TechScape: Turns out there's another problem with AI – its environmental toll

Publication Date: 2023-08-01

Author: Chris Stokel-Walker

Section: Technology

Tags: Technology, TechScape newsletter, Artificial intelligence (AI), Computing, Climate crisis, Fossil fuels, features

Article URL: https://www.theguardian.com/technology/2023/aug/01/techscape-environment-cost-ai-artificial-intelligence

Technology never exists in a vacuum, and the rise of cryptocurrency in the last two or three years shows that. While plenty of people were making extraordinary amounts of money from investing in bitcoin and its competitors, there was consternation about the impact those get-rich-quick speculators had on the environment. Mining cryptocurrency was environmentally taxing. The core principle behind it was that you had to expend effort to get rich. To mint a bitcoin or another cryptocurrency, you had to first "mine" it. Your computer would be tasked with completing complicated equations that, if successfully done, could create a new entry on to the blockchain. People began working on an industrial scale, snapping up the high-powered computer chips, called GPUs (graphics processing units), that could mine for crypto faster than your off-the-shelf computer components at such pace that Goldman Sachs estimated 169 industries were affected by the 2022 chip shortage. And those computer chips required more electricity to power them; bitcoin mining alone uses more electricity than Norway and Ukraine combined. The environmental cost of the crypto craze is still being tallied – including by the Guardian this April. The AI environmental footprint A booming part of tech – which uses the exact same GPUs as intensely, if not moreso, than crypto mining – has got away with comparatively little scrutiny of its environmental impact. We are, of course, talking about the AI revolution. Generative AI tools are powered by GPUs, which are complex computer chips able to handle the billions of calculations a second required to power the likes of ChatGPT and Google Bard. (Google uses its own similar technology, called tensor processing units, or TPUs.) There should be more conversation about the environmental impact of AI, says Sasha Luccioni, a researcher in ethical and sustainable AI at Hugging Face, which has become the de facto conscience of the AI industry. (Meta recently released its Llama 2 open-source large language model through Hugging Face.) "Fundamentally speaking, if you do want to save the planet with AI, you have to consider also the environmental footprint [of AI first]," she says. "It doesn't make sense to burn a forest and then use AI to track deforestation." Counting the carbon cost Luccioni is one of a number of researchers trying – with difficulty – to quantify AI's environmental impact. It's difficult for a number of reasons, among them that the companies behind the most popular tools, as well as the companies selling the chips that power them, aren't very willing to share details of how much energy their systems use. There's also an intangibility to AI that stymies proper accounting of its environmental footprint. "I think AI is not part of these pledges or initiatives, because people think it's not material, somehow," she says. "You can think of a computer or something that has a physical form, but AI is so ephemeral. Even for companies trying to make efforts, I don't typically see AI on the radar." That ephemerality also exists for end users. We know that we're causing harm to the planet when we turn on our cars because we can see or smell the fumes coming out of the exhaust after we turn the key. With AI, you can't see the cloud-based servers being

queried, or the chips rifling through their memory to complete the processing tasks asked of it. For many, the huge volumes of water coursing through pipes inside data centres, deployed to keep the computers powering the AI tools cool, are invisible. You just type in your query, wait a few seconds, then get a response. Where's the harm in that? Putting numbers to the problem Let's start with the water use. Training GPT-3 used by 3.5m litres of water through datacentre usage, according to one academic study, and that's provided it used more efficient US datacentres. If it was trained on Microsoft's datacentres in Asia, the water usage balloons to closer to 5m litres. Prior to the integration of GPT-4 into ChatGPT, researchers estimated that the generative AI chatbot would use up 500ml of water – a standard-sized water bottle – every 20 questions and corresponding answers. And ChatGPT was only likely to get thirstier with the release of GPT-4, the researchers forecast. Estimating energy use, and the resulting carbon footprint, is trickier. One third-party analysis by researchers estimated that training of GPT-3, a predecessor of ChatGPT, consumed 1,287 MWh, and led to emissions of more than 550 tonnes of carbon dioxide equivalent, similar to flying between New York and San Francisco on a return journey 550 times. Reporting suggests GPT-4 is trained on around 570 times more parameters than GPT-3. That doesn't mean it uses 570 times more energy, of course – things get more efficient – but it does suggest that things are getting more energy intensive, not less. For better or for worse Tech boffins are trying to find ways to maintain AI's intelligence without the huge energy use. But it's difficult. One recent study, published earlier this month, suggests that many of the workarounds already tabled end up trading off performance for environmental good. It leaves the AI sector in an unenviable position. Users are already antsy about what they see as a worsening performance of generative AI tools like ChatGPT (whether that's just down to their perception or based in reality isn't yet certain). Sacrificing performance to reduce ecological impact seems unlikely. But we need to rethink AI's use – and fast. Technology analysts Gartner believe that by 2025, unless a radical rethink takes place in how we develop AI systems to better account for their environmental impact, the energy consumption of AI tools will be greater than that of the entire human workforce. By 2030, machine learning training and data storage could account for 3.5% of all global electricity consumption. Pre-AI revolution, datacentres used up 1% of all the world's electricity demand in any given year. So what should we do? Treating AI more like cryptocurrency – with an increased awareness of its harmful environmental impacts, alongside awe at its seemingly magical powers of deduction – would be a start. If you want to read the complete version of the newsletter please subscribe to receive TechScape in your inbox every Tuesday.