

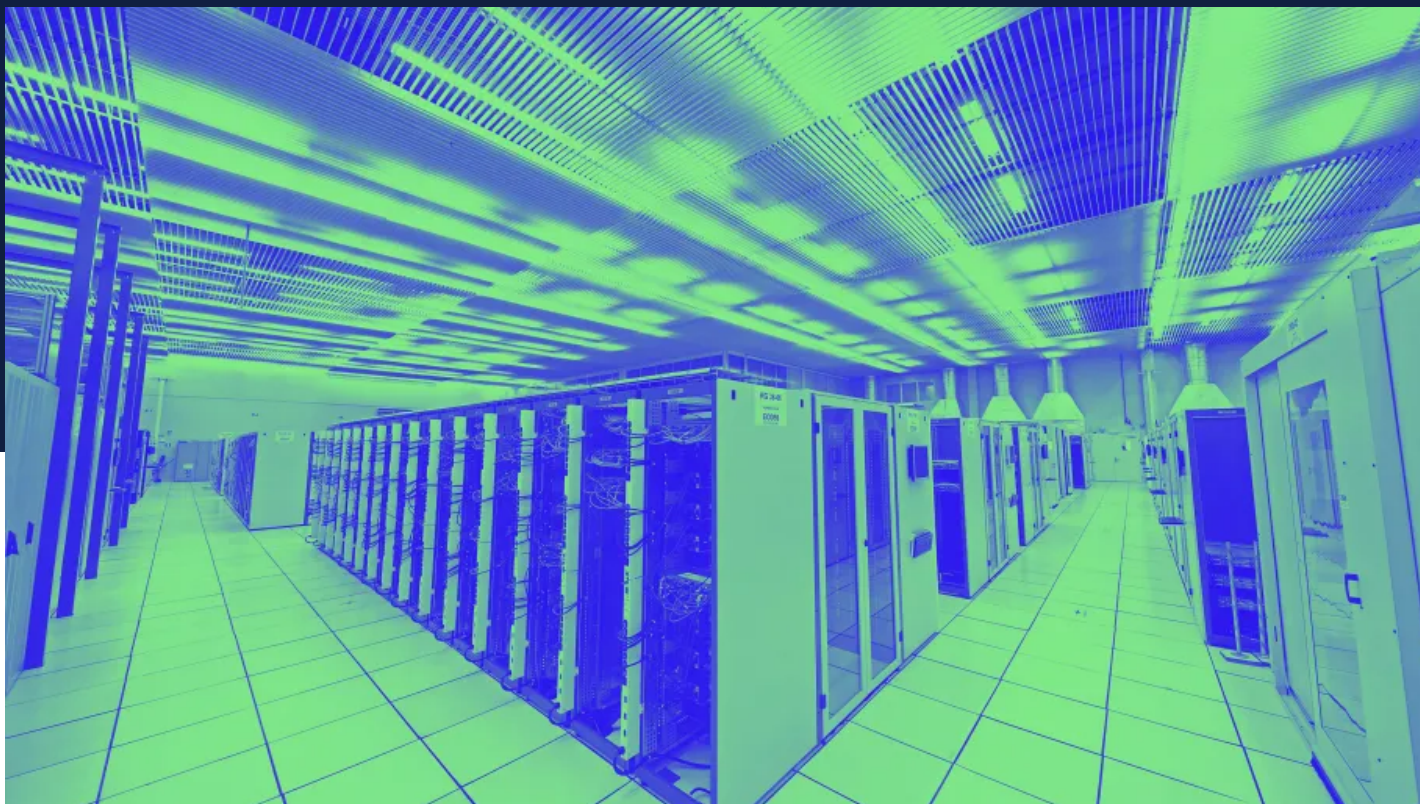
## ARTIFICIAL INTELLIGENCE

## Training a single AI model can emit as much carbon as five cars in their lifetimes

Deep learning has a terrible carbon footprint.

By Karen Hao

June 6, 2019



A data center

commodity. Now it seems the metaphor may extend even further. Like its fossil-fuel counterpart, the process of deep learning has an outsize environmental impact.

In a [new paper](#), researchers at the University of Massachusetts, Amherst, performed a life cycle assessment for training several common large AI models. They found that the process can emit more than 626,000 pounds of carbon dioxide equivalent—nearly five times the lifetime emissions of the average American car (and that includes manufacture of the car itself).

Advertisement

## Common carbon footprint benchmarks

in lbs of CO2 equivalent

Roundtrip flight b/w NY and SF (1 passenger)	1,984
Human life (avg. 1 year)	11,023
American life (avg. 1 year)	36,156
US car including fuel (avg. 1 lifetime)	126,000
Transformer (213M parameters) w/ neural architecture search	626,155

Chart: MIT Technology Review • Source: Strubell et al. • [Created with Datawrapper](#)

It's a jarring quantification of something AI researchers have suspected for a long time. "While probably many of us have thought of this in an abstract, vague level, the figures really show the magnitude of the problem," says Carlos Gómez-Rodríguez, a computer scientist at the University of A Coruña in Spain, who was not involved in the research. "Neither I nor other researchers I've discussed them with thought the environmental impact was that substantial."

## The carbon footprint of natural-language processing

 **Subscribe & save 17% for exclusive access to our NEW 15 Climate Tech Companies list.**

The paper specifically examines the model training process for [natural-language processing](#) (NLP), the subfield of AI that focuses on teaching machines to handle human language. In the last two years, the NLP community has reached several noteworthy performance milestones in machine translation, sentence completion, and other standard benchmarking tasks. [OpenAI's infamous GPT-2 model](#), as one example, excelled at writing convincing fake news articles.

But such advances have required training ever larger models on sprawling data sets of sentences scraped from the internet. The approach is computationally expensive—and highly energy intensive.

The researchers looked at four models in the field that have been responsible for the biggest leaps in performance: the Transformer, ELMo, [BERT](#), and GPT-2. They trained each on a single GPU for up to a day to measure its power draw. They then used the number of training hours listed in the model's original papers to calculate the total energy consumed over the complete training process. That number was converted into pounds of carbon dioxide equivalent based on the average energy mix in the US, which closely matches the energy mix used by Amazon's AWS, the largest cloud services provider.

MIT Technology Review

SUBSCRIBE

# The estimated costs of training a model once

In practice, models are usually trained many times during research and development.

	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

*Note: Because of a lack of power draw data on GPT-2's training hardware, the researchers weren't able to calculate its carbon footprint.*  
Table: MIT Technology Review • Source: Strubell et al. • Created with [Datawrapper](#)

They found that the computational and environmental costs of training grew proportionally to model size and then exploded when additional tuning steps were used to increase the model’s final accuracy. In particular, they found that a tuning process known as neural architecture search, which tries to optimize a model by incrementally tweaking a neural network’s design through exhaustive trial and error, had extraordinarily high associated costs for little performance benefit. Without it, the most costly model, BERT, had a carbon footprint of roughly 1,400 pounds of carbon dioxide equivalent, close to a round-trip trans-America flight for one person.

What’s more, the researchers note that the figures should only be considered as baselines. “Training a single model is the minimum amount of work you can do,” says Emma Strubell, a PhD candidate at the University of Massachusetts, Amherst, and the lead author of the paper. In practice, it’s much more likely that AI researchers would develop a new model from scratch or adapt an existing model to a new data set, either of which can require many more rounds of training and tuning.

To get a better handle on what the full development pipeline might look like in terms of carbon footprint, Strubell and her colleagues used a model they’d produced in a [previous paper](#) as a case study. They found that the process of building and testing a final paper-worthy model required training 4,789 models over a six-month period. Converted to CO<sub>2</sub> equivalent, it emitted more than 78,000 pounds and is likely representative of typical work in the field.

Advertisement

MIT Technology Review

SUBSCRIBE

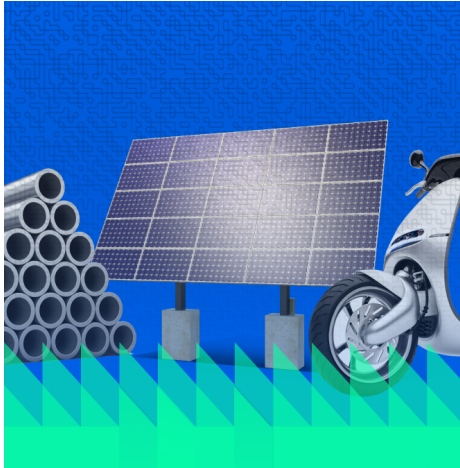
companies and institutions that have abundant access to computational resources can leverage this to obtain a competitive

advantage,” Gómez-Rodríguez says. “This kind of analysis needed to be done to raise awareness about the resources being spent [...] and will spark a debate.”

“What probably many of us did not comprehend is the scale of it until we saw these comparisons,” echoed Siva Reddy, a postdoc at Stanford University who was not involved in the research.

### The privatization of AI research

The results underscore another growing problem in AI, too: the sheer intensity of resources now required to produce paper-worthy results has made it increasingly challenging for people working in academia to continue contributing to research.



#### Subscribe for Exclusive List Access

Discover the climate tech companies dedicated to making our world a better place for generations to come.

**SUBSCRIBE & SAVE 17%**

“This trend toward training huge models on tons of data is not feasible for academics—grad students especially, because we don’t have the computational resources,” says Strubell. “So there’s an issue of equitable access between researchers in academia versus researchers in industry.”

Strubell and her coauthors hope that their colleagues will heed the paper’s findings and help level the playing field by investing in developing more efficient hardware and algorithms.

Reddy agrees. “Human brains can do amazing things with little power consumption,” he says. “The bigger question is how can we build such machines.” **T**

by Karen Hao



#### Subscribe for Exclusive List Access

Discover the climate tech companies dedicated to making our world a better place for generations to come.

MIT Technology Review

**SUBSCRIBE**

## DEEP DIVE

## ARTIFICIAL INTELLIGENCE

**DeepMind's cofounder: Generative AI is just a phase. What's next is interactive AI.**

"This is a profound moment in the history of technology," says Mustafa Suleyman.

By Will Douglas Heaven

**Deepfakes of Chinese influencers are livestreaming 24/7**

With just a few minutes of sample video and \$1,000, brands never have to stop selling their products.

By Zeyi Yang

**AI hype is built on high test scores. Those tests are**

MIT Technology Review

SUBSCRIBE

can't do.

By Will Douglas Heaven

Achieve "ML democratization" and generate proof points by intaking problem statements, not technical requirements.

By David Kang

## STAY CONNECTED

Illustration by Rose Wong

**Get the latest updates from  
MIT Technology Review**

Discover special offers, top stories, upcoming events, and more.

Enter your email

[Privacy Policy](#)**The latest iteration of a legacy**

Founded at the Massachusetts Institute of Technology in 1899, MIT Technology Review is a world-renowned, independent media company whose insight, analysis, reviews, interviews and live events explain the newest technologies and their commercial, social and political impact.

MIT Technology Review

SUBSCRIBE

Advertise with MIT Technology Review



Elevate your brand to the forefront of conversation around emerging technologies that are radically transforming business. From event sponsorships to custom content to visually arresting video storytelling, advertising with MIT Technology Review creates opportunities for your brand to resonate with an unmatched audience of technology and business elite.

[ADVERTISE WITH US](#)[About us](#)[Careers](#)[Custom content](#)[Advertise with us](#)[International Editions](#)[Republishing](#)[MIT News](#)[Help & FAQ](#)[My subscription](#)[Editorial guidelines](#)[Privacy policy](#)[Terms of Service](#)[Write for us](#)[Contact us](#)

© 2023 MIT Technology Review

MIT Technology Review

SUBSCRIBE

Technology  
Review