



Paul Deaton <paul.deaton@gmail.com>

How much electricity does AI consume? [2025 summary]

1 message

Hannah Ritchie from By the Numbers <hannahritchie@substack.com>

Tue, May 5, 2026 at 12:55 AM

Reply-To: Hannah Ritchie from By the Numbers

<reply+38ztt&21p5m&&16c9cf26184405c30c74109b554f58e1e014a2613fa6a2cba32d4ba4ae3d7bdf@mg1.substack.com>

To: paul.deaton@gmail.com

Forwarded this email? [Subscribe here](#) for more

How much electricity does AI consume? [2025 summary]

What share of electricity is consumed by data centres? What's the energy footprint of ChatGPT and other chatbots?

HANNAH RITCHIE

MAY 5



READ IN APP ↗

Over the past year, I've written [a few articles](#) about the energy (and carbon) footprint of individual use of artificial intelligence, mostly in the form of asking LLMs (or chatbots) questions.

The problem is that until recently, the latest figures from organisations such as the International Energy Agency were for 2024. It's hard to have a serious conversation about AI's energy use with data from before AI really took off.

But a few weeks ago, the IEA published [its latest report](#) on Energy and AI, with estimates for 2025, and updates of its future projections.¹

I want to run through the high-level numbers here.

I'll make one caveat here so I don't have to repeat myself throughout: things become increasingly uncertain once we get into medium-term

projections. I am not here to present these projections as ground truth — other analysts produce outlooks that are quite different — but to reflect what the latest IEA report says in an understandable way.

How much of global electricity is used for data centres and AI?

Last year, around **1.5% of the world's electricity** was used to power data centres.

Now data centres are not just AI: they're facilities that contain the servers and IT infrastructure behind all of our digital services. That's everything from Substack and the rest of the internet, to Netflix streaming, Google Maps, online banking, and messaging friends.

AI data centres are dedicated facilities for running AI models. The distinction between the two is not always that clean-cut, but the chart below shows the breakdown.

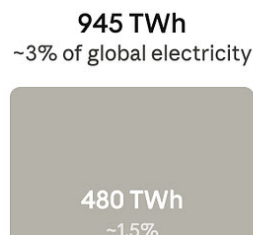
Non-AI data centres consumed more than twice as much electricity as AI-focused ones.

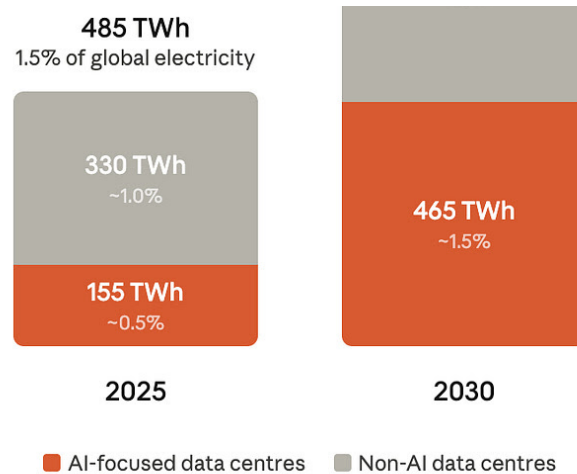
By my estimates, **AI consumed around 0.5% of the world's electricity in 2025**.

Electricity is just a subset of total energy use, so data centres consume less than 0.5% of final energy, and AI less than 0.2%. That's useful to keep in mind: the AI debate is an electricity one, not an energy one.

How much of global electricity is used for data centres?

Measured in terawatt-hours (TWh). Estimates for 2025; IEA projections for 2030.





Note: Projections for 2030 are based on the IEA's base case, and come with significant uncertainty.
Data source: Calculated based on IEA, Key questions on Energy and AI (2026) and Ember (2026).

I've also included the IEA's base-case projection of data centre demand in 2030 (I'll look at this in more detail later).

Most of the *growth* in data centre demand will come from AI-focused facilities. In the IEA's base scenario, data centres grow to 3% of global electricity in 2030. AI centres then use about the same amount as non-AI ones.

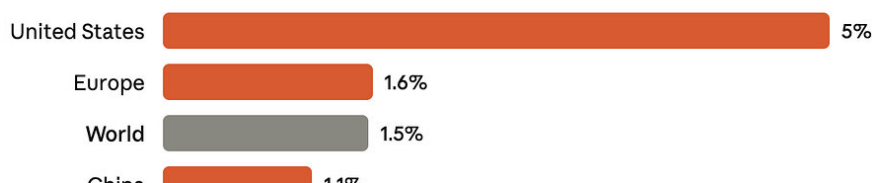
Data centre demand is very much US-centric

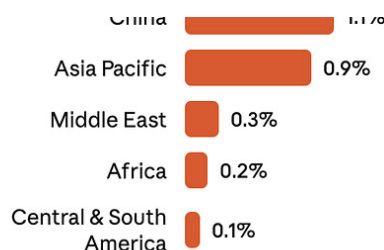
1.5% of the world's electricity doesn't seem like much. But in some places, that share is far higher.

In the chart, you can see the share of electricity used for data centres in different regions. 5% of the US's electricity is used for data centres. For AI-focused data centres specifically, we're probably talking around 2%.

Share of electricity used by data centres, 2025

Data centre electricity consumption as a percentage of total electricity generation.





Note: Regional electricity totals are based on Ember (2026). Regional boundaries may not align exactly with IEA definitions.

Data source: Calculated based on IEA, Key questions on Energy and AI (2026) and Ember (2026).

But in fact, this demand is even more locally concentrated. Beneath Europe's 1.6% figure, we have Ireland, where data centres account for [more than 20%](#) of its electricity consumption. Beneath the 5% US figure, there are a number of states where data centres make up [more than 10%](#) of demand, and in states such as Virginia, it's more than a quarter.²

This is really the key challenge with growing data centre demand: it's geographically concentrated, meaning the entire world's demand is being served by a small number of electricity grids. You'd also see bottlenecks and local grid issues if you were trying to serve all of the world's demand for microwaves, kettles, or washing machines from a handful of places.

How have the IEA's projections changed?

What's the IEA now saying about *future* demand?

It publishes four scenarios: its base case, a lift-off case where AI grows much faster, and two cases with lower power demand, either due to even more rapid efficiency improvements or lower-than-expected demand.

You can see these four scenarios in the chart below. Projections to 2030 are already pretty uncertain, but the IEA stresses that those out to 2035 are even more so (what it calls "explorations").

There aren't huge differences between the four scenarios in 2030. In all scenarios, global power demand for data centres basically doubles from 2025 levels.

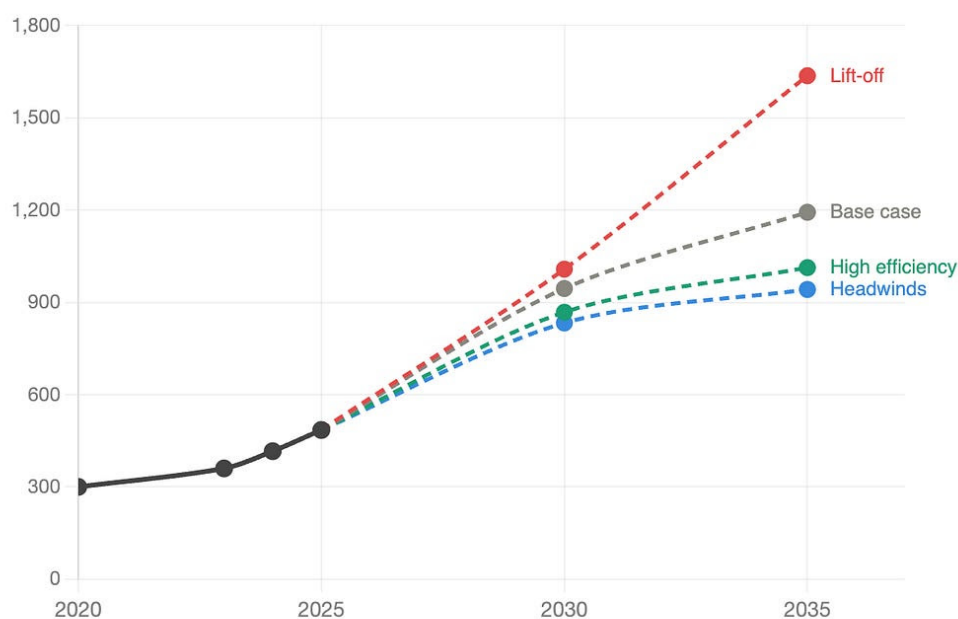
In the base case, global electricity demand reaches 945 TWh. The high and

low scenarios are within around 100 TWh of this. The reason these scenarios vary so little is that much of this supply is already “locked in”. It takes several years to build a data centre, there are near-term bottlenecks in supply chains and chips, and there are *local* bottlenecks in how quickly new power supplies can be built.

After 2030, the IEA’s projections diverge as some of these nearer-term barriers become unlocked.

IEA scenarios of electricity demand from data centres

Global data centre electricity consumption (TWh). Four IEA scenarios for 2030 and 2035.



Note: 2035 figures are exploratory given the high level of uncertainty around data centre demand growth.

Data source: IEA, Key questions on Energy and AI (2026).

Data centres could be a modest driver of electricity growth globally, but a large one in the US

Data centres are not the only driver of new electricity demand. Countries will need more electricity for transport, heating, and industry. Low- and middle-income countries are still seeing rising electricity demand as people move out of poverty and raise their standards of living.

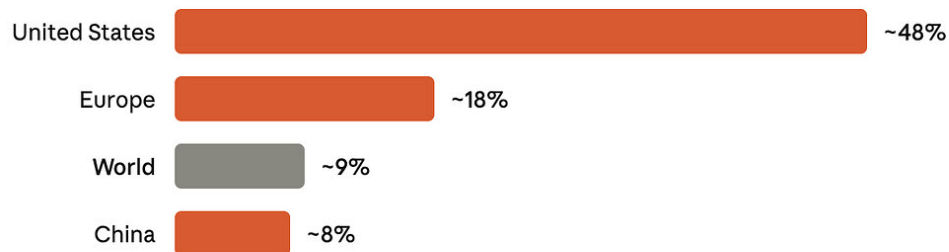
How much of the power demand growth in the next five years could come from data centres?

The IEA projects around one-tenth (9%) of it.

In the US, it could be as much as half. That's for two reasons: most data centres are being built in the US, so most of the growth in *global* demand is really being centralised there; and growth in electricity demand for other uses is slower than elsewhere. The US is [not seeing](#) the electric vehicle boom that other countries are; almost everyone has air conditioning, and it's not a country adding huge amounts of new industrial capacity.

Data centres' share of electricity demand growth to 2030

How much of total electricity demand growth from 2025 to 2030 comes from data centres?



Note: Based on IEA Base Case for data centre growth. Regional electricity growth estimated from recent trends.

Data source: Calculated based on IEA, Key questions on Energy and AI (2026) and Ember (2026).

What's the energy footprint of individual queries?


I've written [several articles](#) on the footprint of individual LLM queries.

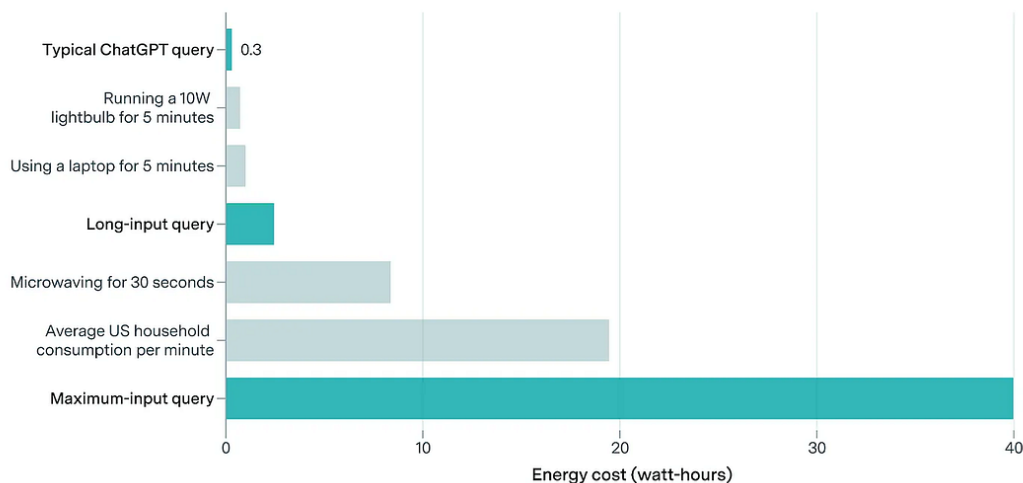
A key takeaway from the numbers was that asking a chatbot a question — which is what most people were using AI for in their day-to-day lives — consumes very little energy.

Tech companies have not been very transparent about the energy use of their AI models (and I think they should be), but the numbers seemed to converge around 0.3 watt-hours (Wh) per typical text query. To put this into context, asking ChatGPT or Gemini 10 simple questions [is equivalent](#) to about 10

seconds of microwaving or mere seconds of showering.

This power consumption varies by the size of the text query, as this chart from Epoch AI shows.

Energy consumption per ChatGPT query is small compared to everyday electricity use 



Pessimistic estimates of the energy usage of ChatGPT with GPT-4o across for different query lengths: typical (<100 words), long (~7,500 words), and maximum context length (~75,000 words), with an average response length of 400 words.

CC-BY

epoch.ai

What does the IEA say about this in its latest report?

It quotes two things. First, the company-reported figures I discussed previously: 0.24 Wh per median text query using Google Gemini ([published by Google](#)) and around 0.34 Wh per average ChatGPT query (mentioned by Sam Altman).³

Second, AI Energy Score benchmarks for different levels of tasks. These figures only measured the energy used by the GPUs (basically, the chips); on top of this, energy is needed for networks, cooling, lighting etc. In hyperscale data centres, GPUs can consume 50% to 60% of a facility's total energy. So the total energy use per query could be as much as double this.

Both these measures are included in the chart below.

The AI Energy Score benchmarks are in decent agreement with the company-reported figures. A medium text query uses very little energy: 0.05

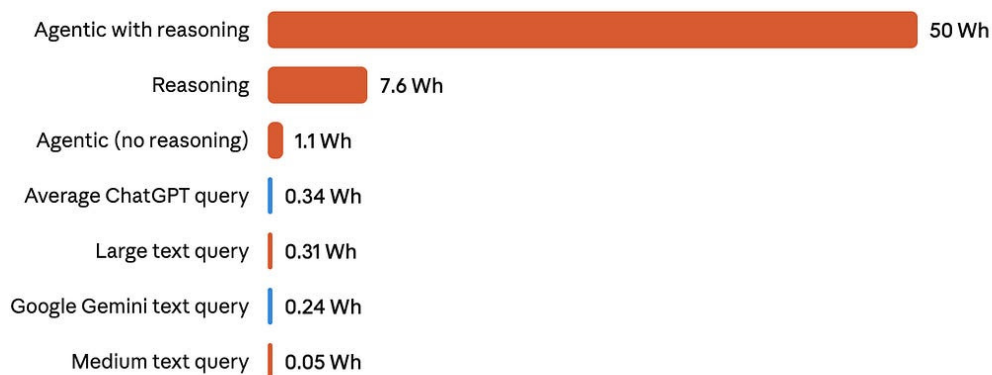
Wh for GPUs, which could be around 0.1 Wh in total. A large text query consumes around 0.3 Wh on GPUs (up to 0.6 Wh in total).

Depending on the length of the text query, asking an LLM a question is somewhere in this 0.1 to 0.6 Wh range. Company-reported figures and benchmark scores are all in this ballpark.

Electricity consumption per AI query varies a lot by task

Marginal electricity consumption of individual LLM queries, measured in watt-hours (Wh).

■ AI Energy Score benchmarks (GPU only) ■ Company-reported figures



Note: AI Energy Score values are GPU-electricity only; agentic estimates assume 4–6 sequential LLM calls. Company-reported figures likely reflect full-system electricity, not GPU only, so the two sources are not directly comparable.

Data source: IEA, Key questions on Energy and AI (2026), based on Hugging Face (2025); Google (2025); OpenAI (2025).

As we'd expect, more complex tasks use more energy. Once we get into the realm of agentic, reasoning, or agentic-with-reasoning tasks, we enter the range of several to tens of watt-hours.

Agents are AI tools — Claude Code is one example — which can plan, iterate and call on tools autonomously. In the IEA methodology, this involves 4 to 6 sequential calls and responses, so you can see how this is inherently more complex than a simple text query.

What do these numbers mean for individual footprints?

Many people are using AI for medium- to long-form text queries, such as asking a quick question or requesting a short fact-check or correction. Their energy use is very small, even if they're asking tens or hundreds of questions a day. A hundred questions have a footprint of around 30 Wh. That's roughly the amount of electricity the average American consumes in just over a

minute (or for the average European, every two and a half minutes).⁴

The footprint of someone who uses agents heavily is not so negligible.

Let's say they do 4 agentic queries per hour (how many you can do in an hour is limited by the fact that complex tasks can take 15 minutes or more to complete). And they do this for 6 hours a day. That's 24 per day. We'll assume that the total electricity use per query is actually 100 Wh (50 Wh multiplied by two).

They'll consume 2,400 Wh (or 2.4 kWh). That's like running a tumble dryer for one cycle, or driving an electric car eight miles. It's around 7% of the average American's electricity use (but a much smaller share of total energy use).

It's not blowing up their footprint, but it's not nothing either.

One follow-up question is the energy use of image and video generation. Unfortunately, I don't know, and this report doesn't include numbers on this.

But it leads to one of the most interesting — and unanswered — points in the report.

The “missing” energy puzzle

The IEA report includes some useful napkin math.

Let's assume a text query to a chatbot consumes 1 Wh of electricity (this would be a fairly large text query, but it's a nice round number).

If the world made 10 billion queries a day — similar to the number of Google searches, and four times the number of queries that ChatGPT reports — then we'd be consuming 10 GWh of power a day, or 3.65 TWh per year.⁵

But AI-focused data centres consumed 155 TWh in 2025. So, all of the world's text queries accounted for around 2% of this electricity consumption. Where did the other 98% go?

That's the puzzle, and we don't know because companies don't release

breakdowns of exactly where their compute and power are going.

Some of it will be for training the models. But even some of the largest models consume less than a terawatt-hour during training (or at least that's what's reported).⁶

Another explanation is image and video generation. Video generation in particular is more energy-intensive per task, but without good data on per-video energy costs or volumes, it's hard to know how much they account for.

Perhaps the most plausible explanation is that the largest energy consumer is the more diffuse deployment of AI across the internet: AI-generated summaries in Google Search results, AI-driven algorithms on social media, content moderation, advertising, and translation.

Enterprise and industrial use of AI is also increasing, ranging from AI models in Microsoft Office, Teams, Gemini, and Salesforce to AI in drug discovery, weather forecasting, and financial modelling.

The most visible form of AI use is asking a chatbot a question. But in reality, many are using it all the time, in subtle (and even involuntary) ways. The energy consumption of these interactions is likely far higher than that of the average person's chatbot requests.

To be clear, this isn't about whether the total AI figure is right; the IEA's estimate is not adding up queries, but from a bottom-up model based on server shipments and their power draw. The puzzle is that we know how much electricity AI infrastructure uses overall, but companies don't disclose what they're actually being used for.

✓ Subscribed

1 IEA (2026), Key Questions on Energy and AI, IEA, Paris <https://www.iea.org/reports/key-questions-on-energy-and-ai>, Licence: CC BY 4.0

-
- 2 In 2023, 25% of its electricity consumption was for data centres. Some projections suggest that this has increased to as much as 40% (although these more recent estimates are less certain).

<https://qz.com/us-states-data-centers-electricity-use-share-ai-tech-1851708857>

<https://jlarc.virginia.gov/landing-2024-data-centers-in-virginia.asp>

-
-
- 3 I never put a lot of onus on the Altman announcement, because there was no report or methodology backing it up. It was really just a single quote in a broader post.

However, Epoch AI also did some independent analysis and landed on a figure of around 0.3 Wh, giving more support to this.

<https://blog.samaltman.com/the-gentle-singularity>

<https://epoch.ai/gradient-updates/how-much-energy-does-chatgpt-use>


-
-
-
- 4 This is calculated based on data from [Ember](#), which you can find here: https://ourworldindata.org/grapher/per-capita-electricity-generation?tab=line&country=USA~OWID_EU27&mapSelect=~USA

It includes total electricity generation per person, including industrial and commercial uses.

-
-
-
-
- 5 $1 * 10 \text{ billion} = 10 \text{ billion Wh (or 10 GWh) per day.}$

Multiplied by 365 days is 3650 GWh (or 3.65 TWh).

-
-
-
-
-
- 6 Grok 4, for example, is reported to have consumed 0.31 TWh.


 SHARE

 LIKE

 COMMENT

 RESTACK

© 2026 Hannah Ritchie
[Unsubscribe](#)

 **Start writing**