PolluNet

Damon Aliomrany

Computer Science

Portland State University

Portland, Oregon, USA

doman9[@pdx.edu](mailto:lvn2@pdx.edu)

# 

# **ABSTRACT**

The main idea in this paper is to estimate the carbon footprint of ML. Later the data will be used in compression to determine whether ML is destructive to nature.

# **KEYWORDS**

Machine learning, ML, Artificial intelligence, AI, Model training, Carbon footprint, Carbon emission, power consumption, pollution.

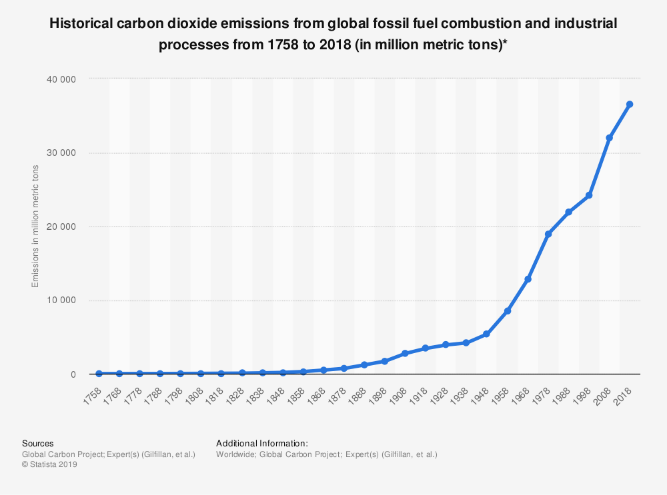
# **ACM Reference format:**

Damon Aliomrany. 2020. PolluNet. In *Proceedings of CS 410/510: Explorations of Data Science. Portland, OR, USA, 4 pages.*

# **1 Introduction**

Mankind has always been in a search of utopia by any means. Often, the sheen of this idea leaves us blind. Regularly, the cutting edge technology of the century looks promising to elevate our society to this utopia. Once, technologies such as “Mechanical” and “Electrical” were used to solve not all but some of the world’s top issues. Nonetheless, these technologies introduced new sets of complications.

Currently, we are in the digital era and the task to create a utopia on earth is on the shoulders of computer science, specifically machine learning. Although the image of an advanced world is pretty, we should not be negligent of the consequences. Some of these consequences have been discussed, immensely. However, a less conversed topic is the amount of pollution that ML and AI make.



**Figure 1: Historical carbon dioxide emissions from fossil fuel**

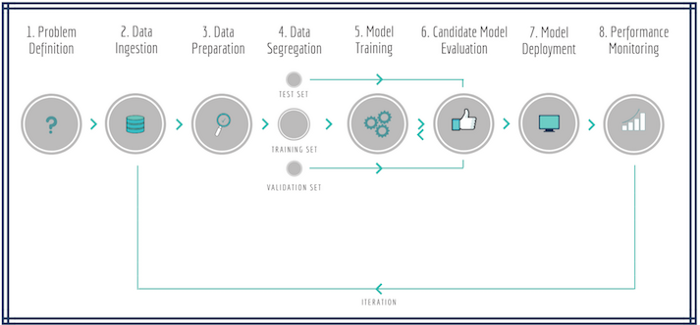
Figure 1 shows the problematic growth rate of carbon emissions from fossil fuel consumption throughout 3 centuries. It is explicitly clear that the share of technological revolutions is not negligible. But what part of the recent decade of progressing towards machine learning plays in this rapid movement to the catastrophe?

To answer this question first I need to explain how machine learning is being trained.

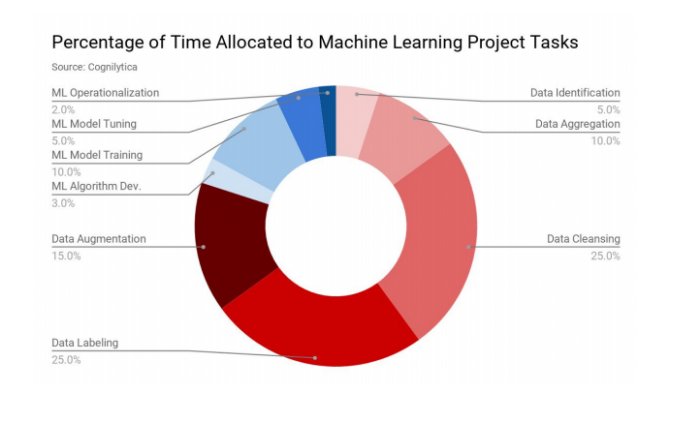
# **2 How an ML model works**

“Machine learning is programming computers to optimize a performance criterion using example data or past experience” [1](#_REFERENCES). ML tries to imitate the human brain and beyond to determine an answer to a question. This question might be “What shape is in the image?”, “What sentence is correct to express a specific idea?”, “Which gene is mutated?”, “What might be the price of a stock, tomorrow?”, “How many different sounds are in the audio file?”, etc. To acquire a proper response by a human to any of these questions, a collection of past observations is required. In machine learning, this collection is called the training data set.

An ML model is similar to an inexperienced child. It needs to be trained by an Algorithm to obtain the skillset of guessing the outcome of some inputs, correctly. However, the training process consumes a large amount of time and energy, especially when higher accuracy is desirable. However, the training is not the most time-consuming part of the ML cycle.



**Figure 2: Machine learning cycle**



**Figure 3: Time consumption percentage of ML tasks**

Data scientists who use machine learning in their problems, mostly spend their time dealing with the data itself. Although, only 10% of the consumed time is related to the model training, yet the amount of it cannot be unnoticed.

# **3 Duration of model training in the machine learning**

This article tends to investigate the carbon emission of the training ML models. For this purpose, I introduce some models and the time required to train them. Then the power consumed by the hardware used in those projects and consequently the amount of carbon emission during the training can be calculated.

## **3.1 Models**

## **3.1.1 Transformer** [**2**](#_REFERENCES)

The authors in the introduction mention “The Transformer is a model architecture eschewing recurrence and instead relying entirely on an attention mechanism to draw global dependencies between input and output. The Transformer allows for significantly more parallelization and can reach a new state of the art in translation quality after being trained for as little as twelve hours on eight P100 GPUs.”

Under chapter 6.1 they report “On the WMT 2014 English-to-German translation task, the big transformer model (Transformer (big) in Table 2) outperforms the best previously reported models (including ensembles) by more than 2.0 BLEU, establishing a new state-of-the-art BLEU score of 28.4. The configuration of this model is listed in the bottom line of Table 3. Training took 3.5 days on 8 P100 GPUs.”

Others picked this model to create more sophisticated codes, one of them is “**Evolved transformer or** **NAS**” [3](#_REFERENCES). Authors say “Specifically, to train a Transformer to peak performance on WMT’14 En-De requires ∼300K training steps, or 10 hours, in the base size when using a single Google TPU V.2 chip, as we do in our search” under chapter 3.3.

Unfortunately, public information about TPU is very limited. Google has not released power consumption rates related to TPUs. By estimation, 300K of 10 hours of TPU compares to 274120 hours on 8 P100 GPUs.

## **3.1.2 ELMo** [**4**](#_REFERENCES)

The duration of training is not mentioned in the original paper of this model. However, in the “Energy and Policy Considerations for Deep Learning in NLP” paper, Chapter 2.1, authors say “Peters et al. (2018) report that ELMo was trained on 3 NVIDIA GTX 1080 GPUs for 2 weeks (336 hours)” [5](#_REFERENCES)

## **3.1.3 BERT** [6](#_REFERENCES)

According to the authors in Appendix A2 “Training of BERT-base was performed on 4 Cloud TPUs in Pod configuration (16 TPU chips total). Training of BERT-large was performed on 16 Cloud TPUs (64 TPU chips total). Each pretraining took 4 days to complete.”

Nvidia claims BERT can be trained in 47 minutes using 92 DGX-2H each having 16 V100 GPUs [**7**](#_REFERENCES). This means the training time is almost three days with one DGX-2H card.

## **3.1.4 GPT-2** [**8**](#_REFERENCES)

The paper of this model does not talk about the duration of the training, too. But authors of “Energy and Policy Considerations for Deep Learning in NLP” say “The large model described in Radford et al. (2019) has 1542M parameters and is reported to require 1 week (168 hours) of training on 32 TPUv3 chips” in chapter 2.1. Since the power consumption of TPU cannot be calculated I ignore this model.

## **3.2 Shortage of data**

These models are a few of many state-of-art models existing to the date. For more information and a list of all the models go to this [webpage](https://paperswithcode.com/sota).

The reason most of these models are not in my article is that most of the ML researchers do not care about the power consumption of their model during the training. Authors of the paper “Estimation of energy consumption in machine learning” [**9**](#_REFERENCES) say “Computer architecture researchers have been investigating energy consumption for decades, especially to be able to deliver state-of-the-art energy efficient processors. Machine learning researchers, on the other hand, have been mainly focused on producing high accurate models without considering energy consumption as an important factor”

# **4 Power consumption and carbon emission**

Now that we have the information regarding how long training of the ML model takes, we can show the power consumption during the training period. Components involved in the process of training are CPU, GPU, RAM. Consequently, the average power consumption of them can be identified as pc, pg, and pr. Servers can have multiple GPUs (and CPUs) so let G and C be the count of them, let t be the duration of the system being running. For the cooling system required to cool down the system and power usage effectiveness of the servers, I use PUE coefficient of 1.58. The equation will be . The carbon emission of consuming 1 KW power is 0.99 pound [10](#_REFERENCES).

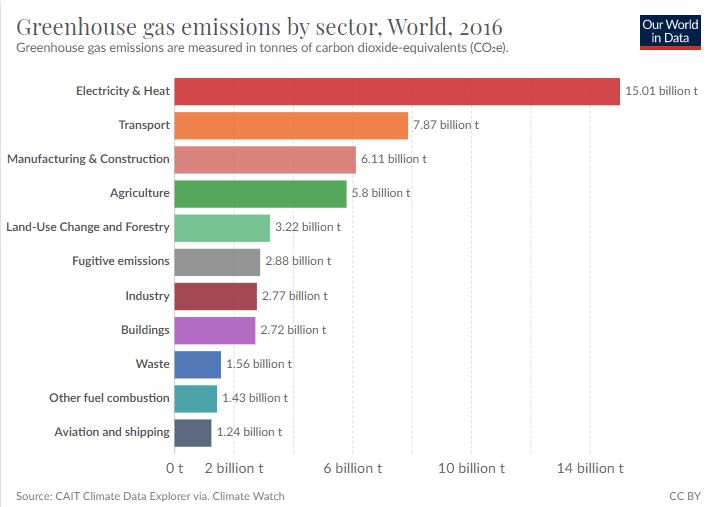
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | GPU | Duration(h) | Power consumption(kWh) | Carbon emission(lbs.) |
| Transformer base | P100x8 | 12 | 27 | 26.73 |
| Transformer big | P100x8 | 84 | 201 | 198.99 |
| NAS | P100x8 | 274120 | 656347 | 649783.53 |
| ELMo | P100x3 | 336 | 275 | 272.25 |
| BERT | 92x(V100x16) | 0.78 | 1507 | 1491.93 |

# **5 Comparing the carbon emissions**

Now that we have the amount of carbon emission of these models let us compare it to pollution made by other industries to see how big is the impact of machine learning on our environment.

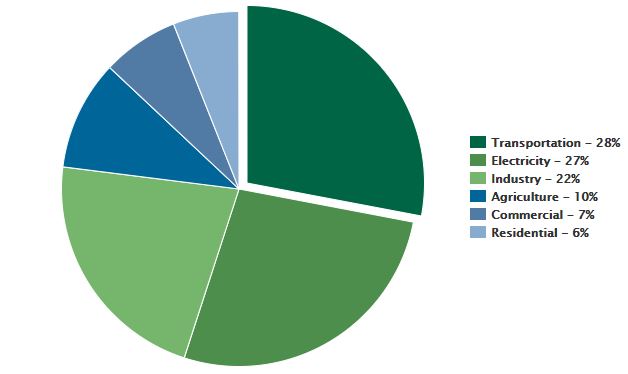
# **5.1 Transportation**

Transportation is the next most pollution generating sector after the “electricity and heat” in the world.



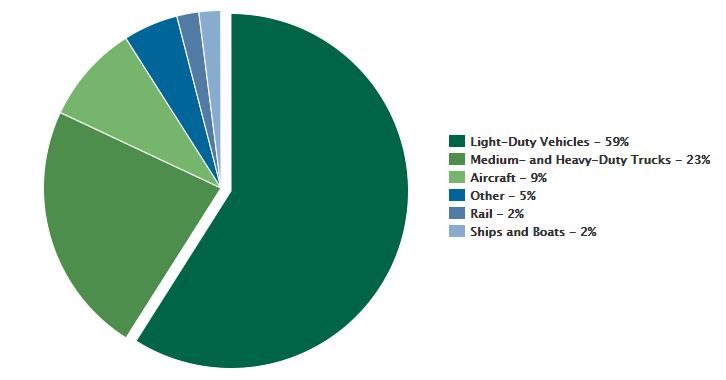
**Figure 4: Greenhouse gas emissions by sector** [**11**](#_REFERENCES)

If we separate electricity from heat (residential and industrial), transport is producing the most amount of greenhouse gas in the US.



**Figure 5: 2018 U.S. GHG Emissions by Sector** [**12**](#_REFERENCES)

There are two major ways of moving people around, cars, and airplanes. Between them, light-duty vehicles pollute the environment more. The reason is that almost every family has one car to commute to work or other necessary places on a daily basis.



**Figure 6: 2018 U.S. Transportation Sector GHG Emissions by Source** [**12**](#_REFERENCES)

The following sections discuss the amount of carbon made by these two transportation ways.

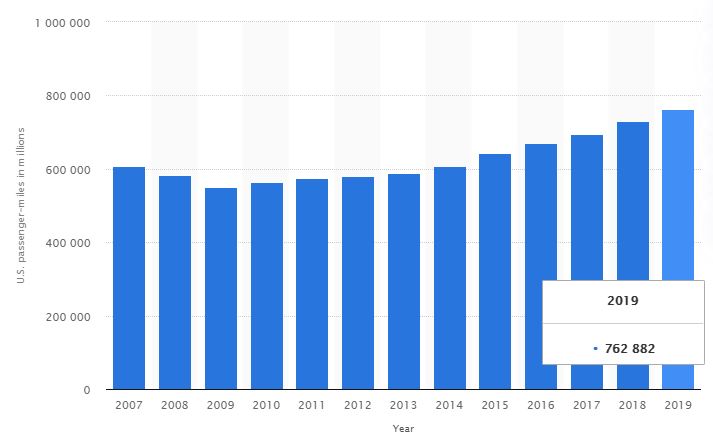
# **5.1.1 Cars**

Based on the United States Environmental Protection Agency (EPA) “A typical passenger vehicle emits about 4.6 metric tons of carbon dioxide per year” [13](#_REFERENCES)The assumption is a typical passenger car travels 11500 miles annually burning a gallon of fuel every 22 miles, every gallon of burned fuel emits 8887 grams of carbon dioxide. This is almost 10241.5 lbs. of carbon emission for an average car, annually. An article on “MIT technology review” has compared the carbon emission of training the NAS model with the carbon emitted from an average passenger car within its lifespan (11.8 years) [14](#_REFERENCES). The article concludes the ratio is 1 (NAS) to 5 (Car) [15](#_REFERENCES). However, in the year 2019, the number of vehicles operating on the roads of the US was 278 million [14](#_REFERENCES). Although these vehicles can be trucks or old cars with higher MPG, let us use the same carbon emission for them. The assumptions give us the enormous number of 2,847,137,000,000 lbs. of emitted carbon for all the vehicles in the US, annually! That is almost 4.4 million times of the carbon emitted from training the NAS model.

# **5.1.2 Airplanes**

In the previous section, I compared the carbon emission of training a sophisticated ML model and vehicles in the United States. Vehicles play a major part in transportation within the US (Almost 82%). The next sector is aircraft with a share of 9% of total transportation.

There are many different aircraft flying. The fuel burned during a flight varies based on the speed, weather, and different parts of a flight such as taxiing, take-off, cruising, and landing. The average amount of emitted carbon dioxide per passenger-mile is 0.24 kg for a short flight and 0.18 kg for a long flight. Let us use the mean of these which is 0.21 kg or 0.46 lbs.



**Figure 7: U.S. air traffic passenger-miles** [**16**](#_REFERENCES)

Considering the numbers from this section and figure 7, the carbon emitted from air traveling in the US in the year 2019 is 350,925,720,000 lbs. which is 434,665 times of the carbon emission of training the NAS model.

# **6 Conclusion**

The comparison made in section 5 is between a single machine learning model and an entire industrial section within the US. This is not a fair comparison; however, the point of this comparison is to show how big is the bad impact of transportation on our environment compared to the ML model training considering there are not so many state-of-art ML models existing to the day, maybe thousands of them. Not all these model needs a huge amount of time to be trained, even if they did still the carbon emission of them will be 434 times less than pollution made by air traveling or 4400 times less than an average car in a year.

All these aside, mankind needs to be gentle and thoughtful about the environment and its future whether the machine learning (or any other computer-related activity e.g. gaming) is polluting our beautiful earth or other industries. Do not forget, we are trying to improve our lives using AI, not put an end to it!

# **ACKNOWLEDGMENTS**

We would like to thank Professor Kristin Tufte, the instructor for this course, for providing advice, ideas, and resources for this project.

# **REFERENCES**

[[1]](#_2_How_an) Ethem Alpaydin. 2020. Introduction to machine learning, fourth edition

[[2]](#_3.1.1__) Vaswani et al. 2017. [Attention Is All You Need](https://arxiv.org/pdf/1706.03762.pdf)

[[3]](#_3.1__) So et al. 2019. [The Evolved Transformer](https://arxiv.org/pdf/1901.11117.pdf)

[[4]](#_3.1.2__) Peters et al. 2018. [Deep contextualized word representations](https://arxiv.org/pdf/1802.05365.pdf)

[[5]](#_3.2__) Strubell et al. 2019. [Energy and Policy Considerations for Deep Learning in NLP](https://arxiv.org/pdf/1906.02243.pdf)

[[6]](#_3.1.3__) Devlin et al. 2019. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](https://arxiv.org/pdf/1810.04805.pdf)

[[7]](#_3.3__) [NVIDIA - NVIDIA Clocks World’s Fastest BERT Training Time and Largest Transformer Based Model, Paving Path For Advanced Conversational AI](https://developer.nvidia.com/blog/training-bert-with-gpus/)

[[8]](#_3.1.4__) Radford et al. 2019. [Language Models are Unsupervised Multitask Learners](https://www.ceid.upatras.gr/webpages/faculty/zaro/teaching/alg-ds/PRESENTATIONS/PAPERS/2019-Radford-et-al_Language-Models-Are-Unsupervised-Multitask-%20Learners.pdf)

[[9]](#_3.2___1) García-Martín et al. 2019. [Estimation of energy consumption in machine learning](https://www.sciencedirect.com/science/article/pii/S0743731518308773)

[[10]](#_4_Power_consumption) [EIA - How much carbon dioxide is produced per kilowatthour of U.S. electricity generation?](https://www.eia.gov/tools/faqs/faq.php?id=74&t=11#:~:text=In%202018%2C%20total%20U.S.%20electricity,of%20CO2%20emissions%20per%20kWh.)

[[11]](#_5.1__) [Our world in data - CO₂ and Greenhouse Gas Emissions](https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions)

[[12]](#_5.1__) [EPA - Fast Facts on Transportation Greenhouse Gas Emissions](https://www.epa.gov/greenvehicles/fast-facts-transportation-greenhouse-gas-emissions)

[[13]](#_5.1.1__) [EPA - Greenhouse Gas Emissions from a Typical Passenger Vehicle](https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle)

[[14]](#_5.1.1__) [USA today - Old cars everywhere: Average vehicle age hits all-time high](https://www.usatoday.com/story/money/cars/2019/06/28/average-vehicle-age-ihs-markit/1593764001/)

[[15]](#_5.1.1__) [MIT technology review - Training a single AI model can emit as much carbon as five cars in their lifetimes](https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/)

[[16]](#_5.1.2__) [Statista - U.S. air traffic passenger-miles from 2007 to 2019](https://www.statista.com/statistics/185744/us-passenger-miles-in-air-traffic-since-1990/)