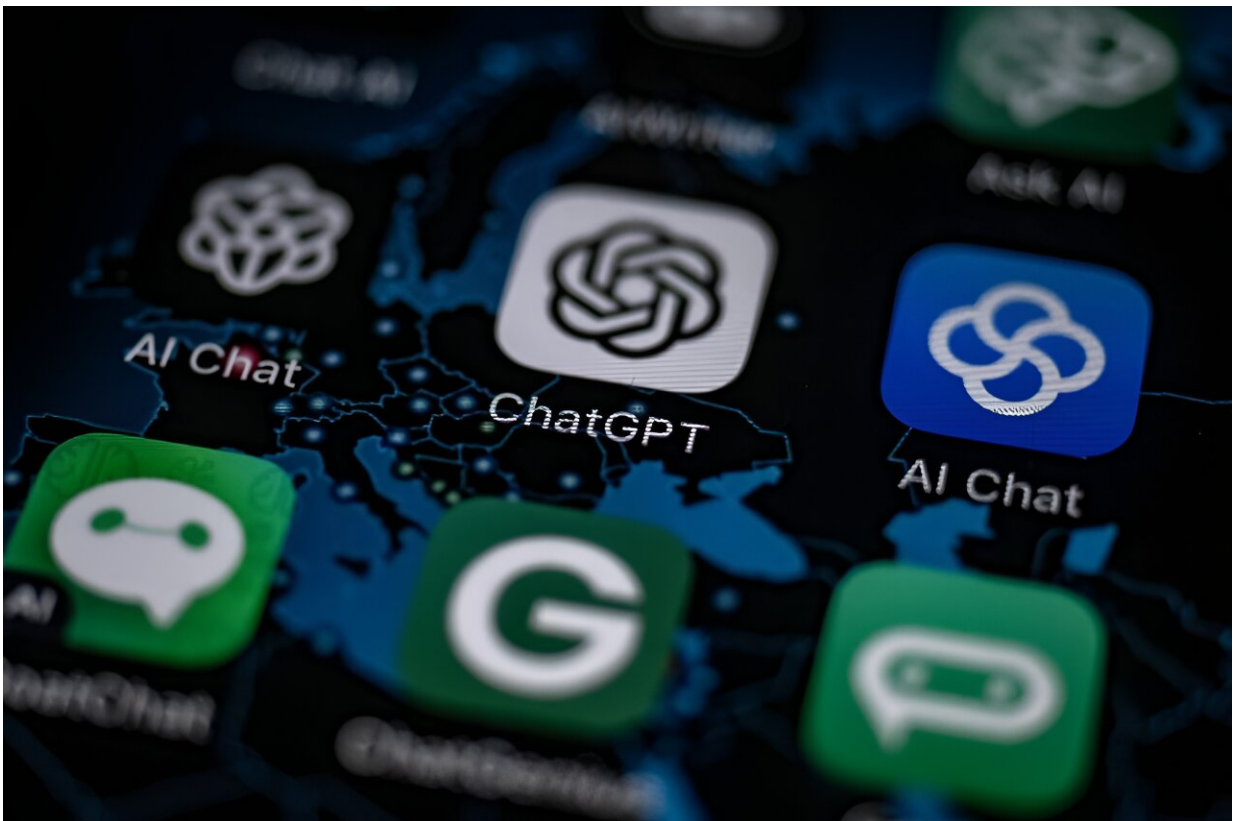


Tech's carbon footprint: can AI revolutionize responsibly?

September 3 2023, by Alex PIGMAN



The transformation of the world's data servers to become AI-ready is well underway.

Across the globe, data servers are humming, consuming both megawatts and precious natural resources to bring life to our digital world.

The planet's 8,000 or so [data centers](#) are the foundation of our online existence, and will grow ever further with the advent of artificial intelligence—so much so that research estimates that by 2025, the IT industry could use 20 percent of all electricity produced, and emit up to 5.5 percent of the world's carbon emissions.

This poses a real—and to some, increasingly urgent—question about the industry's carbon footprint as startups and companies fall behind Silicon Valley's latest forward march.

"Pandora's box is open," said Arun Iyengar, CEO of Untether AI, a highly specialized chip-making company that strives to make AI more energy efficient.

"We can utilize AI in ways that enhance the climate requirements or we can ignore the climate requirements and find ourselves facing the consequences in a decade or so in terms of the impact."

The transformation of the world's [data servers](#) to AI readiness is already well underway, in what one Google executive called a "once-in-a-generation inflection point in computing."

But the scope of the mission is huge.

The creation of generative AI tools such as GPT-4, which powers ChatGPT, or Google's Palm2, behind the bot Bard, can be broken into two key stages, the actual "training" and then the execution (or "inference").

In 2019, University of Massachusetts Amherst researchers trained several large language models, and found that training a single AI model can emit the CO₂ emission equivalent of five cars over their lifetimes.

A more recent study by Google and the University of California, Berkeley, reported that training GPT-3 resulted in 552 metric tons of carbon emissions, equivalent to driving a passenger vehicle 1.24 million miles (2 million kilometers).

OpenAI's latest generation model, GPT-4, is trained on around 570 times more parameters—or inputs—than GPT-3, and the scale of these systems will only grow as AI becomes more powerful and ubiquitous.

Nvidia, AI's chip giant, provides the processors that are indispensable for training, known as GPUs. And while they are more energy efficient than typical chips, they remain formidable consumers of power.

The ChatGPT 'problem'

The other side of generative AI is deployment, or inference: when the trained model is applied to identify objects, respond to text prompts or whatever the use case may be.

Deployment doesn't necessarily need the computing heft of a Nvidia chip, but taken cumulatively, the endless interactions in the [real world](#) far outweigh training in terms of workload.

"Inference is going to be even more of a problem now with ChatGPT, which can be used by anyone and integrated into daily life through apps and web searches," said Lynn Kaack, assistant professor of computer science at the Hertie School in Berlin.



Chip giant Nvidia provides the processors that are indispensable for AI training -- and while they are more energy efficient than typical chips, they remain formidable consumers of power.

The biggest cloud companies insist that they are committed to being as energy efficient as possible.

Amazon Web Services pledges to be carbon-neutral by 2040 while Microsoft has pledged to be carbon-negative by 2030.

The latest evidence that the companies are serious about energy efficiency is reassuring.

Between 2010 and 2018, global data center energy use rose by only 6 percent, despite a 550 percent increase in workloads and computing instances, according to the International Energy Agency.

'Backwards' thinking

Silicon Valley's AI tycoons believe discussions of AI's current carbon footprint are beside the point, and underplay its revolutionary potential.

The naysayers have it "backwards," Nvidia CEO Jensen Huang told reporters on a recent visit to his company's headquarters in California.

The mass deployment of AI and faster computing will in the end diminish the need to go to the world's data clouds, he argued.

AI's superpowers will turn your laptop, car or the device in your pocket into an energy-efficient supercomputer without the need to "retrieve" data from the cloud.

"In the future, there'll be a little tiny model that sits on your phone and 90 percent of the pixels will be generated, 10 percent will be retrieved, instead of 100 percent retrieved—and so you're going to save (energy)," he said.

OpenAI's Sam Altman meanwhile believes that AI will soon enough be able to build humanity a completely new future.

"I think once we have a really powerful super intelligence, addressing climate change will not be particularly difficult," Altman said recently.

"This illustrates how big we should dream... Think about a system where you can say, 'Tell me how to make a lot of clean energy cheaply, tell me how to efficiently capture carbon, and tell me how to build a factory to

do this at planetary scale."

But some experts worry that the mad dash for AI has elbowed out fears about the planet, at least for now.

"Large corporations are spending a lot of money right now deploying AI. I don't think they are thinking about the environmental impact yet," said Untether AI's Iyengar.

But, he added, "I think that is coming."

© 2023 AFP

Citation: Tech's carbon footprint: can AI revolutionize responsibly? (2023, September 3) retrieved 25 May 2024 from <https://techxplore.com/news/2023-09-tech-carbon-footprint-ai-revolutionize.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.