The Social Costs of AI

Literature Review

Tim Mensinger*

Bonn Graduate School of Economics

Version: January 8, 2022

Preamble

This document is intended to serve as a reference guide to the existing literature and available online tools. Connections between your research question and the existing literature are highlighted and formalized where necessary. The abstract below is intended to clarify that we are analyzing the same research question.

All materials needed to reproduce this project can be found online in a *private* repository.[1]

Abstract

Modern AI research and its application in industry have developed rapidly over the last decades. We argue that the AI community has (implicitly) defined its success in terms of a metric which focuses exclusively on improving model accuracy. In this paper, we further argue that by neglecting other social criteria (e.g. environmental or economic), the community can drift into a socially suboptimal equilibrium. However, in order to accurately model the trade-offs between social criteria and model improvement, one must develop methods to measure the (environmental) impact of running machine learning models.

*tmensinger[at]uni-bonn.de

[1] Online repository: **https://github.com/timmens/social-cost-ai**. To access the repository one needs a GitHub account and an invitation.

1 Literature Review

There is a small, recent literature that addresses questions about the social costs of AI. In the following sections, I will review the works that I think are most similar to what you talked about. The order is roughly chronological. If you want to read only one paper, read Henderson *et al.* (2020, [1]). All of the following papers also discuss mitigation strategies, which I ignore in my summaries.

1.1 Quantifying the Carbon Emissions of Machine Learning

Lacoste *et al.* (2019, [2])

Summary: Explains that the AI community —research and industry— is consuming more and more computational resources, which translates directly into increasing energy demand. The paper then argues that an informed discussion of the tradeoffs that arise due to the (potentially significant) climate impact of machine learning requires a methodology that approximates the impact of a model (training, deployment). They then present their own implementation of such an approximation model. The model, or "ML CO_2 impact calculator" as they call it, is hosted online.[2] The calculator gives a rough estimate of the carbon emissions caused by energy use. It does this using information about the type of hardware, the hours the model was running, and the CO2 efficiency of the local power grid. Certain cloud computing infrastructures are also allowed (Google, AWS, Azure). For this estimate to be accurate, many assumptions must be made. Some of these are discussed in Henderson *et al.* (2020, [1]). In my view, this calculator can be used to obtain an initial estimate. Because of the underlying assumptions, the estimate will tend to be positively biased. Thus, if one uses conservative values for energy efficiency, the estimate should represent an upper bound of the true value.

Formula: The formula is not described in the paper, and I have not managed to find the time to extract the (mathematical) formula from their code.

Computational implementation: I have not looked at the computational implementation of the calculator in detail. Therefore, I cannot judge whether the implementation is trustworthy.

[2] `https://mlco2.github.io/impact/`

1.2 Energy and Policy Considerations for Modern Deep Learning Research

Strubell *et al.* (2020, [3])

Summary: The paper presents a strategy for (pointwise) estimating the CO_2 equivalents and economic costs of training AI models. They then apply this strategy to a set of NLP models and report the costs of training and the development costs. The development costs take into account that there are certain hyperparameters that need to be optimized outside of the training process. They find that a single training process of a standard NLP model ($BERT_{base}$) generates about as much CO_2 equivalents as a flight from NY to SF. In their case study, they consider a simpler model, for which they find that full development (repeated training of 5000 models instead of 1 model) increases the cost by a factor of 2000. The factor is not one-to-one with the number of repetitions, since there was a possibility that jobs were canceled early. They conclude that to compare training costs for different models, independent of hardware and local power grid specifications, one should use FPO, as mentioned in Schwartz *et al.* (2020, [4]).

Formula: Let p_c, p_g, p_r denote the average power consumption in Watts of the CPU, GPU and DRAM (memory), respectively. Let g denote the number of GPUs used for training. Let 1.58 denote the PUE (Power Usage Effectiveness) coefficient. The total power consumption at a given instance is then given by

$$p_t = \frac{1.58(p_c + p_r + g p_g)}{1000}, \tag{1}$$

where a dividing by 1000 converts Watts to KiloWatts. To calculate the CO_2 equivalents, one can use

$$CO_2 e = 0.954 p_t\,. \tag{2}$$

1.3 Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine learning

Henderson *et al.* (2020, [1])

Summary: The paper makes similar arguments as above as to why methods to accurately measure the carbon impact of machine learning models are needed; see Strubell *et al.* (2020, [3]), Schwartz *et al.* (2020, [4]), and Lacoste *et al.* (2019, [2]). The main contributions are a (legal) policy viewpoint on this topics and a very advanced software

implementation to measure the carbon impact of machine learning models in real time. Compared to the aforementioned work, their approach is to model energy consumption in as much detail as possible.

The micro-based idea of modeling each component of the system in minute detail is only beneficial if we can trust the micro-level data. Otherwise, aggregation can lead to an accumulation of errors. Particularly problematic is that we lose the information about the direction of the bias, while we knew above that we overestimated the true value.

Formula: The formula is similar to that used in Strubell *et al.* (2020, [3]); see Section 1.2. PUE stands for the power usage effectiveness coefficient. Consider the set of all (computer) processes spawned during a model run and denote it by $\mathcal{P}$. Compared to Section 1.2, here e denotes energy, while p denotes the percentage of each resource used by the attributable process. For example, for a process ρ, the CPU consumes energy e_{CPU}, and $p_{CPU} = p_{CPU}(\rho)$ is the utilization of the CPU by process ρ. The total energy consumption of process ρ by the CPU is then given by $p_{CPU}e_{CPU}$, where we omit the dependence on ρ for clarity. The total energy consumption is then defined by a double sum over the different components and processes, multiplied with the PUE coefficient:

$$e_{total} = PUE \sum_{\rho \in \mathcal{P}} (p_{dram}e_{dram} + p_{cpu}e_{cpu} + p_{gpu}e_{gpu}) \tag{3}$$

Computational Implementation: The computational implementation of their calculator seems to be the most advanced. However, it is still in development stage and buggy. Unfortunately, development appears to have ceased in July 2021. Unless development continues or is taken over by someone else, I cannot recommend using this software without making your own corrections.

1.4 Green AI

Schwartz *et al.* (2020, [4])

Summary: The paper argues that the field of AI is overly focused on *model accuracy*, leading to rapidly increasing computational costs, especially as the returns to accuracy are diminish relative to computational costs. This then leads to crowding out of researchers or companies with low resources as well as high economic and environmental costs ($\approx$ computational costs). They refer to this trend as *Red AI* and introduce their idea of *Green AI*, which considers other metrics that should take into account social costs (economic or environmental). The propose FPO (Floating Point Operations) as a specific metric to assess the quality of a new model, since it is hardware and locations

independent. They acknowledge the work of Lacoste *et al.* (2019, [2]) and Henderson *et al.* (2020, [1]), but believe that their direct measure of CO_2 equivalents is inferior because it loses the link to the model, as much of the variability in the measure is due to local energy efficiency and hardware choice.

Bibliography

[1] 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," *Journal of Machine Learning Research*, 2020. [Online]. Available: http://jmlr.org/papers/v21/20-312.html.

[2] A. Lacoste, A. Luccioni, V. Schmidt, and T. Dandres, "Quantifying the carbon emissions of machine learning," *ArXiv Preprint*, 2019. [Online]. Available: https://arxiv.org/abs/1910.09700.

[3] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for modern deep learning research," *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/7123.

[4] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green AI," *Communications of the ACM*, 2020. [Online]. Available: https://cacm.acm.org/magazines/2020/12/248800-green-ai/fulltext.

[5] "ML CO2 impact calculator," 2019. [Online]. Available: https://mlco2.github.io/impact/#home.

[6] "ML CO2 impact data and code source," 2019. [Online]. Available: https://github.com/mlco2.