

Become a Full-Stack Data Scientist 🧑🔬 Avail Data Science Scholarship 🎓

Avail Now

✕

[Home](#)

The Carbon Footprint of AI and Deep Learning



👤 [Prateek Majumder](#) — Published On March 1, 2022
[Beginner](#) [Deep Learning](#)

This article was published as a part of the [Data Science Blogathon](#).

There are immense computational costs of Deep Learning and AI. Artificial intelligence algorithms, which power some of technology’s most cutting-edge applications, such as producing logical stretches of text or creating visuals from descriptions, may need massive amounts of computational power to train. This, in turn, necessitates a vast quantity of power, prompting many to fear that the carbon footprint of these increasingly popular ultra-large A.I. systems would render them environmentally unsustainable.

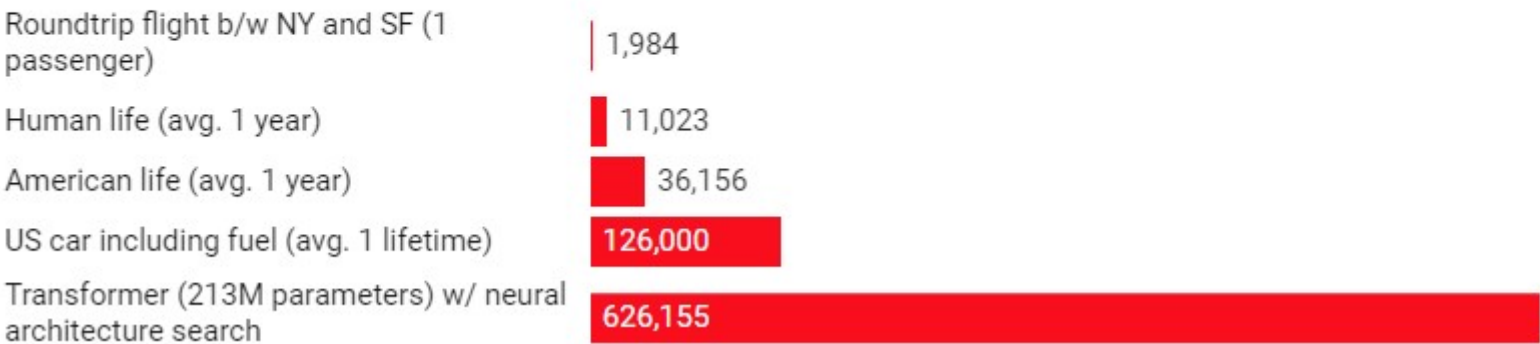
AI and Energy

The artificial intelligence sector is sometimes likened to the oil industry: data, like oil, can be a tremendously profitable commodity if collected and processed. Researchers from the University of Massachusetts, Amherst, conducted a life cycle analysis for training several typical big AI models in a recent publication. They discovered that the procedure may produce almost 626,000 pounds of CO2 equivalent.

(Source: MIT Technology Review)

Common carbon footprint benchmarks

in lbs of CO2 equivalent



(Image: <https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/>)

Modern AI models consume a tremendous amount of energy, and their energy demands are rapidly increasing. The computing resources required to create a best-in-class AI model have doubled every 3 to 4 months in the deep learning era.

From: Rob Toews for Forbes Jun 17, 2020,11:54am EDT ([Link Below](#)).

Today, AI has a significant carbon footprint, and if current market trends continue, it will soon be much worse. Artificial intelligence might become an adversary in the battle against climate change in the years ahead unless we are ready to examine and revise today's AI development programme.

Understanding the Carbon Footprint

Artificial intelligence is becoming a more vital part of the research, health, and even our everyday life. Deep learning is used in chatbots, digital assistants, and streaming service movie and music recommendations. Deep learning is a process in which computer models are trained to spot patterns in data.



(Image: <https://www.pexels.com/photo/female-software-engineer-coding-on-computer-3861972/>)

The GPT-3 model has a staggering 175 billion parameters. To put this amount in context, its previous model GPT-2 had just 1.5 billion parameters when it was introduced, which was considered cutting-edge at the time. GPT-2 took a few dozen petaflop-days to train, which was already a large amount of computing input, but GPT-3 took thousands.

The datasets utilised to train these algorithms are becoming increasingly large. After being trained on a dataset of 3 billion words, the BERT model achieved best-in-class NLP performance in 2018. Based on a training set of 32 billion words, XLNet surpassed BERT. GPT-2 was trained on a dataset of 40 billion words not long after that. GPT-3 was trained using a weighted dataset of around 500 billion words, which dwarfed all prior attempts.

Such exponential increase in training data leads to the rising carbon footprint of AI and Deep Learning. For each piece of data they are fed during training, neural networks perform a lengthy sequence of mathematical operations (both forward and reverse propagation), adjusting their parameters in sophisticated ways. Larger datasets necessitate increased computation and energy requirements.

The process of deploying AI models to take action in real-world settings, known as inference, requires considerably more energy than training. Indeed, Nvidia believes that inference, rather than training, accounts for 80% to 90% of the cost of a neural network.

Research regarding The Carbon Footprint of AI and Deep Learning

Given the everyday consequences of climate change, the consensus is growing on the necessity for AI research ethics to include a focus on limiting and offsetting the study’s carbon imprint. Along with time, accuracy, and other factors, researchers should include the cost of energy in research paper results. In a recent study report published by MIT researchers, the process of deep learning outsizing the environmental effect was further underlined. Researchers conducted a life cycle study for training many typical big AI models in the article “Energy and Policy Considerations for Deep Learning in NLP.

Transformer, ELMo, BERT, and GPT-2 are four models in the field that have been responsible for the most significant improvements in performance. They trained for up to a day on a single GPU to determine its power usage. They then calculated the total energy spent during the whole training procedure using the number of training hours stated in the model’s original papers. Based on the average energy mix in the United States, which roughly matches the energy mix utilised by Amazon’s AWS, the largest cloud services provider, that amount was translated into pounds of carbon dioxide equivalent.

| | Date of original paper | Energy consumption (kWh) | Carbon footprint (lbs of CO2e) | Cloud compute cost (USD) |
|---|------------------------|--------------------------|--------------------------------|--------------------------|
| Transformer (65M parameters) | Jun, 2017 | 27 | 26 | \$41-\$140 |
| Transformer (213M parameters) | Jun, 2017 | 201 | 192 | \$289-\$981 |
| ELMo | Feb, 2018 | 275 | 262 | \$433-\$1,472 |
| BERT (110M parameters) | Oct, 2018 | 1,507 | 1,438 | \$3,751-\$12,571 |
| Transformer (213M parameters) w/ neural architecture search | Jan, 2019 | 656,347 | 626,155 | \$942,973-\$3,201,722 |
| GPT-2 | Feb, 2019 | - | - | \$12,902-\$43,008 |

(Image: <https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/>)

The researchers discovered that the environmental costs of training increased in direct proportion to model size. It grew rapidly when more tuning steps were applied to improve the model’s ultimate accuracy. Neural architecture search, in particular, has large associated costs for little performance advantage. Neural architecture search is a tuning procedure that attempts to improve a model by progressively altering the design of a neural network via extensive trial and error. The researchers also stated that these data should only be used as a starting point. In practice, AI researchers either create a new model from start or adapt an existing model to a new data set, both of which need many additional rounds of training and tweaking.

Reducing the Carbon Footprint

- Make use of computationally efficient machine learning algorithms. This is possible because quantization technologies can bring CPUs up to speed with GPUs and TPUs. One possibility is to use resource-constrained devices, such as IoT devices, for edge computing.
- Don't train a model from scratch if it is not necessary. Even though the majority of researchers don't train from scratch, training a heavy model from scratch can burn up a lot of CPU/GPU power. Hence, using a pre-trained model can reduce the carbon footprint significantly as less electricity is necessary. Azure offers various sophisticated pre-trained models for vision, speech, language & search. These are already trained on massive datasets using many compute hours training the model, the end-user only has to transfer the knowledge of these models to their own dataset. This way only a few compute hours and power are spent. There is a side note to this, model inferences have been indicated to be a bigger consumer of energy than the training itself based on research done by Nvidia in 2019.
- When necessary, use Automated ML. Whereas machine learning was formerly primarily a laboratory activity, an increasing number of businesses are beginning to employ technology outside of the laboratory. This implies that determining which model is the most efficient and quickly eliminating the less efficient models not only saves money on the computational side but also minimizes the carbon footprint.
- The use of federated learning (FL) methods can be advantageous in the area of carbon footprint. For example, a study conducted by experts at the University of Cambridge, FL has an advantage because of the cooling requirements of data centers. Though GPUs and TPUs are becoming more efficient in terms of processing power given per unit of energy spent, the requirement for a powerful and energy-consuming cooling system persists, so the FL can always benefit from hardware developments.
- Deep learning and particular optical neural networks take a lot of energy. Second, the model inference is a significant energy user. Nvidia estimates that model inference costs 80-90 percent of the model cost. One solution is to use customized processors that increase the speed and efficiency of training and testing neural networks. As a result, having the proper hardware is critical.

Conclusion

AI appears to be destined for a dual function. On the one hand, technology has the potential to help mitigate the consequences of the climate problem, such as through smart grid design, low-emission infrastructure development, and climate change modelling. AI, on the other hand, is a big carbon emitter.

The media shown in this article is not owned by Analytics Vidhya and are used at the Author's discretion.

[blogathon](#) [Carbon Footprint](#)

The advertisement is a vertical banner with a dark blue background. At the top left is the Analytics Vidhya logo. Below it, the text 'DATA SCIENCE IMMERSIVE BOOTCAMP' is written in white. The middle section features three yellow horizontal bars, each with a white icon and text: a person icon for '100% Batch Placed', a money bag icon for '250% Avg. Salary Hike', and a briefcase icon for '8.3 L.P.A. Avg Salary'. At the bottom is a white button with the text 'Enroll Now'.

About the Author



Prateek Maiumder

We use cookies on Analytics Vidhya websites to deliver our services, analyze web traffic, and improve your experience on the site. By using Analytics Vidhya, you agree to our [Privacy Policy](#) and [Terms of Use](#). [Accept](#)

Our Top Authors



Download

Analytics Vidhya App for the Latest blog/Article



Previous Post

[Gaussian Naive Bayes Algorithm for Credit Risk Modelling](#)

Next Post

[What are Graph Neural Networks, and how do they work?](#)

Leave a Reply

Your email address will not be published. Required fields are marked *

Comment

Name*

Email*

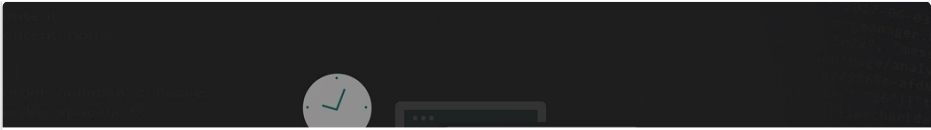
Website

☒ Notify me of follow-up comments by email.


☒ Notify me of new posts by email.


Submit

Top Resources



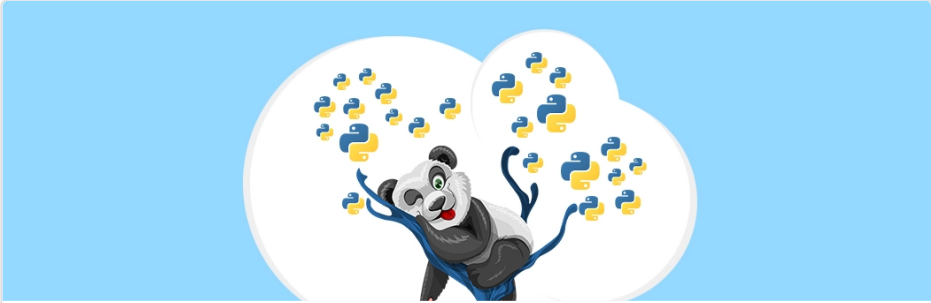
[Python Tutorial: Working with CSV file for Data Science](#)

 [Harika Bonthu](#) - AUG 21, 2021




[Boost Model Accuracy of Imbalanced COVID-19 Mortality Prediction Using GAN-based..](#)

[Bala Gangadhar Thilak Adiboina](#) - OCT 07, 2020




[Joins in Pandas: Master the Different Types of Joins in..](#)


[Abhishek Sharma](#) - FEB 27, 2020





[Understanding Random Forest](#)



[Sruthi E R](#) - JUN 17, 2021



Download App



Analytics Vidhya

[About Us](#)[Our Team](#)[Careers](#)[Contact us](#)





Companies

[Post Jobs](#)[Trainings](#)[Hiring Hackathons](#)[Advertising](#)

Data Scientists

[Blog](#)[Hackathon](#)[Discussions](#)[Apply Jobs](#)

Visit us



© Copyright 2013- 2022 Analytics Vidhya.

[Privacy Policy](#) | [Terms of Use](#) | [Refund Policy](#)