

Is generative AI bad for the environment? A computer scientist explains the carbon footprint of ChatGPT and its cousins

May 23 2023



AI chatbots and image generators run on thousands of computers housed in data centers like this Google facility in Oregon. Credit: [Tony Webster/Wikimedia, CC BY-SA](#)

Generative AI is the hot new technology behind chatbots and image generators. But how hot is it making the planet?

As an AI researcher, I often worry about the [energy](#) costs of building

artificial intelligence models. The more powerful the AI, the more energy it takes. What does the emergence of increasingly more powerful generative AI models mean for society's future carbon footprint?

"Generative" refers to the ability of an AI algorithm to produce complex data. The alternative is "discriminative" AI, which chooses between a fixed number of options and produces just a single number. An example of a discriminative output is choosing whether to approve a loan application.

Generative AI can create much more complex outputs, such as a sentence, a paragraph, an image or even a short video. It has long been used in applications like smart speakers to generate audio responses, or in autocomplete to suggest a [search query](#). However, it only recently gained the ability to generate humanlike language and realistic photos.

Using more power than ever

The exact energy cost of a single AI model is difficult to estimate, and includes the energy used to manufacture the computing equipment, create the model and use the model in production. In 2019, researchers found that creating a generative AI model called BERT with 110 million parameters consumed the energy of a round-trip transcontinental flight for one person.

The number of parameters refers to the size of the model, with larger models generally being more skilled. Researchers estimated that creating the much larger GPT-3, which has 175 billion parameters, consumed 1,287 megawatt hours of electricity and generated 552 tons of carbon dioxide equivalent, the equivalent of 123 gasoline-powered passenger vehicles driven for one year. And that's just for getting the model ready to launch, before any consumers start using it.

Size is not the only predictor of carbon emissions. The open-access BLOOM model, developed by the BigScience project in France, is similar in size to GPT-3 but has a much lower carbon footprint, consuming 433 MWh of electricity in generating 30 tons of CO₂eq. A study by Google found that for the same size, using a more efficient model architecture and processor and a greener data center can reduce the carbon footprint by 100 to 1,000 times.

Larger models do use more energy during their deployment. There is limited data on the carbon footprint of a single generative AI query, but some industry figures estimate it to be four to five times higher than that of a search engine query. As chatbots and image generators become more popular, and as Google and Microsoft incorporate AI language models into their search engines, the number of queries they receive each day could grow exponentially.

AI bots for search

A few years ago, not many people outside of research labs were using models like BERT or GPT. That changed on Nov. 30, 2022, when OpenAI released ChatGPT. According to the latest available data, ChatGPT had over 1.5 billion visits in March 2023. Microsoft incorporated ChatGPT into its search engine, Bing, and made it available to everyone on May 4, 2023. If chatbots become as popular as search engines, the energy costs of deploying the AIs could really add up. But AI assistants have many more uses than just search, such as writing documents, solving math problems and creating marketing campaigns.

Another problem is that AI models need to be continually updated. For example, ChatGPT was only trained on data from up to 2021, so it does not know about anything that happened since then. The carbon footprint of creating ChatGPT isn't [public information](#), but it is likely much higher than that of GPT-3. If it had to be recreated on a regular basis to

update its knowledge, the energy costs would grow even larger.

One upside is that asking a chatbot can be a more direct way to get information than using a search engine. Instead of getting a page full of links, you get a direct answer as you would from a human, assuming issues of accuracy are mitigated. Getting to the information quicker could potentially offset the increased energy use compared to a search engine.

Ways forward

The future is hard to predict, but large generative AI models are here to stay, and people will probably increasingly turn to them for information. For example, if a student needs help solving a math problem now, they ask a tutor or a friend, or consult a textbook. In the future, they will probably ask a chatbot. The same goes for other expert knowledge such as legal advice or medical expertise.

While a single large AI model is not going to ruin the environment, if a thousand companies develop slightly different AI bots for different purposes, each used by millions of customers, the energy use could become an issue. More research is needed to make generative AI more efficient. The good news is that AI can run on renewable energy. By bringing the computation to where green energy is more abundant, or scheduling computation for times of day when renewable energy is more available, emissions can be reduced by a factor of 30 to 40, compared to using a grid dominated by fossil fuels.

Finally, societal pressure may be helpful to encourage companies and research labs to publish the carbon footprints of their AI models, as some already do. In the future, perhaps consumers could even use this information to choose a "greener" chatbot.

Provided by The Conversation

Citation: Is generative AI bad for the environment? A computer scientist explains the carbon footprint of ChatGPT and its cousins (2023, May 23) retrieved 26 April 2024 from <https://techxplore.com/news/2023-05-generative-ai-bad-environment-scientist.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.