Engineering xxx (xxxx) xxx

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

News & Highlights

Artificial Intelligence, Like Cryptocurrency, Eats Energy—Lots of It

Mitch Leslie

Senior Technology Writer

In the month of August 2023, more than 180 million users logged on to ChatGPT (Fig. 1) [1], the famous artificial intelligence (AI) chatbot [2]. Developed by San Francisco, CA, USA-based OpenAI, ChatGPT can write prose in the style of Ernest Hemingway, generate computer code, help jobseekers craft their resumes, furnish reasonably accurate medical diagnoses, and perform a range of other sophisticated tasks [2–5]. People who sign up for the free version of ChatGPT—as well as anyone else accessing the many AI-enhanced search engines and programs that are now available—do not have to pay for the assistance. But it is not actually free.

Providing the huge amount of computational power required by ChatGPT and the multitude of other AI systems now being used around the world takes a huge amount of electricity and water and results in a lot of pollution [6]. Exactly how much remains "a matter of debate," said Deep Jariwala, an associate professor of electrical and systems engineering at the University of Pennsylvania in Philadelphia, PA, USA. But every estimate so far suggests the amounts are already enormous [7]. "The numbers are daunting in absolute terms," said Benjamin Lee, professor of electrical and systems engineering and Jariwala's colleague at the University of Pennsylvania, PA, USA.

And AI's use is growing exponentially. Google (Mountain View, CA, USA) and Microsoft (Redmond, WA, USA) have created their own rivals to ChatGPT, for instance, that they have integrated into their search engines [8]. AI is becoming indispensable in many sectors of the economy, including medicine, agriculture, pharmaceutical research, and education, and it underlies the function of more and more products and services such as cell phones and online shopping websites like Amazon's (Seattle, WA, USA) [9]. "AI consumes a lot of electricity, and if it becomes more a part of our daily lives, it will consume a lot more," said Jesse Dodge, a research scientist at the Allen Institute for AI in Seattle, WA, USA.

Although the use of AI is increasing exponentially, its energy and resource consumption do not have to, said Shaolei Ren, an associate professor of electrical and computer engineering at the University of California, Riverside, CA, USA. Researchers and companies are working on solutions, including more efficient computer circuits, streamlined software, and improved algorithms, to reduce AI's climate and resource costs. Google, Microsoft, and Amazon and other companies whose cloud computing services run most AI models claim they have taken steps to boost the efficiency and sustainability of their operations [10]. Even allowing customers to choose where and when the models run could make a positive impact [11,12].

The huge environmental footprint of cryptocurrencies—whose well-studied energy consumption and greenhouse gas emissions exceed those of many countries—has drawn plenty of attention and provoked substantial criticism for many years [13,14]. But the first research report raising the alarm about the environmental downsides of AI did not come out until 2019 [15]. AI requires so much energy because what it does is inherently more computationally demanding than just crunching numbers or searching the internet, said Jariwala. AI systems pore over massive amounts of data, hunting for patterns that can allow them to make predictions [16]. AI models used to run on laptops, Jariwala said, but they now need so much processing power that they must rely on large data centers. And these data centers devour energy (Fig. 2) [17,18].

Researchers have estimated the overall energy use of individual cryptocurrencies such as Bitcoin and Ether [13,14], but gauging the power consumption and greenhouse emissions of AI has been more difficult because the necessary data are hard to obtain. AI researchers and companies are scattered around the world, and no central authority tracks their energy consumption, Dodge said. In addition, companies are reluctant to share details about their AI use that might help their competitors.

Still, the estimates researchers have produced are provocative. Before AI models can start writing poems in the style of Shakespeare or searching magnetic resonance imaging (MRI) scans to detect tumors, they undergo training, during which they learn to interpret data. This process is energy intensive. A 2021 study estimated that training GPT-3, the AI model initially used to power ChatGPT, produced the equivalent of 550 tonnes of CO_2 and consumed 1287 MW·h of electricity, roughly the same amount as 120 homes in the United States would use in a year [19,20]. Once AI systems begin operating, they continue to consume electricity and generate CO_2 [21]. A study that had access to data from Google estimated that taken together, AI training and operation account for 10%–15% of the company's annual 15.4 TW·h power usage [20,22]. One of the first studies to assess AI's overall power consumption drew on projected sales by NVIDIA (Santa Clara, CA, USA), the company whose servers dominate the AI market. Using those figures, the study predicted that by 2027, AI could be consuming between 85 TW·h and 134 TW·h of electricity per year, roughly the same amount as the country of Argentina [23,24]

https://doi.org/10.1016/j.eng.2023.11.006
2095-8099/ All rights reserved.

Please cite this article as: M. Leslie, Artificial Intelligence, Like Cryptocurrency, Eats Energy—Lots of It, Engineering, https://doi.org/10.1016/j.eng.2023.11.006

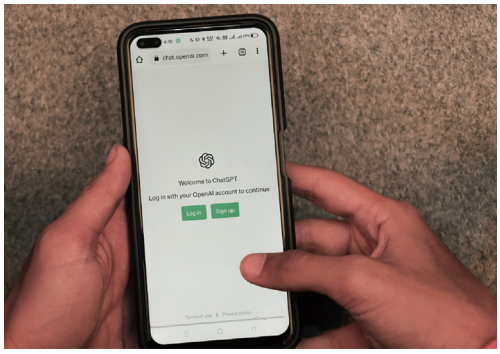

Fig. 1. Millions of people log onto the chatbot ChatGPT every day to access its surprisingly human-like output. The newest version can speak with users, search the internet, and analyze images. But these innovative abilities, and those of similar artificial intelligence (AI) models being employed by Google, Microsoft, Amazon, and other companies, appear to come at a steep environmental cost. Credit: Sanket Mishra (CC0).

Fig. 2. Large, cloud-accessed data centers like this one built and operated by Google in Council Bluffs, IA, USA, can draw more than 100 MW at any given time—consuming roughly as much electricity as 80 000 homes—and have cooling systems that can guzzle up to 3×10^9 L of water a year. Google claims that all its data centers will be running entirely on renewable energy by 2030. Credit: Chad Davis (CC BY 2.0).

The environmental toll of AI goes beyond energy. Building data centers and manufacturing computer chips and servers produces greenhouse gases, for instance, and obtaining the raw materials for the electronics entails environmentally damaging mining [25]. Using data from Facebook (Menlo Park, CA, USA), Lee and colleagues estimated in a 2022 report that this so-called embodied carbon (all greenhouse gas emissions not directly connected to data center operation) accounts for between 30% and 70% of the total carbon footprint of certain AI models [26]. And as one article cheekily commented, AI has "a drinking problem," with data centers requiring large amounts of water to keep cool [27,28]. Ren said that he and his team were the first to try to estimate AI's water use. Starting with Microsoft's reported annual water consumption, they estimated that GPT-3—which trained on Microsoft's cloud computing platform—needed about half a liter of water every time it answered from 20 to 50 user questions [29]. ChatGPT Plus, OpenAI's updated, faster, and more versatile chatbot that costs 20 USD per month, now relies on an even larger AI model, GPT-4, which is likely to be thirstier. So "overall, our estimate is conservative," Ren said. Given the 180 million-plus monthly visitors to the ChatGPT website [1], its water consumption could quickly add up to a substantial amount.

AI's expanding energy demand is part of a larger problem—computation of all types is gobbling increasing amounts of power and needs to dramatically improve efficiency, said Jariwal. A 2020 report from the Semiconductor Research Company (Durham, NC, USA), the industry's main research organization, projects that if current energy use trends continue, by 2040 computation alone will require more electricity than the entire world will be able to produce [30].

Improving computer hardware and software is one way to help reduce the energy use of computation in general and of AI specifically, said Jariwal. Using smaller AI models that can be trained on less data is another strategy that is gaining traction [31]. Optimizing where and when data centers train and run AI models may also make a difference. In a 2022 paper, for instance, Dodge and colleagues estimated carbon emissions for training different AI models on Microsoft's cloud computing service Azure. The amount of CO_2 produced varies with the location within the United States and around the world, their study showed [12]. Training an AI model in data centers in the central United States, a region more reliant on fossil fuels, generates about twice as much CO_2 on average as training it in facilities in the western United States, which draw on plentiful hydropower. The time of day also matters, the researchers found. For one model, delaying training until more renewable energy was online cut CO_2 emissions by up to 80%.

Users who are going to train their AI models on cloud services like Azure can already choose where and when they run, Dodge said, and they can pick times and places that will produce the least amount of CO_2. "That is a really impactful decision people can make," he said. Ren and his team have shown that a similar time and geography calculation can also cut the water consumption of AI [32].

Another counterintuitive way to reduce AI's environmental impact is to use it less often, said Dodge. Although companies are rushing to apply AI, it is not ready for some uses, particularly if they involve human safety. For example, an AI model that can detect tumors on MRI scans may be a useful tool for radiologists, but it is not ready to replace radiologists, Dodge said, because such models still make many mistakes. Companies and researchers should think carefully about their rationale for choosing to use AI, he said. "Is AI the right tool here? I hope people will ask that more often."

References

[1] Tong A. Exclusive: ChatGPT traffic slips again for third month in a row [Internet]. London: Reuters; 2023 Sep 7 [cited 2023 Sep 30. Available from: https://www.reuters.com/technology/chatgpt-traffic-slips-again-third-month-row-2023-09-07/.

[2] Mackenzie D. Surprising advances in generative Artificial Intelligence prompt amazement—and worries. Engineering 2023;25:9–11.

[3] Bilton N. ChatGPT made me question what it means to be a creative human [Internet]. New York City: Vanity Fair; 2022 Dec 9 [cited 2023 Sep 30]. Available from: https://www.vanityfair.com/news/2022/12/chatgpt-question-creative-human-robotos.

[4] Gunnell M. 10 awesomely practical tasks you can do with ChatGPT [Internet]. San Francisco: PC World; 2023 Apr 10 [cited 2023 Sep 30]. Available from: https://www.pcworld.com/article/1681251/10-awesome-things-you-can-do-with-chatgpt.html.

[5] Leonard A. 'Dr. Google' meets its match in Dr. ChatGPT [Internet]. Washington, DC: NPR; 2023 Sep 16 [cited 2023 Sep 30]. Available from: https://www.npr.org/sections/health-shots/2023/09/16/1199924303/chatgpt-ai-medical-advice.

[6] Stokel-Walker C. The generative AI race has a dirty secret [Internet]. London: Wired UK; 2023 Feb 10 [cited 2023 Sep 30]. Available from: https://www.wired.co.uk/article/the-generative-ai-search-race-has-a-dirty-secret.

[7] Luccioni S. The mounting human and environmental costs of AI [Internet]. New York City: Ars Technica; 2023 Apr 12 [cited 2023 Sep 30]. Available from: https://arstechnica.com/gadgets/2023/04/generative-ai-is-cool-but-lets-not-forget-its-human-and-environmental-costs/.

[8] Bohannon M. Google adding AI to search engine for some users—closing in on microsoft's AI Push [Internet]. New York City: Forbes; 2023 May 10 [cited 2023 Sep 30]. Available from: https://www.forbes.com/sites/mollybohannon/2023/05/10/google-adding-ai-to-search-engine-for-some-users-closing-in-on-microsofts-ai-push/.

[9] Wong M. The Internet's next great power suck [Internet]. New York City: The Atlantic; 2023 Aug 23 [cited 2023 Sep 30]. Available from: https://www.theatlantic.com/technology/archive/2023/08/ai-carbon-emissions-data-centers/675094/.
[10] Petrenko CM. AI complicates sustainability profiles of tech companies—ESG insight [Internet]. New York City: Wall Street Journal; 2023 Sep 21 [cited 2023 Sep 30]. Available from: https://www.wsj.com/tech/ai-complicates-sustainability-profiles-of-tech-companies-esg-insight-681e0441.
[11] Xu T. These simple changes can make AI research much more efficient [Internet]. Cambridge: MIT Tech Review; 2022 Jul 6 [cited 2023 Sep 30]. Available from: https://www.technologyreview.com/2022/07/06/1055458/ai-research-emissions-energy-efficient/.
[12] Dodge J, Prewitt T, Tachet des Combes R, Odmark E, Schwartz R, Strubell E, et al. Measuring the carbon intensity of AI in cloud instances. 2022. arXiv: 2206.05229.
[13] Leslie M. Will cryptocurrencies break the energy bank? Engineering 2020;6(5):489–90.
[14] Bitcoin energy consumption index [Internet]. Amsterdam: Digiconomist; 2023 Sep 30 [cited 2023 Sep 30]. Available from: https://digiconomist.net/bitcoin-energy-consumption.
[15] Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP. 2019. arXiv: 1906.02243.
[16] What is AI? [Internet]. New York City: McKinsey & Company; 2023 Apr 24 [cited 2023 Sep 30]. Available from: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai.
[17] Loten A. Rising data center costs linked to AI [Internet]. New York City: Wall Street Journal; 2023 Jul 13 [cited 2023 Sep 30]. Available from: https://www.wsj.com/articles/rising-data-center-costs-linked-to-ai-demands-fc6adc0e.
[18] Kumar A, Davenport T. How to make generative AI greener [Internet]. Cambridge: Harvard Business Review; 2023 Jul 20 [cited 2023 Sep 30]. Available from: https://hbr.org/2023/07/how-to-make-generative-ai-greener.
[19] Patterson D, Gonzalez J, Le Q, Liang C, Munguia L, Rothchild D, So D, et al. Carbon emissions and large neural network training. 2021. arXiv: 2104:10350.
[20] Saul J, Bass D. Artificial Intelligence is booming—so is its carbon footprint [Internet]. New York City: Bloomberg News; 2023 Mar 9 [cited 2023 Sep 30]. Available from: https://www.bloomberg.com/news/articles/2023-03-09/how-much-energy-do-ai-and-chatgpt-use-no-one-knows-for-sure#xj4y7vzkg.
[21] Desislavov R, Martínez-Plumed F, Hernández-Orallo J. Trends in AI inference energy consumption: beyond the performance-vs-parameter laws of deep learning. Sustain Comput Infor 2023;38:100857.
[22] Patterson D, Gonzalez J, Hölzle U, Le Q, Liang C, Munguia L, et al. The carbon footprint of machine learning training will plateau, then shrink. 2022. arXiv:2204:05149.
[23] Paul A. AI could consume as much energy as argentina annually by 2027 [Internet]. New York City: Popular Science; 2023 Oct 12 [cited 2023 Oct 16]. Available from: https://www.popsci.com/technology/ai-energy-use-study/.
[24] de Vries A. The growing energy footprint of Artificial Intelligence. Joule 2023;7(10):2191–4.
[25] Leslie M. Will massive appetite for minerals stall clean energy transition? Engineering 2023: In press.
[26] Wu CJ, Raghavendra R, Gupta U, Acun B, Ardalana N, et al. Sustainable AI: environmental implications, challenges and opportunities. 2022. arXiv: 2111:00364.
[27] Gandhi R. AI chatbots may be fun, but they have a drinking problem [Internet]. Mumbai: Times of India; 2023 May 31 [cited 2023 Sep 30]. Available from: https://timesofindia.indiatimes.com/india/ai-chatbots-may-be-fun-but-they-have-a-drinking-problem/articleshow/100634953.cms?from=mdr.
[28] O'Brien. Fingerhut H. Artificial intelligence technology behind ChatGPT was built in Iowa, with a lot of water [Internet]. New York City: Associated Press; 2023 Sep 9 [cited 2023 Sep 30]. Available from: https://apnews.com/article/chatgpt-gpt4-iowa-ai-water-consumption-microsoft-f551fde98083d17a7e8d904f8be822c4.
[29] Li P, Yang J, Islam MA, Ren S. Making AI less "thirsty": uncovering and addressing the secret water footprint of AI models. 2022. arXiv: 2304:03271.
[30] Decadal Plan for Semiconductors. Abridged report [Internet]. Durham: Semiconductor Research Corporation; 2020 Oct [cited 2023 Sep 30]. Available from: https://www.src.org/about/decadal-plan/.
[31] The bigger-is-better approach to AI is running out of road [Internet]. London: Economist; 2023 Jun 21 [cited 2023 Sep 30]. Available from: https://www.economist.com/science-and-technology/2023/06/21/the-bigger-is-better-approach-to-ai-is-running-out-of-road.
[32] Li P, Yang J, Werman A, Ren S. Toward environmentally equitable AI via geographical load balancing. 2023. arXiv: 2307:05494.