Dense(128, activation="relu"), tf.keras.layers.Dropout(0.2),
        tf.keras.layers.Dense(10),])

loss_fn = tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True)
model.compile(optimizer="adam", loss=loss_fn, metrics=["accuracy"])

tracker = EmissionsTracker()
tracker.start()
model.fit(x_train, y_train, epochs=10)
emissions: float = tracker.stop()
print(emissions)
```

Code Carbon python package

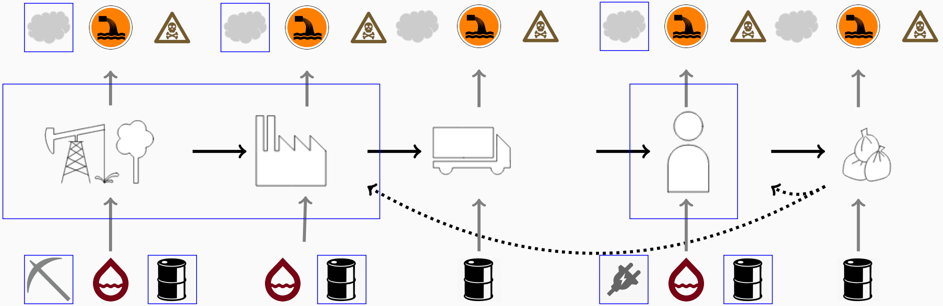
https://github.com/mlco2/codecarbon

```
import tensorflow as tf

from codecarbon import track_emissions
@track_emissions(project_name="mnist")
def train_model():
    mnist = tf.keras.datasets.mnist
    (x_train, y_train), (x_test, y_test) = mnist.load_data()
    x_train, x_test = x_train / 255.0, x_test / 255.0
    model = tf.keras.models.Sequential([tf.keras.layers.Flatten(input_shape=(28, 28)),
    tf.keras.layers.Dense(128, activation="relu"), tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(10)])
    loss_fn = tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True)
    model.compile(optimizer="adam", loss=loss_fn, metrics=["accuracy"])
    model.fit(x_train, y_train, epochs=10)
    return model


if __name__ == "__main__":
    model = train_model()
```

Code Carbon python package

```
[codecarbon INFO @ 11:15:30] Energy consumed for RAM : 0.000018 kWh.
                             RAM Power : 5.737926006317139 W
[codecarbon INFO @ 11:15:30] Energy consumed for all CPUs : 0.000044 kWh.
                             Total CPU Power : 14.0 W
[codecarbon INFO @ 11:15:30] 0.000061 kWh of electricity used since the beginning.
3.56889761642147e-06
```

Pollution (Emissions in air, water, soil)



Natural resources

Impact categories assessed by MLCA

- *Abiotic Depletion Potential* (ADP), measured in kgSbeq [van Oers et al., 2020, Bruijn et al., 2002]
- *Primary Energy demand* (PE), measured in MJ [Frischknecht et al., 2015]
- *Global Warming Potential* (GWP), measured in gCO_2eq [Forster et al., 2023]

Clément Morand, Aurélie Névéol, Anne-Laure Ligozat. MLCA: a tool for Machine Learning Life Cycle Assessment. 2024 International Conference on ICT for Sustainability (ICT4S), Jun 2024, Stockholm, Sweden. ⟨hal-04643414⟩ https://github.com/blubrom/MLCA

Modeling the production impacts of a server

Bottom Up approach (Machine = graphics card + CPU + memory + motherboard + storage + Power supply + casing)

Graphics card = GPU + Memory + Base

- GPU modeled by die size
- memory modeled by memory size
- base impacts computed from [Loubet et al., 2023]

$$\begin{aligned} Graphics\ card_{impact} = & die_{size} * die_{impact_{per-cm^2}} & + \\ & memory_{size} * memory_{impact_{perGB}} & + \\ & base_{impact} & \end{aligned}$$

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

Example (NVIDIA A100 SMX4 80GB)

GWP: 330 kgCO_2 eq PE: 3900 MJ ADP: 0.027 kgSb eq

Attributing production impacts to a specific task

linear attribution

$$\texttt{total available hours} = 365 * 24 * \texttt{replacement rate} * \texttt{average usage}$$

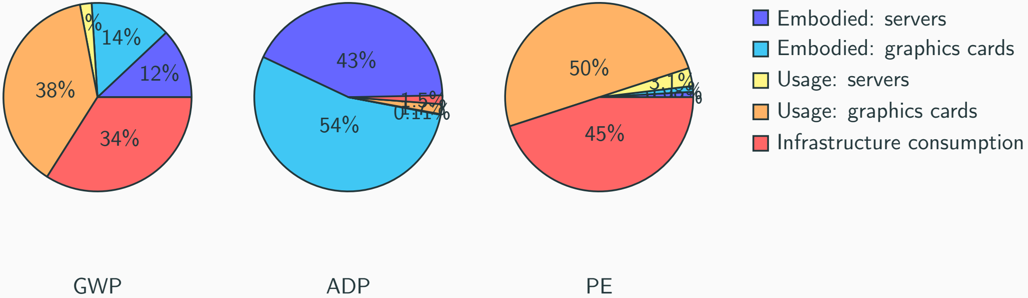
$$\texttt{embodied}_{impact} = \texttt{manufacturing}_{impact} \frac{\texttt{hours usage}}{\texttt{total available hours}}$$

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

Example (Jean Zay cluster [Luccioni et al., 2023])

- replacement rate of 6 years
- 85% average usage over the lifetime

Distribution of the impact of the Bloom model



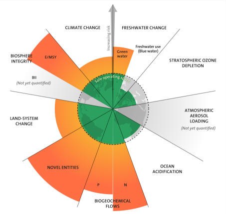
Putting impacts in perspective

Need for a global perspective [Rasoldier et al., 2022], [Hauschild, 2015]

COP21 · CMP11
PARIS 2015
UN CLIMATE CHANGE CONFERENCE

IPCC

- 2 tCO_2 e/person/year [Talalkhokh and Laugier, 2024]



[a]

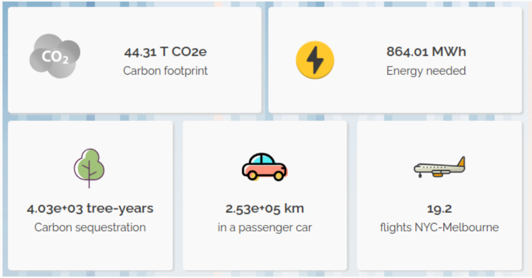
Planetary boundaries [Sala et al., 2020]

- $PB_{GWP} = 985$ $kgCO_2$ eq/person/year
- $PB_{ADP} = 3.17E\text{-}02$ kgSbeq/person/year

[a] Credit: "Azote for Stockholm Resilience Centre, based on analysis in Persson et al 2022 and Steffen et al 2015".

Perspectives on the impact of the BLOOM model

Green-Algorithms assessment of BLOOM



http://calculator.green-algorithms.org/

MLCA assessment of BLOOM

- GWP: 59tCO_2 eq
  - annual emissions of 59 person (PB_{GWP})
  - annual emissions of 29 person (SNBC)
- ADP: 1.2 kgSb eq
  - annual resource extraction of 38 person (PB_{ADP})
- PE: 9,800,000 MJ

Adding this info in your research paper

https://doi.org/10.1016/j.jbi.2022.104073

Table 2
Overall results on test corpus.

	Precision	Recall	F-Measure	CO_2 equivalent (g.)
Private Model (*MERLOT, teacher model*)	0.852	0.862	0.857	123
Public Model (*DEFT*)	0.592	0.383	0.465	22
Dictionary-based Model (*JDM*)	0.153	0.062	0.089	–
Dictionary-based Model (*UMLS*)	0.246	0.168	0.200	–
Privacy-Preserving Mimic Model (*DEFT, student model*)	0.604	0.743	0.666	30
Privacy-Preserving Mimic Model (*CAS, student model*)	**0.628**	**0.806**	**0.706**	169
Privacy-Preserving Mimic Model (*CépiDc, student model*)	0.580	0.710	0.638	394

Adding this info in your research paper

This algorithm runs in 12h on 2 GPUs NVIDIA GTX 1080 Ti and 12 CPUs Xeon E5-2683 v4, and draws 35.74 kWh. Based in France, and ran 3 times in total, this has a carbon footprint of 1.83 kg CO2e, which is equivalent to 2.00 tree-months (calculated using green-algorithms.org v2.2 [1]).

[1] Lannelongue, L., Grealey, J., Inouye, M., Green Algorithms: Quantifying the Carbon Footprint of Computation. Adv. Sci. 2021, 2100707.

Pitfalls to avoid when using the tools

- Wrong impact factor used (In France, value is between 20 and 100 gCO_2 eq/kWh)
- Wrong order of magnitude
  - Wrong hardware specification (e.g., incorrect name or TDP for a processing unit)
  - Wrong use rate (using one CPU core should consume less than a GPU)
  - Consumption from another process atributed to the process measured

Summary:

- 3 easy-to-use tools to measure impact
  - Green Algorithms and CodeCarbon only assess energy and carbon footprint
  - MLCA also accounts for hardware production and other impact categories
- avoid pitfalls
- Assessments are partial

Bruijn, H., Duin, R., Huijbregts, M. A. J., Guinee, J. B., Gorree, M., Heijungs, R., Huppes, G., Kleijn, R., Koning, A., van Oers, L., Sleeswijk, A. W., Suh, S., and de Haes, H. A. U. (2002).
Handbook on Life Cycle Assessment - Operational Guide to the ISO Standards.
Springer Dordrecht.

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., Lunt, D., Mauritsen, T., Palmer, M., Watanabe, M., Wild, M., and Zhang, H. (2023).
The earth's energy budget, climate feedbacks and climate sensitivity.
In Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., editors, Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment

Report of the Intergovernmental Panel on Climate Change, page 923–1054. Cambridge University Press.

Frischknecht, R., Wyss, F., Knöpfel, S. B., Lützkendorf, T., and Balouktsi, M. (2015).
Cumulative energy demand in lca: the energy harvested approach.
International Journal of Life Cycle Assessment 20, page 957–969.

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

Loubet, P., Vincent, A., Collin, A., Dejous, C., Ghiotto, A., and Jego, C. (2023).
Life cycle assessment of ict in higher education: a comparison between desktop and single-board computers.
The International Journal of Life Cycle Assessment, pages 1–19.

Luccioni, A. S., Viguier, S., and Ligozat, A.-L. (2023).
Estimating the carbon footprint of BLOOM, a 176b parameter language model.
Journal of Machine Learning Research, 24(253):1–15.

Rasoldier, A., Combaz, J., Girault, A., Marquet, K., and Quinton, S. (2022).
How realistic are claims about the benefits of using digital technologies for GHG emissions mitigation?
In Eighth Workshop on Computing within Limits 2022. LIMITS.
https://limits.pubpub.org/pub/real.

Sala, S., Crenna, E., Secchi, M., and Sanyé-Mengual, E. (2020).
Environmental sustainability of european production and consumption assessed against planetary boundaries.
Journal of Environmental Management, 269:110686.

Talalkhokh, M. and Laugier, F. (2024).
Mise à plat méthodologique de la révision de l'objectif d'émissions moyennes par personne à l'échelle mondiale en 2050.
Technical report.
note réalisée par la 2tonnes Compagnie.

van Oers, L., Guinée, J. B., and Heijungs, R. (2020).
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, 25:294–308.