

Can the climate survive the insatiable energy demands of the AI arms race?

New computing infrastructure means big tech is likely to miss emissions targets but they can't afford to get left behind in a winner takes all market

Dan Milmo, Alex Hern and Jillian Ambrose

Thu 4 Jul 2024 06.00 BST

Share



Google's emissions have climbed nearly 50% in five years due to AI energy demand. Photograph: Google Handout/EPA

The artificial intelligence boom has driven big tech share prices to fresh highs, but at the cost of the sector's climate aspirations.

Google admitted on Tuesday that the technology is threatening its environmental targets after revealing that datacentres, a key piece of AI infrastructure, had helped increase its greenhouse gas emissions by 48% since 2019. It said "significant uncertainty" around reaching its target of net zero emissions by 2030 – reducing the overall amount of CO₂ emissions it is responsible for to zero – included "the uncertainty around the future environmental impact of AI, which is complex and difficult to predict".

It follows Microsoft, the biggest financial backer of ChatGPT developer OpenAI, admitting that its 2030 net zero “moonshot” might not succeed owing to its AI strategy.

So will tech be able to bring down AI’s environmental cost, or will the industry plough on regardless because the prize of supremacy is so great?

Why does AI pose a threat to tech companies’ green goals?

Datacentres are a core component of training and operating AI models such as Google’s Gemini or OpenAI’s GPT-4. They contain the sophisticated computing equipment, or servers, that crunch through the vast reams of data underpinning AI systems. They require large amounts of electricity to run, which generates CO₂ depending on the energy source, as well as creating “embedded” CO₂ from the cost of manufacturing and transporting the necessary equipment.

According to the International Energy Agency, total electricity consumption from datacentres could double from 2022 levels to 1,000 TWh (terawatt hours) in 2026, equivalent to the energy demand of Japan, while research firm SemiAnalysis calculates that AI will result in datacentres using 4.5% of global energy generation by 2030. Water usage is significant too, with one study estimating that AI could account for up to 6.6bn cubic metres of water use by 2027 – nearly two-thirds of England’s annual consumption.

What do experts say about the environmental impact?

A recent UK government-backed report on AI safety said that the carbon intensity of the energy source used by tech firms is “a key variable” in working out the environmental cost of the technology. It adds, however, that a “significant portion” of AI model training still relies on fossil fuel-powered energy.

Indeed, tech firms are hoovering up renewable energy contracts in an attempt to meet their environmental goals. Amazon, for instance, is the world’s largest corporate purchaser of renewable energy. Some experts argue, though, that this pushes other energy users into fossil fuels because there is not enough clean energy to go round.

“Energy consumption is not just growing, but Google is also struggling to meet this increased demand from sustainable energy sources,” says Alex de Vries, the founder of Digiconomist, a website monitoring the environmental impact of new technologies.

Is there enough renewable energy to go round?

Global governments plan to triple the world’s renewable energy resources by the end of the decade to cut consumption of fossil fuels in line with climate targets. But the ambitious pledge, agreed at last year’s COP28 climate talks, is already in doubt and experts fear that a sharp increase in energy demand from AI datacentres may push it further out of reach.

The IEA, the world’s energy watchdog, has warned that even though global renewable energy capacity grew by the fastest pace recorded in the past 20 years in 2023, the world may only double its renewable energy by 2030 under current government plans.

The answer to AI’s energy appetite may be for tech companies to invest more heavily in building new renewable energy projects to meet their growing power demand.

How soon can we build new renewable energy projects?

Onshore renewable energy projects such as wind and solar farms are relatively fast to build – they can take less than six months to develop. However, sluggish planning rules in many developed countries alongside a global logjam in connecting new projects to the power grid could add years to the process. Offshore windfarms and hydro power schemes face similar challenges in addition to construction times of between two and five years.

This has raised concerns over whether renewable energy can keep pace with the expansion of AI. Major tech companies have already tapped a third of US nuclear power plants to supply low-carbon electricity to their datacentres, according to the Wall Street Journal. But without investing in new power sources these deals would divert low-carbon electricity away from other users leading to more fossil fuel consumption to meet overall demand.

Will AI's demand for electricity grow for ever?

Normal rules of supply and demand would suggest that, as AI uses more electricity, the cost of energy rises and the industry is forced to economise. But the unique nature of the industry means that the largest companies in the world may instead decide to plough through spikes in the cost of electricity, burning billions of dollars as a result.

The largest and most expensive datacentres in the AI sector are those used to train “frontier” AI, systems such as GPT-4o and Claude 3.5 which are more powerful and capable than any other. The leader in the field has changed over the years, but OpenAI is generally near the top, battling for position with Anthropic, maker of Claude, and Google's Gemini.

Already, the “frontier” competition is thought to be “winner takes all”, with very little stopping customers from jumping to the latest leader. That means that if one business spends \$100m on a training run for a new AI system, its competitors have to decide to spend even more themselves or drop out of the race entirely.

Worse, the race for so-called “AGI”, AI systems that are capable of doing anything a person can do, means that it could be worth spending hundreds of billions of dollars on a single training run – if doing so led your company to monopolise a technology that could, as [OpenAI says](#), “elevate humanity”.

Won't AI firms learn to use less electricity?

Every month, there are new breakthroughs in AI technology that enables companies to do more with less. In March 2022, for instance, a DeepMind project called Chinchilla showed researchers how to train frontier AI models using radically less computing power, by changing the ratio between the amount of training data and the size of the resulting model.

But that didn't result in the same AI systems using less electricity; instead, it resulted in the same amount of electricity being used to make even better AI systems. In economics, that phenomenon is known as “Jevons' paradox”, after the economist who noted that the improvement of the steam engine by James Watt, which allowed for much less coal to be used, instead led to a huge increase in the amount of the fossil fuel burned in England. As the price of steam power plummeted following Watt's invention, new uses were discovered that wouldn't have been worthwhile when power was expensive.

