Al Is Eating Data Center Power Demand—and It's Only Getting Worse

A new analysis of Al hardware being produced and how it is being used attempts to estimate the vast amount of electricity being consumed by Al.



SAVE

AI'S ENERGY USE already represents as much as 20 percent of global datacenter power demand, research published Thursday in the journal Joule shows. That demand from AI, the research states, could double by the end of this year, comprising nearly half of all total data-center electricity consumption worldwide, excluding the electricity used for bitcoin mining.

The new research is published in a commentary by Alex de Vries-Gao, the founder of Digiconomist, a research company that evaluates the environmental impact of technology. De Vries-Gao started Digiconomist in the late 2010s to explore the impact of bitcoin mining, another extremely energy-intensive activity, would have on the environment. Looking at AI, he says, has grown more urgent over the past few years because of the widespread adoption of ChatGPT and other large language models that use massive amounts of energy. According to his research, worldwide AI energy demand is now set to surpass demand from bitcoin mining by the end of this year.

ADVERTISEMENT

Science Newsletter

Your weekly roundup of the best stories on health care, the climate crisis, new scientific discoveries, and more. Delivered on Wednesdays.

SIGN UP

By signing up, you agree to our <u>user agreement</u> (including <u>class action waiver and arbitration provisions</u>), and acknowledge our <u>privacy policy</u>.

"The money that bitcoin miners had to get to where they are today is peanuts compared to the money that <u>Google</u> and <u>Microsoft</u> and all these big tech companies are pouring in [to AI]," he says. "This is just escalating a lot faster, and it's a much bigger threat."

The development of AI is already having an impact on Big Tech's climate goals. Tech giants have <u>acknowledged in recent sustainability reports</u> that AI is largely responsible for driving up their energy use. Google's greenhouse gas emissions, for instance, have increased 48 percent since 2019, complicating the company's goals of reaching net zero by 2030.

"As we further integrate AI into our products, reducing emissions may be challenging due to increasing energy demands from the greater intensity of AI compute," Google's 2024 sustainability report <u>reads</u>.

Last month, the International Energy Agency released a <u>report</u> finding that data centers made up 1.5 percent of global energy use in 2024—around 415 terrawatt-hours, a little less than the yearly energy demand of Saudi Arabia. This number is only set to get bigger: Data centers' electricity consumption has grown four times faster than overall consumption in recent years, while the amount of investment

in data centers has nearly doubled since 2022, driven largely by massive expansions to account for new AI capacity. Overall, the IEA predicted that data center electricity consumption will grow to more than 900 TWh by the end of the decade.

But there's still a lot of unknowns about the share that AI, specifically, takes up in that current configuration of electricity use by data centers. Data centers power a variety of services—like hosting cloud services and providing online infrastructure—that aren't necessarily linked to the energy-intensive activities of AI. Tech companies, meanwhile, largely keep the energy expenditure of their software and hardware private.

Some attempts to quantify AI's energy consumption have started from the user side: calculating the amount of electricity that goes into a single ChatGPT search, for instance. De Vries-Gao decided to look, instead, at the supply chain, starting from the production side to get a more global picture.

The high computing demands of AI, De Vries-Gao says, creates a natural "bottleneck" in the current global supply chain around AI hardware, particularly around the Taiwan Semiconductor Manufacturing Company (TSMC), the undisputed leader in producing key hardware that can handle these needs. Companies like Nvidia outsource the production of their chips to TSMC, which also produces chips for other companies like Google and AMD. (Both TSMC and Nvidia declined to comment for this article.)

De Vries-Gao used analyst estimates, earnings call transcripts, and device details to put together an approximate estimate of TSMC's production capacity. He then looked at publicly available electricity consumption profiles of AI hardware and estimates on utilization rates of that hardware—which can vary based on what it's being used for—to arrive at a rough figure of just how much of global data-center demand is taken up by AI. De Vries-Gao calculates that without increased production, AI will consume up to 82 terrawatt-hours of electricity this year—roughly around the same as the annual electricity consumption of a country like

Switzerland. If production capacity for AI hardware doubles this year, as analysts have projected it will, demand could increase at a similar rate, representing almost half of all data center demand by the end of the year.

Despite the amount of publicly available information used in the paper, a lot of what De Vries-Gao is doing is peering into a black box: We simply don't know certain factors that affect AI's energy consumption, like the utilization rates of every piece of AI hardware in the world or what machine learning activities they're being used for, let alone how the industry might develop in the future.

Sasha Luccioni, an AI and energy researcher and the climate lead at open-source machine-learning platform Hugging Face, cautioned about leaning too hard on some of the conclusions of the new paper, given the amount of unknowns at play. Luccioni, who was not involved in this research, <u>says</u> that when it comes to truly calculating AI's energy use, disclosure from tech giants is crucial.

"It's because we don't have the information that [researchers] have to do this," she says. "That's why the error bar is so huge."

And tech companies *do* keep this information. In 2022, Google published a paper on machine learning and electricity use, <u>noting</u> that machine learning was "10%–15% of Google's total energy use" from 2019 to 2021, and predicted that with best practices, "by 2030 total carbon emissions from training will reduce." However, since that paper—which was released before Google Gemini's debut in 2023—Google has not provided any more detailed information about how much electricity ML uses. (Google declined to comment for this story.)

"You really have to deep-dive into the semiconductor supply chain to be able to make any sensible statement about the energy demand of AI," De Vries-Gao says. "If these big tech companies were just publishing the same information that Google was publishing three years ago, we would have a pretty good indicator" of AI's energy use.

You Might Also Like ...

- In your inbox: Upgrade your life with WIRED-tested gear
- "Wi-Fi keeps going down": Trump's RTO mandate is going terribly
- **Big Story**: The worm that <u>no computer scientist can crack</u>
- Yuval Noah Harari: "Prepare to share the planet with AI superintelligence"
- Uncanny Valley: An insider look at the influence of Silicon Valley

Molly Taft is a senior writer for WIRED, covering climate change, energy, and the environment. Previously, they were a reporter and editor at Drilled, an investigative climate multimedia reporting project. Before that, they wrote about climate change and technology for Gizmodo, and served as a contributing editor for the New ... Read more

SENIOR WRITER, CLIMA	SI	ENI	0 R	WRI	TER.	CL	IMA	\ T E
----------------------	----	-----	-----	-----	------	----	-----	-------

TOPICS SCIENCE

ENVIRONMENT

ARTIFICIAL INTELLIGENCE

ENERGY

ELECTRICITY

Science Newsletter

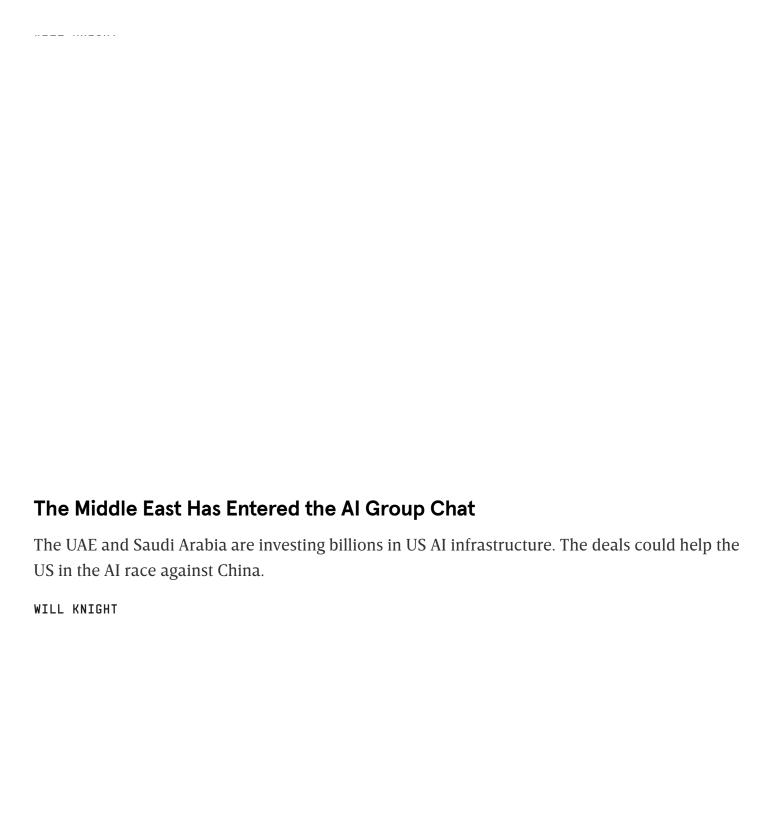
Your weekly roundup of the best stories on health care, the climate crisis, new scientific discoveries, and more. Delivered on Wednesdays.

SIGN UP	

By signing up, you agree to our <u>user agreement</u> (including <u>class action waiver and arbitration provisions</u>), and acknowledge our <u>privacy policy</u>.

READ MORE
These Stantume And Duilding Advanced Al Madele Without Date Conten
These Startups Are Building Advanced Al Models Without Data Centers
A new crowd-trained way to develop LLMs over the internet could shake up the AI industry with a giant 100 billion-parameter model later this year.
WILL KNIGHT





States and Startups Are Suing the US Nuclear Regulatory Commission

Critics of the NRC say its red tape and lengthy authorization timelines stifle innovation, but handing some of its responsibilities to states could undermine public trust and the industry's safety record.

MOLLY TAFT

Singapore's Vision for Al Safety Bridges the US-China Divide

In a rare moment of global consensus, AI researchers from the US, Europe, and Asia came together in Singapore to form a plan for researching AI risks.

WILL KNIGHT



DOGE tested and used Meta's Llama 2 model to review and classity responses from federal workers to the infamous "Fork in the Road" email.

MAKENA KELLY

WIRED

One year for \$2.50 \$1 per month

SUBSCRIBE

✓ YOUR PRIVACY CHOICES