Towards Quantifying the Carbon Emissions of Differentially Private Machine Learning

Anonymous Authors[1]

Abstract

In recent years, machine learning techniques utilizing large scale datasets have achieved remarkable performance. Differential privacy, by means of adding noise, provides strong privacy guarantees for such learning algorithms. The cost of differential privacy is often a reduced model accuracy and a lowered convergence speed. This paper investigates the impact of differential privacy on learning algorithms in terms of their carbon footprint due to either longer run-times or failed experiments. Through extensive experiments, further guidance is provided on choosing the noise levels which can strike a balance between desired privacy levels and reduced carbon emissions.

1. Introduction

With the rising availability of large-scale, diverse datasets, performance of Machine Learning (ML) models have experienced a significant boost across a multitude of domains. This boost is also associated with the availability of extreme-scale datasets, which is heavily linked to individual user contributions achieved via crowd-sourcing. ML algorithms often perform operations directly on raw user data leading to a host of privacy violations. Differential Privacy (DP) (Dwork & Roth, 2014; Abadi et al., 2016) makes progress in this domain by providing strong privacy guarantees for such contributing individuals. This guarantee is achieved by means of noise addition, which can be done at various stages of the ML pipeline including : (1) *Local DP:* Addition to the raw data (2) *Gradient DP:* Addition to gradients after clipping (Abadi et al., 2016) (3) *Addition to Output & Objective DP:* Addition to the final ML model or the loss function (Chaudhuri et al., 2011).

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

1.1. Impact on Climate Change

It is well-known that the computational resource investment requisite for training ML models generates a carbon footprint. This footprint is amplified in privacy-preserving setups where it is harder to reach consistent accuracy due to the addition of noise. Extended and failed runs (especially on larger datasets) actively contribute to an increase in the carbon footprint of ML experiments (Strubell et al., 2019). Therefore, an analysis of the climatic impact of this privacy modulation is critical. While the existing DP literature studies several performance aspects affected by varying privacy requirements, it lacks a comprehensive quantification of the carbon footprint of DP and how it is affected by variable privacy levels. Since DP also provides a mathematical paradigm to quantify the privacy budget of training ML models while tracking the privacy usage across multiple runs, this paper aims at quantifying the Carbon Emissions (CE) associated with varying privacy budgets of differentially private networks. In order to study impact of DP on these emissions, we implement *Gradient DP* (DP-SGD (Abadi et al., 2016)) for natural language processing, image classification, and reinforcement learning domains to identify the privacy implications, model performance and most crucially the carbon footprint of each algorithm. As per our knowledge this is the first attempt to quantitatively benchmark the carbon footprint of differentially private ML models.

1.2. Differential Privacy

Definition 1: Given a randomized mechanism $\mathcal{A} : \mathcal{D} \rightarrow \mathcal{R}$ (with domain $\mathcal{D}$ and range $\mathcal{R}$) and any two neighboring datasets $d_1, d_2 \in \mathcal{D}$ (*i.e.* they differ by a single individual data element), $\mathcal{A}$ is said to be (ε, δ)-differentially private for any subset $S \subseteq \mathcal{R}$ [1].

$$\Pr\left[\mathcal{A}\left(d_1\right) \in S\right] \leq e^{\varepsilon} \cdot \Pr\left[\mathcal{A}\left(d_2\right) \in S\right] + \delta \qquad (1)$$

Here, $\epsilon \geq 0, \delta \geq 0$. A $\delta = 0$ case corresponds to pure differential privacy, while both $\epsilon = 0, \delta = 0$ leads to an infinitely high privacy domain. Finally, $\epsilon = \infty$ provides no privacy guarantees.

[1]In this work, we exclusively use Gaussian noise

The privacy of differentially private models can be quantified with parameters such as epsilon (ε) and delta (δ). Utilizing DP-SGD (Abadi et al., 2016), that is, adding noise to the gradients at each step during training using a clipping factor (S) and noise multiplier (z), the amount of noise added to the model can be linked to to the degree of privacy that the model can achieve. Theoretically, a lower value of ε indicates a higher degree of privacy and this increased privacy degree is understandably, achieved at the expense of model performance due to the addition of the noise. The practical implication of this, however, includes a direct impact on the computational resources required to achieve model performance. Reduced privacy requirements allow the addition of noise with limited power, and hence, models can achieve appropriate performance without any significant resource expense. On the other hand, high privacy requirements necessitate adding a significantly large magnitude of noise which may directly lead to an increase in the number of training passes that the model has to iterate over to achieve the same accuracy. Further, noise addition may even lead to the non-convergence of some systems in the worst case.

1.3. Related Work

Works such as (Strubell et al., 2019; Toews, 2020) discuss how conventional Machine Learning models impact carbon footprint. In particular, (Strubell et al., 2019) discusses how training a single Deep Learning model generates the total lifetime carbon footprint of nearly five cars (as mentioned in (Toews, 2020)) which is more than 17 times the amount of CO_2 emissions generated by an average American per year. Regarding DP, there has been very little considerations on how Privacy-Preserving Machine Learning (PPML) impacts climate change. In (Qiu et al., 2021), a comprehensive study is presented on how local client-side models in Federated learning (FL) could potentially hold quality data required to understand climate change given data privacy concerns due to recent policies like GDPR (Skendžić et al., 2018). However, running local models on multiple client devices and aggregating them globally at the server level requires additional infrastructure in place, thereby causing a detrimental effect on carbon emissions.

1.4. Contributions and Impacts

In this paper, we provide the first benchmark to quantitatively assess how DP-noise affect carbon emissions in three different tasks : (1) a Natural Language Processing (NLP) task using news classification (2) a Computer Vision (CV) task using the MNIST dataset and (3) a Reinforcement Learning (RL) task using the Cartpole control problem. Intuitively, when DP noise is added to ML pipelines, the carbon emissions should increase as the energy required for computations increase. In order to quantify how the addition of noise plays into climate change, we track carbon emissions in the models using the *codecarbon* tool (Schmidt et al., 2021), a joint effort from authors of (Lacoste et al., 2019) and (Lottick et al., 2019). We record the average accuracy of several runs of the considered ML task to assess the behavior of DP-noise.

Given the rise in Privacy-enhancing Technologies and privacy policies, the addition of noise to mask data patterns has become prevalent. We envisage this work to provide an insight on how much noise could result in varying amounts of CO_2 emissions. Hence, our work takes a peek at how the addition of noise could impact a number of industries from healthcare to finance and justice, where sensitive data is commonly in use.

2. Experimental Results

2.1. BERT

In these set of experiments, we evaluate the performance of two experiments on Bidirectional Encoder Representations from Transformers or BERT (Devlin et al., 2019). The model is fine-tuned for topic-classification of news articles. The primary objective of these experiments is to observe the carbon emissions and energy usage of vanilla BERT and DP-BERT (over different privacy levels).

A randomly sampled subset of the AG News Classification (Anand, 2020) is used for this task with a 80/20 train-test split. 15000 instances are used to fine-tune this model. We use BERT in conjunction with the AdamW optimizer and the *bert-base-cased* tokenizer (with a batch size (B) of 32). Finally, we conduct the following two experiments for this task.

2.1.1. Evaluation of DP-BERT's carbon emissions and energy consumed for varying privacy regimes

The aim of this experiment is to analyse any possible association between different levels of privacy and carbon emissions. We run these experiments for 10 epochs each and present our results in Table 1 (averaged over 3 runs). Curiously, the carbon emissions for the $\epsilon = 0.5$ case is comparable to the EU's 2021 passenger vehicle standard (Bandivadekar, 2013).

Epsilon (ε)	CE (g)	EC (Wh)	Accuracy (%)
0.5	26.7 ± 0.63	49.9 ± 1.2	48.5 ± 1.39
2	26.3 ± 0.49	49.3 ± 0.9	52.0 ± 0.73
5	26.1 ± 0.1	48.9 ± 0.9	52.3 ± 0.36
15	25.9 ± 0.09	48.5 ± 0.1	54.2 ± 1.40
∞ **(Non-Private)**	25.2 ± 0.00	47.1 ± 0.27	58.5 ± 5.29

Table 1. **DP-BERT:** Emission-Accuracy trends over change in ϵ for reaching 52% accuracy.

In congruence with existing literature, the accuracy of the differentially private BERT increases consistently with the increase in epsilon. Interestingly, with the increase in the epsilon value – both, CE and EC decrease, though not by a very significant margin. Given that the range of the chosen ε varies considerably, the consequent difference in the carbon emission is not proportionally varied. The practical implication of this invariance can be seen as incurring nearly the same carbon footprint for two versions of a model with different degrees of privacy.

Epsilon (ε)	Epochs	CE (g)	EC (Wh)
0.5	19	153.6	287.3
2	12	96.6	180.6
5	9	80.9	151.3
15	7	56.9	106.5
∞ **(Non-Private)**	6	8.5	16

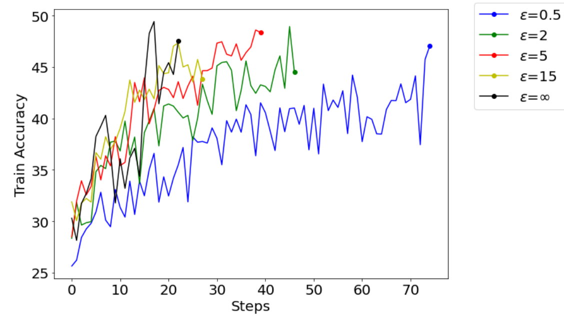
Table 2. Observing the number of epochs needed to achieve the threshold accuracy (T) with different privacy levels

2.1.2. Analyzing the resource expense of achieving a threshold accuracy at different privacy regimes

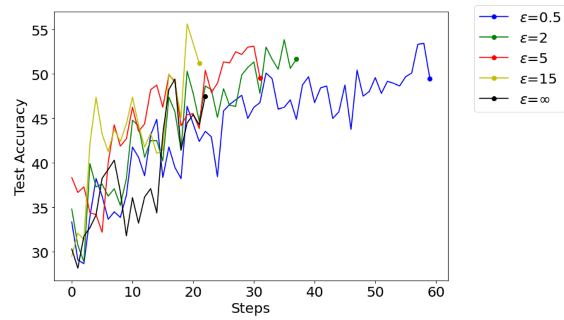
The main aim of this experiment is to evaluate how many resources, in terms of consequent carbon and energy emissions are expended in order to achieve a target or threshold accuracy with different degrees of privacy. As defined in the previous set of experiments, we compute the accuracies over $\varepsilon = 0.5, 2, 5, 15$.We set the target/threshold accuracy (T) to 52% as shown in Table 2.

It can be inferred from Table 2 that the Carbon Emission and Energy Usage required to attain the maximum experimental value of privacy is nearly 18 times the carbon emission required to attain the same threshold accuracy with a non-privacy preserving variant of the model. The practical consequence of this experiment dictates that enhancing the degree of privacy of the model, can incur a huge compute cost, which can invariably increase the carbon footprint of the model's training pipeline.

Additionally, From Figure 1, which present the accuracy curves for the experiment, it is quite evident that the vanilla variant (i.e a model without DP-noise) achieves the threshold accuracy with a significantly smaller carbon footprint than all the footprint of its privacy-preserving variants.



(a) **BERT:** Training Accuracy



(b) **BERT:** Testing Accuracy

Figure 1. **BERT with Gaussian DP:** Training and Testing accuracy trends over change in ϵ where the threshold accuracy (T) is set to 52%.

2.2. MNIST

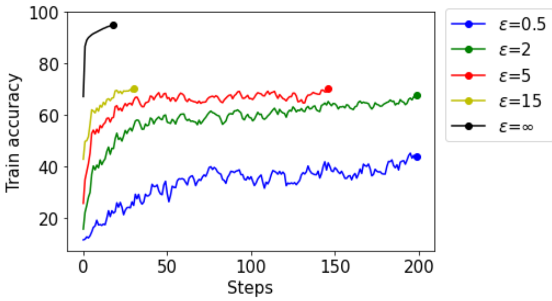
Epsilon (ε)	CE (g)	EC (Wh)
0.5 *	10.53 $\pm$ 2.21	40.41 $\pm$ 0.93
2 *	10.6 $\pm$ 2.43	40.5 $\pm$ 0.53
5	7.85 $\pm$ 1.84	29.93 $\pm$ 0.46
15	1.61 $\pm$ 0.37	6.17 $\pm$ 0.27
∞ **(Non-Private)**	0.08 $\pm$ 7e-04	0.38 $\pm$ 3.3e-03

Table 3. **MNIST:** Emission trends over change in ϵ for reaching 70% accuracy (* *70% accuracy not reached even after 200 epochs.)*

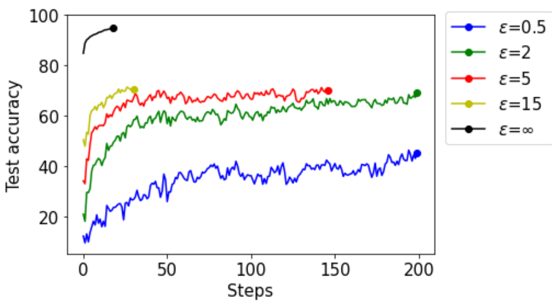
We evaluate our approach on the MNIST dataset (LeCun & Cortes, 2010) with a batch size of 128 using DP-SGD (Abadi et al., 2016). We use a simple multi-layer perceptron (MNIST 2NN) with a two hidden layers of 200 units each (parameters = 199,210) as the network from (McMahan et al., 2017). Our goal is to observe the trend in the CO_2 emissions by allowing the model to train and reach X accuracy with different values of ε (different levels of privacy). We compute the accuracies over $\varepsilon = 0.5, 2, 5, 15$ as shown in Fig. 2. We set the target/threshold accuracy (T) to 70% so that most of the privacy-variant models can achieve under 200 iterations. In Fig. 2 we see that only models with $\varepsilon = 5, 15$ reach 70% accuracy within 200 epochs. Fig. 2 also shows a clear trend on how increasing levels of privacy in ML models increases the amount of computation required to reach T, thereby releasing higher carbon emissions in comparison to the $\varepsilon = \infty$ (baseline) case.

2.3. Cartpole

For the reinforcement learning experiments, we trained a DQN over OpenAI Gym's Cartpole-v0 environment. The Cartpole environment (Barto et al., 1983) consists of an un-actuated joint to a cart. There are two possible actions which involve a force of +1 or -1 being applied to the cart along a friction-less track. The pole starts upright, with the



(a) **MNIST:** Accuracy on Training Set



(b) **MNIST:** Accuracy on Test Set

Figure 2. **MNIST with Gaussian DP:** Training and test accuracy trends during training for multiple ϵ values.

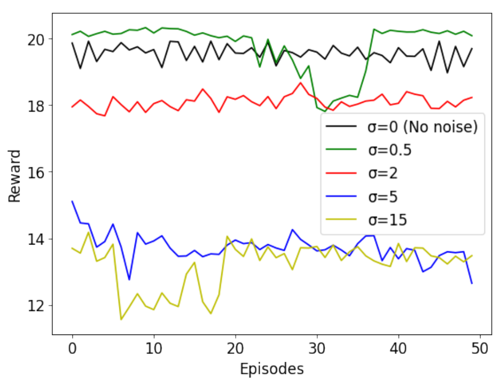


Figure 3. **CartPole with Gaussian DP:** Episodes vs Rewards for the mean reward every 100 episodes

goal of preventing it from falling over. For every time-step that the pole is upright, a reward of +1 is added to the total reward. However, if the pole exceeds 15 degrees from the vertical, or if the cart moves more than 2.4 units from the center, the episode ends.

The DQN's configuration (including the hyperparameters) is the same as the one used in (Wang & Hegde, 2019), and we observed results similar to this paper, with one variant of DP model slightly outperforming the baseline as shown in 3. It consists of a single hidden layer with 16 neurons. For our non-private experiment we obtained a mean reward of 19.94 and carbon emissions of 0.22 g on average (over a 1000 episodes). We provide results of the private variants in Fig. 3. Our setup included multiple experiments.

- Noise addition to DQN's output layer only. (1)
- Noise addition to both, the DQN's output layer and its parameters. The noise added to the parameters is the averaged noise sampled from the *noisebuffer* function during the forward pass. (2)

We varied the value of the variance σ of the distribution to observe its impact on the function approximated by the DQN. As expected, with increasing noise addition to the model (*i.e.*, increasing value of σ), we notice a drop in the average reward. Subsequently, the increased computations lead to higher carbon emissions. We observe that there is a significant increase in CE from Table 5 to Table 6 when the number of episodes increase.

Epsilon* ($\epsilon^* \propto 15\epsilon$)	**Sigma** ($\sigma \propto \frac{1}{\epsilon}$)	**Mean Reward**	**CE (g)**
1	15	4.5 ± 0.6	1.03 ± 0.06
3	5	2.2 ± 0.2	0.96 ± 0.03
7.5	2	19.9 ± 0.5	1.14 ± 0.06
30	0.5	19.4 ± 0.1	1.15 ± 0.06

Table 4. **CartPole:** Emission trends over change in ϵ^* post 1000 episodes in (1) following (Wang & Hegde, 2019)

Epsilon* ($\epsilon^* \propto 15\epsilon$)	**Sigma** ($\sigma \propto \frac{1}{\epsilon}$)	**Mean Reward**	**CE (g)**
1	15	2.3 ± 0.9	0.41 ± 0.01
3	5	10.2 ± 0.8	0.5 ± 0.02
7.5	2	7.6 ± 0.7	0.45 ± 0.02
30	0.5	13.8 ± 0.1	0.48 ± 0.03

Table 5. **CartPole:** Emission trends post 1000 episodes in (2)

Epsilon* ($\epsilon^* \propto 15\epsilon$)	**Sigma** ($\sigma \propto \frac{1}{\epsilon}$)	**Mean Reward**	**CE (g)**
1	15	13.2 ± 0.3	3.51 ± 0.26
3	5	13.7 ± 0.9	2.31 ± 0.28
7.5	2	18.1 ± 0.1	2.72 ± 0.23
30	0.5	19.8 ± 0.6	4.0 ± 0.31

Table 6. **CartPole:** Emission trends post 5000 episodes in (2)

3. Conclusion

We demonstrate and highlight the prominent impact of Privacy-Preserving Machine Learning (PPML) on carbon emissions over three ML domains, namely, CV, NLP and RL. We observe that the stronger privacy regime, *i.e,* a lower ϵ value, ML algorithms always result in higher levels of carbon emissions independent of the ML domain. We conclude that alongside the challenge of obtaining state-of-the-art performance, PPML needs to reduce the number of epochs required to reach the desired performance. This leads us to the following critical questions which we leave as open questions for the future: (1) Can we reduce the number of iterations (including hyperparameter tuning) required to reach a privacy-utility ratio? (2) How much does the size of ML models affect the carbon emissions and the overall performance under PPML?

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. Deep learning with differential privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, Oct 2016. doi: 10.1145/2976749.2978318. URL http://dx.doi.org/10.1145/2976749.2978318.

Anand, A. Ag news classification, 2020. URL https://www.kaggle.com/amananandrai/ag-news-classification-dataset.

Bandivadekar, A. One (vehicle efficiency) table to rule them all. https://theicct.org/blogs/staff/one-vehicle-efficiency-table-rule-them-all, 2013. Accessed: 2021-5-31.

Barto, A. G., Sutton, R. S., and Anderson, C. W. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(5):834–846, 1983. doi: 10.1109/TSMC.1983.6313077.

Chaudhuri, K., Monteleoni, C., and Sarwate, A. D. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12(3), 2011.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL https://www.aclweb.org/anthology/N19-1423.

Dwork, C. and Roth, A. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3–4):211–407, August 2014. ISSN 1551-305X. doi: 10.1561/0400000042. URL https://doi.org/10.1561/0400000042.

Lacoste, A., Luccioni, A., Schmidt, V., and Dandres, T. Quantifying the carbon emissions of machine learning. *Workshop on Tackling Climate Change with Machine Learning at NeurIPS 2019*, 2019.

LeCun, Y. and Cortes, C. MNIST handwritten digit database. 2010. URL http://yann.lecun.com/exdb/mnist/.

Lottick, K., Susai, S., Friedler, S. A., and Wilson, J. P. Energy usage reports: Environmental awareness as part of algorithmic accountability. *Workshop on Tackling Climate Change with Machine Learning at NeurIPS 2019*, 2019.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data, 2017.

Qiu, X., Parcollet, T., Fernández-Marqués, J., de Gusmão, P. P. B., Beutel, D. J., Topal, T., Mathur, A., and Lane, N. D. A first look into the carbon footprint of federated learning. *CoRR*, abs/2102.07627, 2021. URL https://arxiv.org/abs/2102.07627.

Schmidt, V., Goyal, K., Joshi, A., Feld, B., Conell, L., Laskaris, N., Blank, D., Wilson, J., Friedler, S., and Luccioni, S. CodeCarbon: Estimate and Track Carbon Emissions from Machine Learning Computing. 2021. doi: 10.5281/zenodo.4658424.

Skendžić, A., Kovačić, B., and Tijan, E. General data protection regulation — protection of personal data in an organisation. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1370–1375, 2018. doi: 10.23919/MIPRO.2018.8400247.

Strubell, E., Ganesh, A., and McCallum, A. Energy and policy considerations for deep learning in NLP. *CoRR*, abs/1906.02243, 2019. URL http://arxiv.org/abs/1906.02243.

Toews, R. Deep learning's carbon emissions problem. *Forbes*, 06 2020. URL https://www.forbes.com/sites/robtoews/2020/06/17/deep-learnings-climate-change-problem/?sh=6bcb1b9f6b43.

Wang, B. and Hegde, N. Privacy-preserving q-learning with functional noise in continuous state spaces, 2019.