

Our Frankenstein Problem

By John Benson

August 2026

1. Introduction

Artificial intelligence (AI) is both wonderful and horrible, but above all else, it is hungry, very hungry. But don't worry, it won't eat, you, me, our friends or our pets, because it's a picky eater. Although the data centers and powerful computers that power AI consume many resources (think electronic components, building materials and land) most of all it needs huge amounts of electricity.

AI electricity consumption could account for up to 12 percent of all U.S. Power by 2028. Generative AI queries consumed 15 terawatt-hours in 2025 and are expected 347 TWh by 2030.

However, there are two pieces of good news to go with this:

- The electronics that power AI become more energy-efficient every year.
- All of the support facilities within data centers also will become more efficient over time.

But there is one piece of really bad news:

- As AI becomes smarter, it will become integral to more of our automation and many other processes. I don't believe it will (completely) displace humans anytime soon because of our silly and unpredictable creative processes, but if some supercomputer designer comes up with a "silliness coprocessor," watch out!

2. AI's Energy Use vs. Climate Change

For one of the world's largest industries, computing has historically been a fairly small climate polluter. As recently as 2020, data centers and data transmission networks, which underpin both the Internet and high-performance computing, produced just 0.6% of humanity's climate-warming greenhouse gas emissions.¹

But if we build enough data centers fast enough, the emissions from computing could easily take off. And generative AI is driving just this kind of swift buildout. In the U.S., where AI has grown fastest, it has changed the whole trajectory of the electric grid.

2.1. Rapid Data Center Growth vs. Electric Transition

"The data center demand used to be about one to two percent of the United States' electricity consumption," says Vijay Gadepally, senior scientist at the MIT Lincoln Laboratory Supercomputing Center. "As of today, we're somewhere around four or five percent."

¹ IEEE Spectrum, April 2025 Edition, Editor's Note: "AI Is Insatiable, and it's got the munchies for memory chips," Harry Goldstein, <https://ieeexplore.ieee.org/document/11476837> Note that you will probably need to be an IEEE member (as your author is) to access this.

This rapid growth is a big change for the country. For the past 20 years, U.S. energy use has been mostly flat. This has made it easier to gradually replace old, climate-polluting energy sources, like coal-fired power plants, with cheaper, cleaner ones—and helped our greenhouse gas emissions slowly decline since the mid-2000s.

Now our energy needs are growing again. A U.S. Department of Energy study estimates that our data centers' energy use, already at a record high, will at least double and may more than triple by 2028. Similar growth is expected in China and Europe.

It will be very hard to meet this demand with clean energy, says Gadepally. "You can't take ten years to build a data center. It has to be done in a year, year and a half, just because of the economics behind it. And the only power sources that can generally scale that fast are non-renewable sources."

Clean wind and solar power are also a poor fit for data centers, which need "firm" energy that runs around the clock. If developers find they have to add firm energy very quickly, they will likely turn to climate-warming natural gas—as is already happening in data center hotspots like Texas and Virginia.

Author's comment: I (sort of) disagree with the above two paragraphs. Photovoltaic (PV) panels plus battery energy storage systems (BESS) are built from standard components, and there is no shortage of either PV panels nor lithium-ion batteries. Also, these are both ramping up production very rapidly:

Over the past three years, U.S. lithium-ion battery production has grown significantly, driven by electric vehicle (EV) demand, renewable energy storage, and federal policy support.²

Production levels in 2023–2024 were in the low gigawatt-hours (GWh) per year range for the U.S., though exact annual figures are not publicly detailed in the provided sources.

In 2025, the U.S. is projected to see continued growth in both capacity and output, with planned capacity expansions adding hundreds of gigawatt-hours of manufacturing capacity.

Author's comment: Although I believe that natural gas fueled generation will play a role in coming years, I believe it will be a minor role. The least expensive type of electric generation is PV, and batteries have plunged in cost per MW and MWh recently.

BloombergNEF (BNEF) 2025 survey: Global average turnkey BESS price fell to \$117/kWh in 2025, down ~31% from 2024 and ~40% from 2023 This is the lowest since BNEF began tracking in 2017.³

Back to Reference 1:

Some AI companies are experimenting with new, clean sources of firm energy. Google is building geothermal power plants that run on the earth's natural heat, while Microsoft is turning to nuclear power. But, says Gadepally, "I don't know how quickly we can scale technologies like geothermal and nuclear." Both companies, he points out, saw their emissions rise in 2024.

² https://www.statista.com/topics/2049/lithium-ion-battery-industry/#google_vignette

³ Andy Colthorpe, Energy Storage News, "Battery storage system prices continue to fall sharply, BNEF and Ember reports find," December 16, 2025, <https://www.energy-storage.news/battery-storage-system-prices-continue-to-fall-sharply-bnef-and-ember-reports-find/>

If we can't run all our data centers on clean energy, we can try to make AI models consume less energy.

A classic supercomputing program, like a climate model, processes a huge amount of data each time you run it—and is therefore run sparingly. AI is quite different, because the data that users input is tiny. Think, for example, of the text you enter into a language generator like ChatGPT.

What's large is the number of functions the model performs on that data. "Every time you talk to a generative AI model, you're passing in some data that is flowing through hundreds of billions of parameters," says Gadepally. "And in general, the thinking has been the larger the model—that is, the more parameters it has—the more accurate or realistic content it can produce."

This means that the computing power needed to "train" an AI model before its release to the public is immense. Developers perform these hundreds of billions of functions many times over on new data, slowly reinforcing the model to output better results.

2.2. AI Energy Use and Pathways to Efficiency

...Once a model is released, it continues to use a great deal of energy. Roughly speaking, says Gadepally, querying an AI model hundreds of millions of times is the same as training it all over again. "Think about how many millions of people are talking to ChatGPT or Anthropic or DeepSeek on any given day. If you have a billion interactions with ChatGPT, that's like training that model many times, which adds up very quickly."

We may be able to build high-performing AI models with less appetite for computing power. Gadepally's lab, for example, has studied methods to quickly weed out parameters that are not useful, making the training more efficient and shrinking the final model. And this January, DeepSeek, a Chinese AI firm, released a model trained with a fraction of the processing power of its rivals.

"It's a huge step forward," says Gadepally. "They were able to show that, if you're clever and you start to add some optimizations, we can actually do a lot more with less."

Like other models, DeepSeek contains hundreds of billions of parameters, but it never uses more than a fraction of them. When it gets a new query, it first considers what kind of question it's being asked, and then runs the data through only those parameters that are best at this type of task. This saves computing power and energy.

There's a difference, however, between using less energy per query, and using less energy overall. "You're able to reduce a lot of wasteful computation by using their novel architecture," Gadepally says. "The negative, though, is it has also dropped the costs. So now does this pull in an entirely new group of people who weren't planning to use it?" More frequent use could dampen, or even cancel out, the climate benefits of more efficient AI models.

That leaves one more avenue to lower AI's energy use: the data centers themselves. Gadepally's lab has seen impressive results from "power capping," or limiting the electricity each processor is allowed to draw. "And this is something I'm excited to say is now being picked up by basically everyone," he says. "We have seen reductions in energy use for workloads between 15 to 20%. And it doesn't change the answer in any way. It just makes it a little bit slower, by a couple of percent."

The industry is also moving away from energy-hungry air conditioning to cool its processors. "They do something called direct-to-chip now, where you basically have a cold-water loop that goes through your chip," Gadepally says. "That's more efficient." (It also, unfortunately, uses the same clean water we need for drinking and farming.) On the horizon is the even more efficient "immersion cooling," in which processors are submerged in a liquid like mineral oil, which conducts heat but not electricity.

The advantage of focusing on data centers is that they are used for much more than AI. "Every device we use is being processed somewhere," Gadepally points out. "Even the Zoom call for this interview is going through some data center. My phone is probably talking to some app at the same time, and that's using some compute." More efficient data centers lower the energy use, and the climate impact, of all this activity.

"I like to remind everyone that even small changes are okay," he adds. "We're doing billions of these things a day. So even a small change, when you multiply it by the number of times we're doing it, can be quite large."