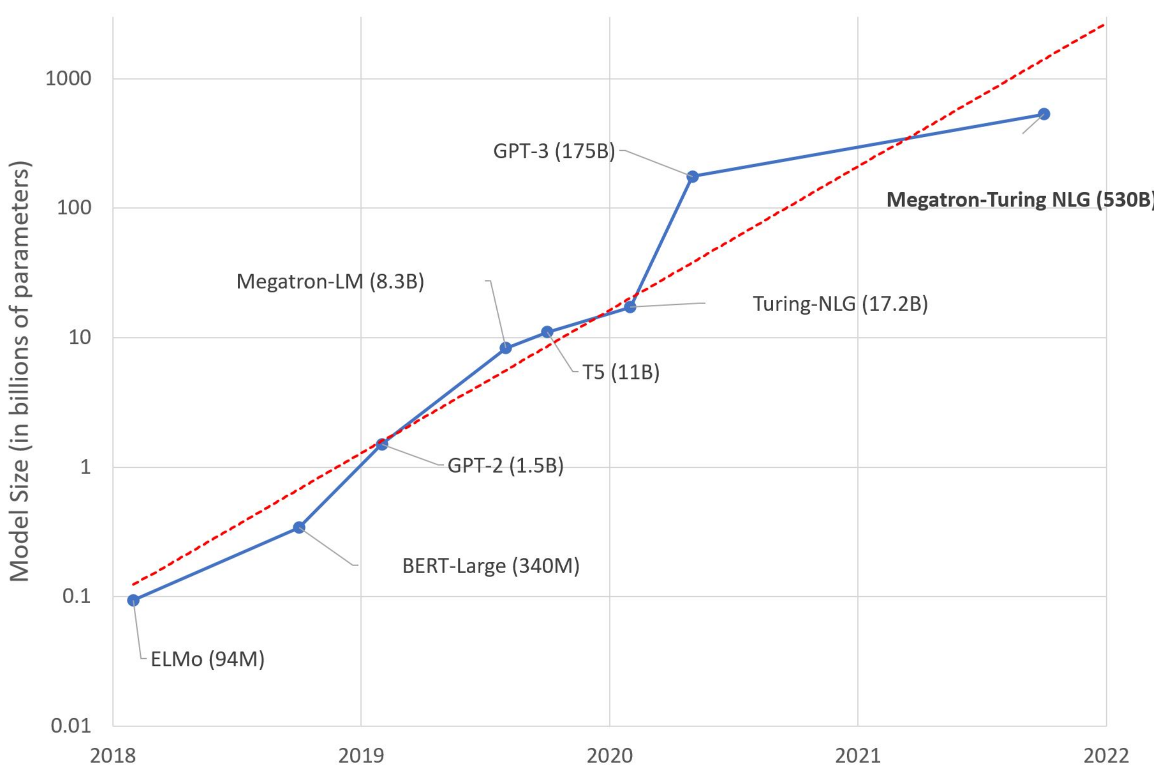
The Carbon Price of AutoML: CO_2

Carbon Cost of ML Training

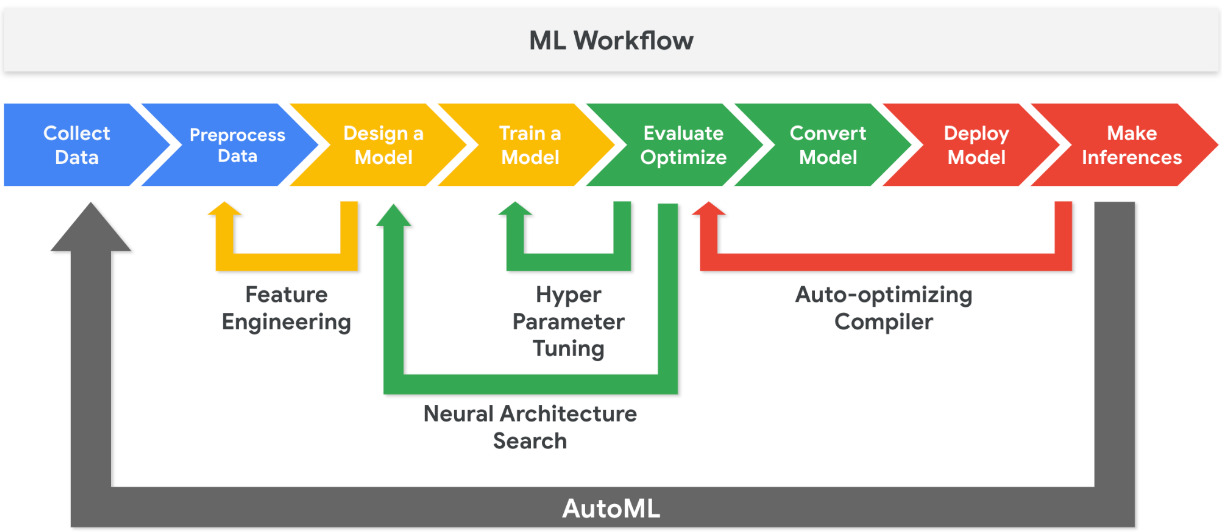
In recent years, machine learning has undergone a bifurcation, a splitting into two scales: big and small. At the small scale, we have TinyML, where model size and efficient computation are crucial for models to be feasibly deployed onto resource-constrained microcontrollers. At the opposite end, we have large-scale neural networks trained on cloud computing clusters, using dozens or even hundreds of GPUs. At smaller scales, such as those in TinyML or on an individual laptop, the carbon footprint of models does not have a significant footprint. However, at a larger scale, these models can have carbon footprints that rival typical emission sources such as planes, trains, and automobiles.

Models developed for natural language processing (NLP) are of particular concern since these seem to exhibit a performance improvement that scales proportionally with model size. This means that NLP researchers will continue to produce larger and larger natural language models, since currently no limit has been reached where performance starts to degrade with size. In principle, this could lead to networks with billions or even trillions of parameters. Computational resources also scale proportionally with the number of parameters, meaning that larger models will begin to have increasingly large carbon footprints.



AutoML Can Exponentially Increases The Carbon Cost

This is quite concerning, especially since these models already account for a substantial portion of carbon emissions from data centers. However, the incorporation of AutoML with large-scale language models such as these adds even more fuel to the fire, so to speak. As we have seen, AutoML allows us to automate the optimization of a model in terms of feature engineering, hyperparameter tuning, neural architecture search (NAS), and auto-optimizing compilation. Since large-scale language models already take a significant amount of resources to train, they take an even more significant amount of resources to optimize, in some cases orders of magnitude more resources.



A study by [Strubell, Ganesh, and McCallum](#) (2019) estimates that training a large-scale transformer with NAS has a carbon footprint equivalent to that of five cars over their working life. Performing NAS on this transformer required over 3000x the resources compared to training the original transformer model. Since NAS is only a subsection of AutoML, one could easily imagine the resources being required for AutoML being even more significant. Not only that, but training this model cost upward of $1 million! Clearly, the cost and carbon footprint of training large-scale models such as these need to be considered in their development.

Chances are, you will not be running NAS on a large-scale language model with billions of parameters. However, it is important to understand the environmental impact that such models are beginning to have due to their inherent scale, and to be cognizant of this when building your own models.

Additional Resources

Good News About the Carbon Footprint of Machine Learning Training
Chasing Carbon: The Elusive Environmental Footprint of Computing