

Training AI models doesn't emit that much

If we just make reasonable comparisons instead of crazy ones



ANDY MASLEY
APR 12, 2026

49 6 5

Share ...

The way the energy cost and emissions of training large new AI models are often talked about and compared to other things makes them seem unreasonably large. Here are some comparisons that were made following GPT-3 and 4, each about training specifically:

- “Training GPT-3 produced 552 tonnes of CO2, equivalent to driving 112 gasoline powered cars for a year” — [Columbia Climate School](#)
- “GPT-4’s footprint is roughly equal to the annual emissions of 1,550 US citizens” — [Shop Without Plastic](#)
- “GPT-3’s emissions are equivalent to the lifetime emissions of 8 cars — or 109 cars’ yearly emissions” — [Truthout](#)

I’m going to argue here that 1) The comparisons people make with training are often silly and misleading, and 2) When you make fair comparisons between training AI models and creating other products, the cost of training does not look unreasonably large at all. The way this is talked about often involves comparisons that would make literally any popular consumer product look ridiculous and wasteful. I think these comparisons are obviously goofy, and if I make more reasonable comparisons while keeping the numbers the same, you will see that training AI models is not some unique environmental catastrophe, and actually just blends into all the other ways society uses energy normally.

- What is training?
- Some bad comparisons
- Two useful comparisons
- A lot of popular coverage of this used terrible comparisons
- State of the art training runs for current models are larger, but don't emit enough to change my point here, and they have way more users
- Should I also include all the other climate costs of training, like the physical infrastructure and failed training runs?

What is training?

Training an AI model basically means creating it. Untrained models are fed huge amounts of data to pick up more and more subtle background patterns. This involves connecting tens of thousands of specialized AI computers very close together to constantly communicate with each other over months, typically all housed in a single large data center. This uses a lot of energy.

Once a model is trained, it can be used over and over and does not need to be retrained. Newer more capable models are trained instead. Models are often used for months before being replaced by newer ones.

Importantly, spending this much energy is a necessary step to creating frontier chatbot models. We currently do not know of a way of creating models with similar capabilities using much much smaller training runs.

This uses a lot of energy by the standards of everyday buildings, but my argument here is that it doesn't use much energy or emit that much by the standard of creating any other popular products.

I see training as being analogous to manufacturing a new product for users to buy. Just like physical objects people buy and use have to be manufactured first, AI models people interact with have to be trained first. Both are necessary steps in creating the final products the users interact

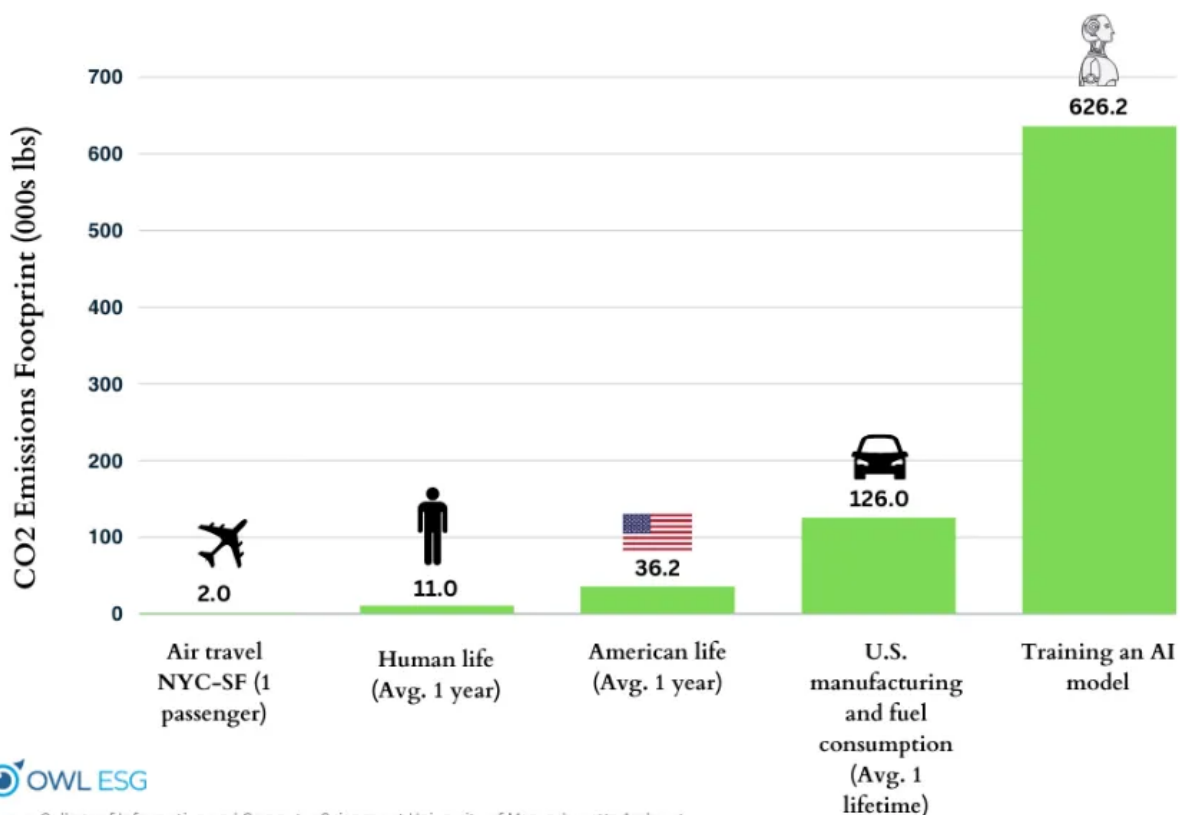
...:tl

Some bad comparisons

Your personal emissions

Take a look at this graph, it and graphs like it have been very popular as a way of showing how bad AI is for the environment:

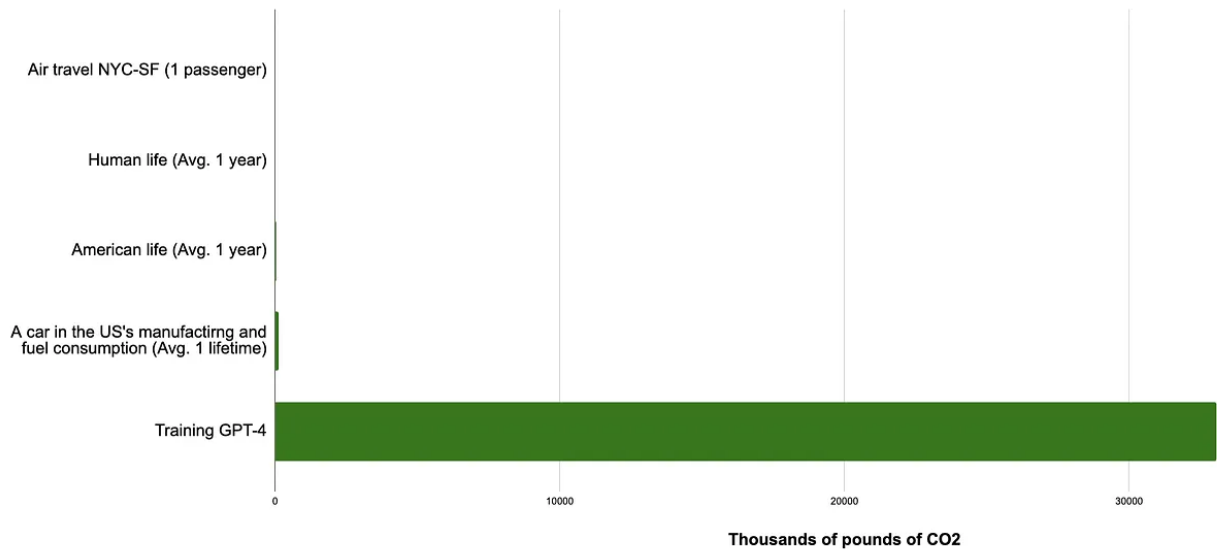
CO2 Emissions Benchmarks



Source: College of Information and Computer Sciences at University of Massachusetts Amherst

[Source](#)

Importantly, this graph is about a very old 2019 model (way before commercial chatbots were available) that didn't get much use. The stats from this old unused model made it into articles about AI's environmental footprint 5 years later. Yet it still gets shared a lot to make points about current consumer chatbots. I'll update the graph with the much much larger estimated training cost of GPT-4. I'm going to start with GPT-4 as an example because this is the first model where everyone started freaking

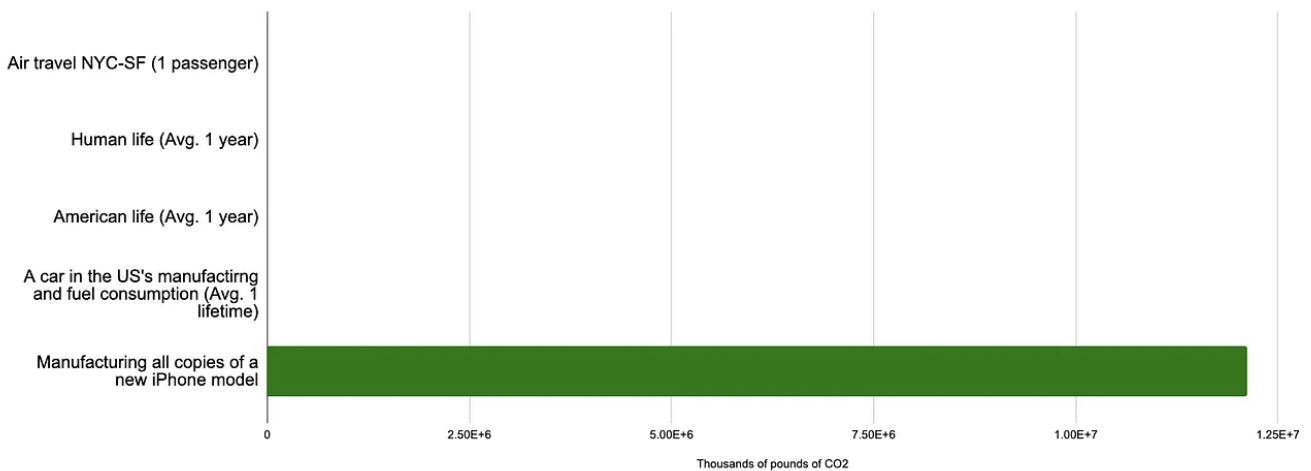


Much bigger!

I'm using a high-end estimate of GPT-4's training, partly to include things like the cost of failed training runs. This cost is 15,000 tonnes of CO2. ¹

Is this a reasonable way to think about GPT-4's emissions? Is it a useful comparison?

How reasonable does this other comparison look, where instead of training GPT-4 the bottom line shows my rough guess for the emissions of manufacturing every new model of the iPhone 16? ²

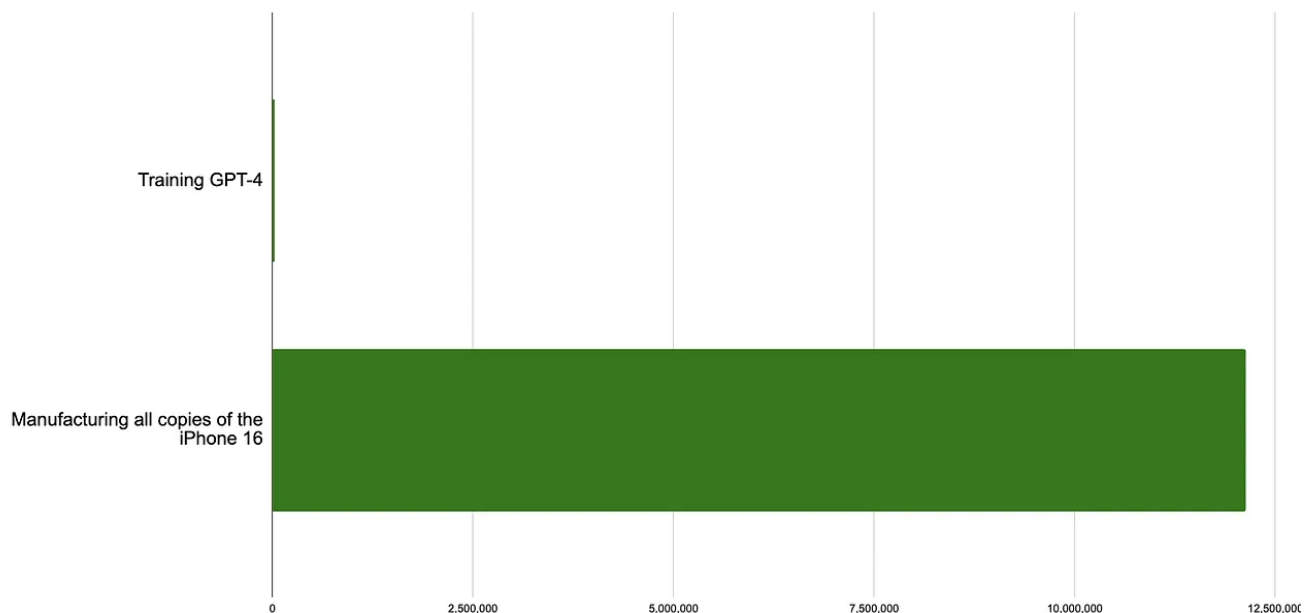


Does this tell you anything about whether you personally should or should not purchase an iPhone?

This site requires JavaScript to run correctly. Please [turn on JavaScript](#) or unblock scripts

This is my big claim: **It is not useful or reasonable to measure the emissions from the creation of a product used by hundreds of millions of people every day against your personal emissions, or the emissions of buying a single plane ticket. That tells you basically nothing about how damaging it is for you personally to use the product, unless you divide by the number of users.**

This seems obvious, but almost all reporting on how much energy AI training runs use compare it to individual things people do. A better comparison is the cost of creating the AI model to the cost of creating other products used by hundreds of millions of people. When you do that, it becomes difficult to understand why the emissions of training AI models are receiving so much scrutiny:



Estimate of manufacturing all copies of iPhone 16

Energy of cities

Another common comparison people make with AI training runs is the electricity costs of whole cities. For example, the [MIT Technology Review](#) notes that training GPT-4 “consumed 50 gigawatt-hours of energy, enough to power San Francisco for three days.” Is this useful? Well, the electricity homes, businesses, and industry in a single city use doesn’t give us too much useful context for how GPT-4 compares to other consumer products.

This site requires JavaScript to run correctly. Please [turn on JavaScript](#) or unblock scripts

Again, take manufacturing iPhones. Manufacturing all iPhone 16s emitted approximately 12 million tonnes of CO₂.³ Generating San Francisco's electricity emits about 890,000 tonnes of CO₂ per year. So all iPhone 16s emitted as much as 13 years of San Francisco electricity.⁴ Does this tell us much about whether you personally should buy an iPhone? Does it add much useful context at all?

Flights

One other common comparison is airplane flights, measured as the total plane's emissions rather than emissions per capita. GPT-4's training emissions were approximately as much as 38 full plane flights from San Francisco to Australia⁵.

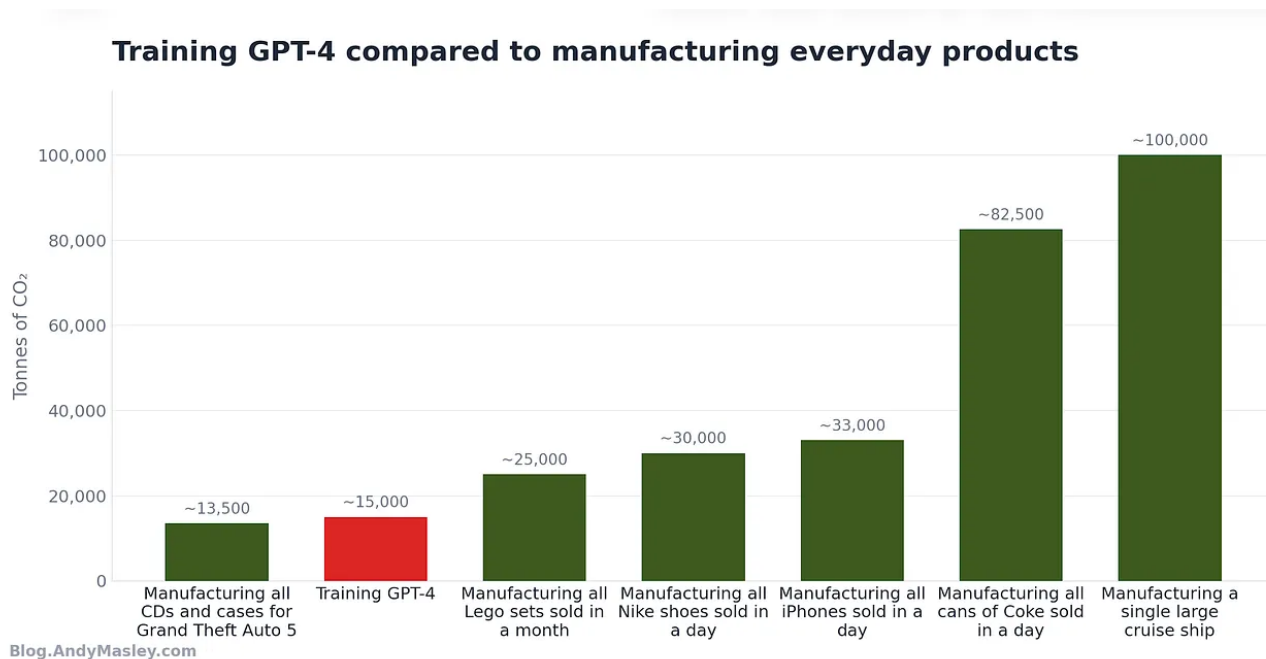
NeurIPS is the largest annual ML/AI research conference. This year it's being hosted in Sydney, Australia. Based on past numbers I'd estimate 15,000 people are going to travel to the conference this year⁶ from all around the world, enough to completely fill 50 planes. If the average attendee is flying from elsewhere, like San Francisco, this means that a single annual AI research conference will emit more than training GPT-4.

Imagine if an AI research conference was held, and it was so useful for attendees that it singlehandedly created GPT-4 from scratch, where only GPT-3.5 existed before. This seems like a pretty worthwhile get-together. Here, I don't think people would even bring up the emissions. Right now few seem to be commenting on the emissions of the conference. I think that a headline "AI company emits as much as 38 planes to create GPT-4" is much more attention grabbing than "AI researchers from around the world take 50 planes to all hang out in Australia" even though the second is much worse for the climate and has likely been significantly less consequential for AI as a field, or everyday people's lives.

Two useful comparisons

Total: Other products

Training is a necessary part of creating a product that hundreds of millions of people interact with every day. One way we can compare training AI models is to the emissions of creating similarly popular consumer products that hundreds of millions of people interact with. Here's a graph of the total emissions of training GPT-4 vs my best estimates of emissions from the production of other consumer products:



Everything here's a rough estimate based on publicly available data.

[Explanation of each estimate here.](#)

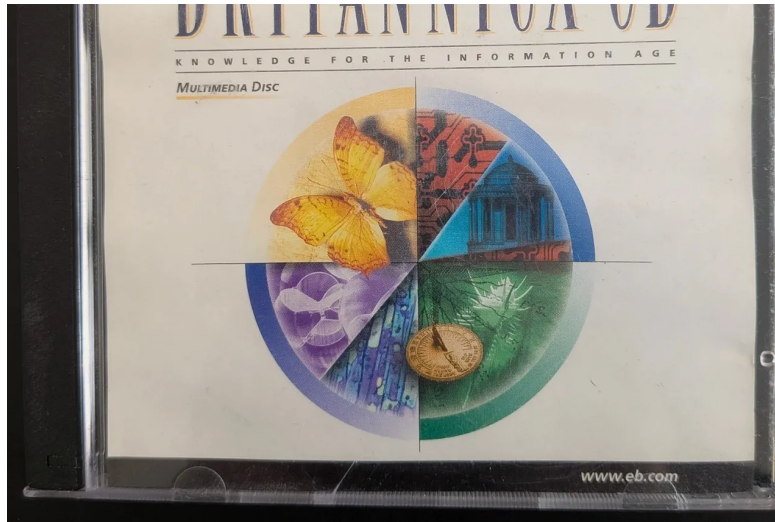
Just like physical CDs and cases had to be manufactured for hundreds of millions of people to play Grand Theft Auto 5, GPT-4 had to be trained before hundreds of millions of people could use it. Both are necessary steps in making a popular product hundreds of millions of people use every day.

Per-user: CDs

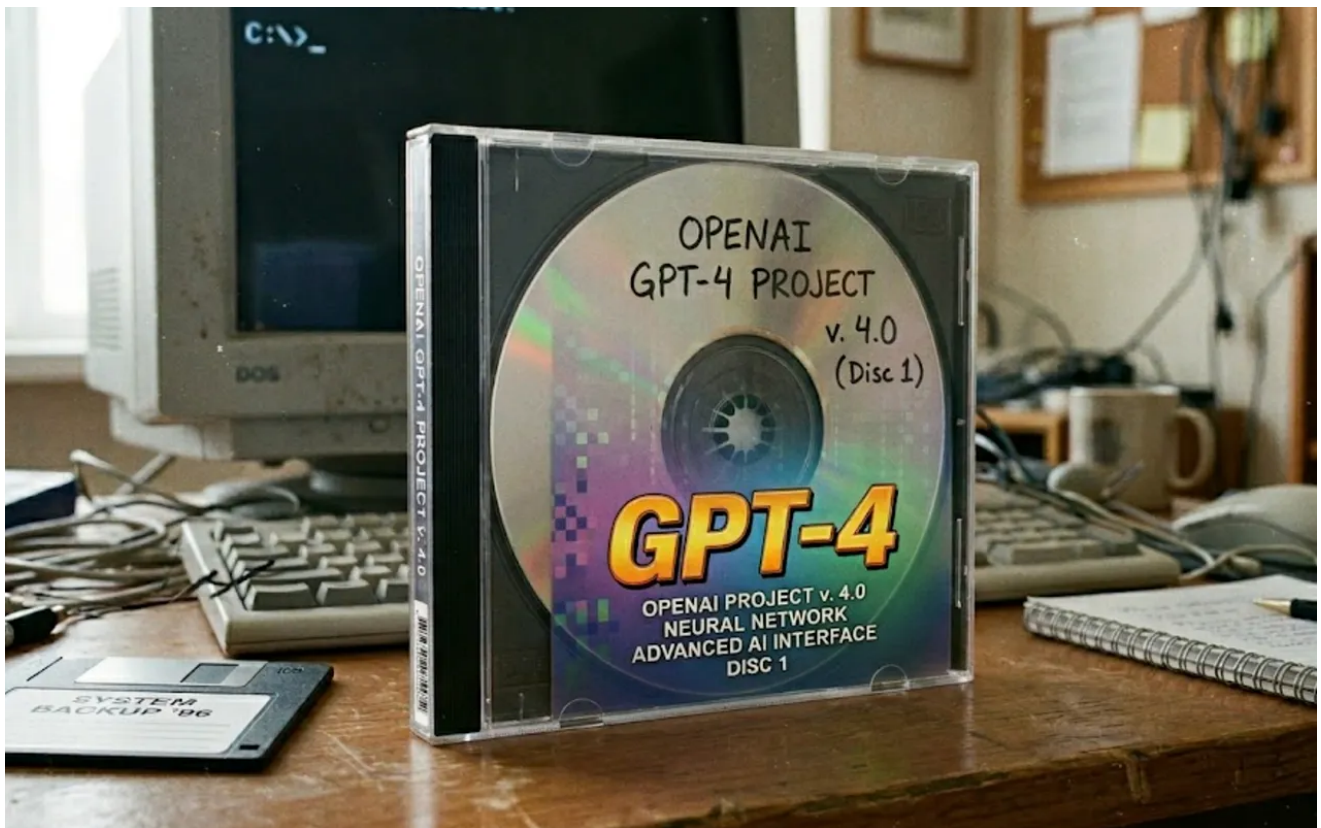
I'm old enough to have memories of going out and buying CDs for computer games I wanted to play. One of the coolest objects I owned as a kid was a CD encyclopedia. The whole world's knowledge at my fingertips!



This site requires JavaScript to run correctly. Please [turn on JavaScript](#) or unblock scripts



There are no exact numbers for how many people were using ChatGPT over the time that GPT-4 was available, maybe somewhere around 100-200 million weekly users is a reasonable guess. Dividing GPT-4's cost of training by this number of users gets to 75-150 grams of CO2 emitted by training per user. This happens to be a little lower than the emissions of manufacturing a single CD. It's as if each weekly user of ChatGPT had purchased a CD to use it.



This is the cost of training everyone was freaking out about when GPT 4

This site requires JavaScript to run correctly. Please [turn on JavaScript](#) or unblock scripts

was released. It's like the company didn't have to spend any energy on training at all, and to use ChatGPT you just had to buy it on a CD, which emits a little to manufacture. I don't think that if GPT-4 didn't involve training at all but came on a CD, anyone would be talking too much about the emissions of manufacturing those CDs, and yet those emissions are the same as the training cost that people speak so ominously about. This is very small!

A lot of popular coverage of this used terrible comparisons

Here's a ton of quotes from different articles on the cost of training GPT-4 and previous models. Not one of them compared training to other consumer products used by hundreds of millions of people, or divided by the number of people to get a per-user number. I haven't found any sources doing either. Read through these and think about whether the reader would come away with a better or worse understanding of the magnitudes involved. Would they know creating a model used by hundreds of millions of people had about the same climate impact as the CDs and cases for Grand Theft Auto 5?:

- "Training GPT-3 produced 552 tonnes of CO2, equivalent to driving 112 gasoline powered cars for a year" — [Columbia Climate School](#)
- "GPT-4's footprint is roughly equal to the annual emissions of 1,550 US citizens" — [Shop Without Plastic](#)
- "GPT-3's emissions are equivalent to the lifetime emissions of 8 cars — or 109 cars' yearly emissions" — [Truthout](#)
- "That's like a car driving 1.2 million miles" — [Carbon Credits](#)
- "500 tons of carbon dioxide emissions is the equivalent of around 600 flights between London and New York" — [Cybernews](#)
- "GPT-3 emitted carbon dioxide equivalent to 500 times the emissions of a New York-San Francisco round trip flight" — [Analytics Vidhya](#)
- "It's the same amount of emissions as a single person taking 550

Credits

- “Training GPT-4 consumed 50 gigawatt-hours of energy, enough to power San Francisco for three days” — [MIT Technology Review](#)
- “15,000 tons is roughly the same as the annual emissions of 938 Americans” — [Towards Data Science](#)
- “Enough energy to power an average U.S. home for over 120 years” — [Truthout](#)
- “Training a single AI model can emit as much carbon as five cars in their lifetimes” — [MIT Technology Review](#) (headline)
- “Training GPT-3 consumed 1,287 megawatt hours of electricity (enough to power about 120 average U.S. homes for a year), generating about 552 tons of carbon dioxide” — [MIT News](#)
- “Training a single large model like GPT-3 can use over 1,200 MWh — enough electricity to power around 120 U.S. homes for a year” — [Climate Impact Partners](#)
- “Training GPT-3 emitted roughly 500 metric tons of carbon dioxide — the equivalent of driving a car from New York to San Francisco about 438 times” — [Climate Impact Partners](#)
- “Training GPT-3 produced 552 tonnes of CO₂, which is equivalent to the emissions from 110 gas-powered cars over a year” — [Smartly.AI](#)
- “GPT-3 emitted carbon emissions equivalent to the lifetime impact of five cars” — [Nature Scientific Reports](#)
- “Training the bigger, more popular AI models like GPT-3 produced 626,000 pounds of carbon dioxide, equivalent to approximately 300 round-trip flights between New York and San Francisco — nearly five times the lifetime emissions of an average car” — [The Sustainable Agency](#)
- “This is similar to the yearly emissions produced by 120 passenger cars or 600 transatlantic flights per person” — [Shop Without Plastic](#)
- “Training AI models can emit more than 626,000 pounds of carbon dioxide equivalent — nearly five times the lifetime emissions of the average American car (including the manufacture of the car itself)” —

- “Training AI models can emit more than 626,000 pounds of carbon dioxide equivalent — nearly five times the lifetime emissions of the average American car” — [Supermicro](#)
- “Training a single AI model can emit 626,000+ pounds of CO2 equivalent... about 5x the lifetime carbon emissions of an average passenger car” and “training a single AI model can consume more electricity than one hundred American homes use in one year” — [Carbon Credits](#)
- “Training such a model requires... This is equivalent to the annual energy consumption of around 160 average American homes” — [Frankly Kranky](#)
- “GPT-3 consumed approximately 1,287 MWh of electricity—enough to power 120 US homes for a year... A single AI model training session can emit as much carbon dioxide as five cars over their entire lifetimes” — [AI Energy Calculator](#)
- “Training ChatGPT-3 is estimated to have required the equivalent energy consumed by an average American household for 120 years” — [California State University Northridge Library Guide](#)
- “GPT-3 consumed a staggering 1,287 megawatt-hours (MWh) during training. This resulted in about 502 metric tons of carbon emissions, equivalent to the emissions from hundreds of gasoline-powered cars in a year” — [PatentPC](#)
- “The process can emit more than 626,000 pounds of carbon dioxide equivalent — nearly five times the lifetime emissions of the average American car” — [ACM Communications](#)
- “The process can emit more than 626,000 lbm of carbon dioxide equivalent — nearly five times the lifetime emissions of the average American car” — [SPE Journal of Petroleum Technology](#)

None of these comparisons make sense, because basically all of them treat training as if it's comparable to activities individual people do, instead of creating a product expected to get hundreds of millions of people using it. If instead we treat training GPT-4 like this, it fades into basically nothing. If you think it would have been worthwhile for the climate to prevent GPT-4

of Lego for 2 weeks. If you're not telling everyone to boycott Lego or Coke or Grand Theft Auto because of the emissions involved in creating them, you shouldn't be telling people to boycott ChatGPT for the emissions involved in creating GPT-4. If you read "Manufacturing all the CDs for Grand Theft Auto 5 emitted as much as 160 American homes" I think you might be a little underwhelmed.

State of the art training runs for current models are larger, but don't emit enough to change my point here, and they have way more users

Everything above uses GPT-4 as an example because that's the model that kicked off all the panic about training emissions. But GPT-4 is old now. Current state of the art models are bigger and more expensive to train. Does this change anything?

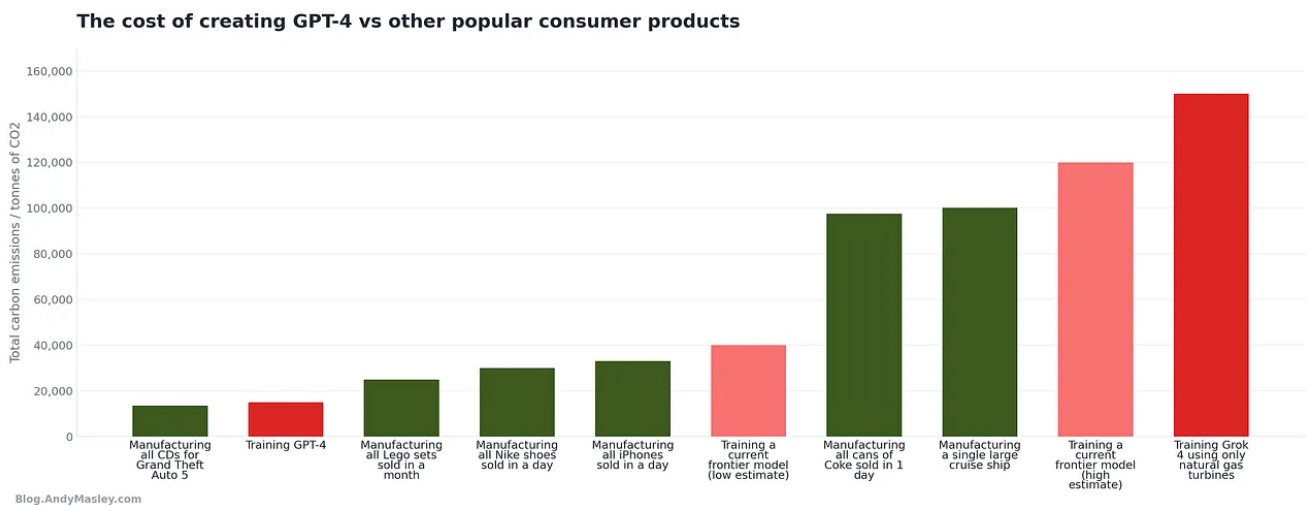
Training costs are growing at about 2.4x per year for frontier models. The largest training run we have decent estimates for is Grok 4, which Epoch AI estimates used about 310 GWh of electricity and cost around \$490 million, roughly 6-7x what GPT-4 cost. But Grok 4 is probably the worst-case scenario for training emissions, because xAI's Colossus data center in Memphis was powered largely by mobile natural gas turbine generators, which emit about 0.49 kg CO₂ per kWh, 1.3x higher than the US grid average of about 0.37 kg CO₂ per kWh, which itself is higher than some of the grids where AI models are typically trained, like Oregon which relies more on hydropower. This is why Epoch estimates Grok 4 emitted about 154,000 tonnes of CO₂, ten times GPT-4's estimated emissions, despite using only about six times as much electricity.

Most other frontier labs don't train on natural gas generators. They primarily train models in large data centers connected to the electrical grid, many of which are in places with more renewable energy. The same 310

US grid, 75% of its actual emissions. On California's grid it would only emit about 62,000 tonnes, 40% of its actual emissions. This means that a frontier training run of similar size to Grok 4, but run on a typical major cloud provider's infrastructure, would likely emit somewhere in the range of 60,000–120,000 tonnes.

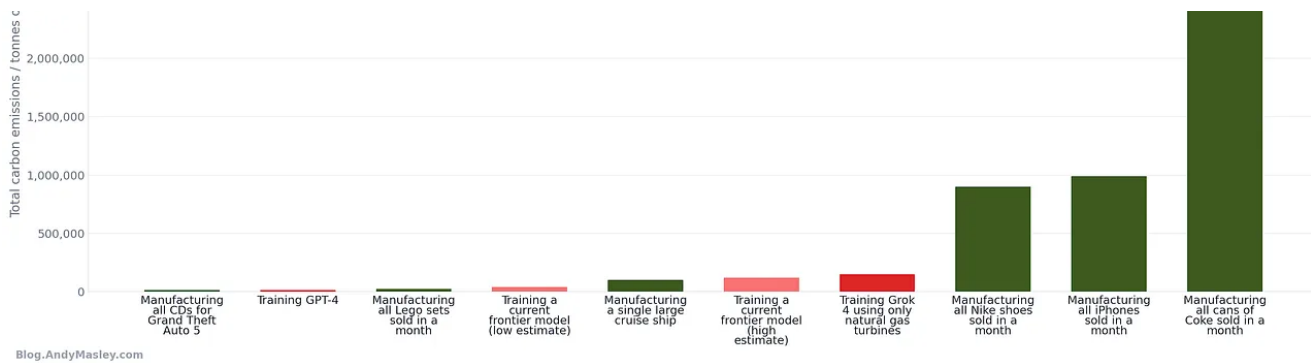
Not every frontier model uses as much compute as Grok 4, either. Grok 4 was trained on 200,000 GPUs and is probably the single largest training run to date. Other state of the art models may use significantly less compute, bringing their energy use and emissions lower. Hardware energy efficiency has been improving at about 40% per year, so the next generation of training runs won't need proportionally more energy even if they use more compute.

So what's a reasonable range for the emissions of training a current frontier model? Probably somewhere between 40,000 and 120,000 tonnes of CO₂, with the wide range driven both by differences in the carbon intensity of the electricity and differences in the amount of energy used.



Is this a lot? Well, these models get used for months at least. I could just up the manufacturing numbers to a month each.

The cost of creating GPT-4 vs other popular consumer products



AI models still just don't stand out much compared to other popular products.

For now I'll use 80,000 tonnes as a round middle estimate for the current emissions of training an average frontier model.

Two things have also changed since GPT-4:

1. There are way more users. When GPT-4 was the main model, ChatGPT had roughly **100–200 million weekly users**. Now ChatGPT has **900 million weekly active users**. That's roughly a 5–9x increase in users over the same period that training costs went up maybe 4–6x. The per-user training footprint has likely stayed roughly flat or even declined. 900 million users over 80,000 tonnes of CO2 for training is about 89 grams. This is still about half a CD. Even under Grok 4's worst-case emissions of 154,000 tonnes, assuming similar user numbers would mean the per-user cost is about 171 grams, roughly one CD.
2. Even at the total level, the comparisons mostly don't change. The most advanced AI models that huge numbers of users interact with for months at a time use about as much energy to create as it takes to manufacture 3–4 days of iPhones.

My basic point has so far not changed with larger models: creating an individual AI product used by hundreds of millions of people produces emissions that are small relative to manufacturing basically any other product used by hundreds of millions of people.

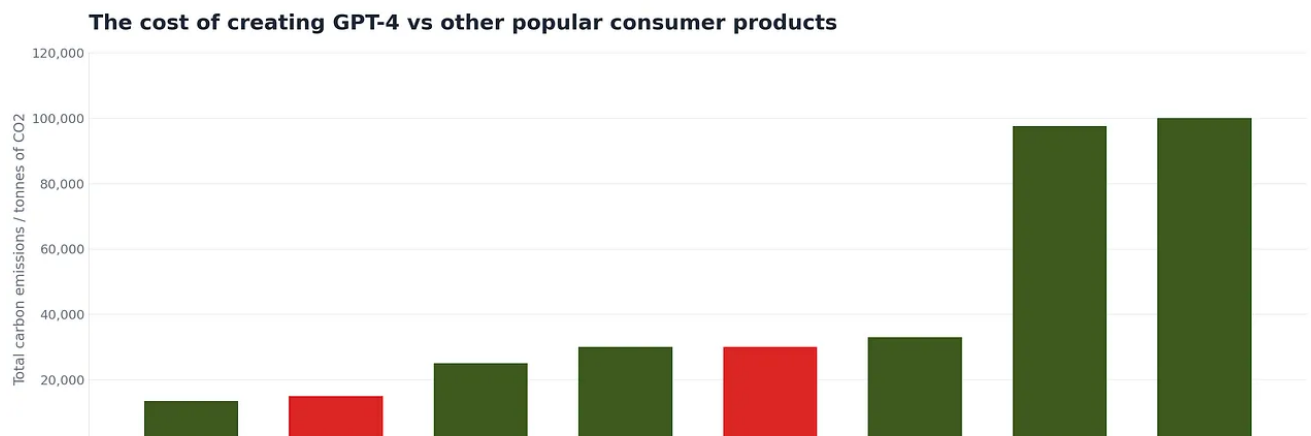
Training runs are going to continue to use more energy, but models are also getting more capable over time. Eventually this may be more like each user

much more capable the investment will, I expect, be worth it.

Should I also include all the other climate costs of training, like the physical infrastructure and failed training runs?

Maybe my comparison to creating other products is unfair here, because there are lots of other analogous steps to manufacturing before the chatbot can be delivered to hundreds of millions of people. OpenAI also had to buy tens of thousands of GPUs, which had to be manufactured in fabs and assembled into servers. Those servers sit in data centers that had to be built out of steel and concrete. What's the total carbon footprint of all that?

Let's try to make this number as big as possible. We can include the embodied hardware emissions, the data center construction, the cooling infrastructure, and the experimental training runs that failed before the final model worked. [Hugging Face found](#) that when they accounted for all of these factors for the model BLOOM, total lifecycle emissions roughly doubled, from 25 tonnes to about 50 tonnes. The estimate I chose for GPT-4's training cost is 3 times as high as many popular estimates already, so it likely includes these costs already, but let's double it anyway to be safe, to get a total of about 30,000 tonnes of CO2 emissions from literally everything required to deliver a model to users.



This site requires JavaScript to run correctly. Please [turn on JavaScript](#) or unblock scripts

Even at this extreme, training GPT-4 would still be smaller than building a single cruise ship. It would still be less than one day of manufacturing Nike shoes. It would still be less than a single day of manufacturing iPhones.

But I also didn't include many of the additional costs of the other products too. I didn't include building the factories or making the mining equipment for extracting raw materials or transporting iPhones to stores or the cost of the retail stores themselves... many accusations that "you're leaving information out" are one-directional, and don't consider that most things we do have lots of hidden costs depending on what we want to compare. Ultimately I think the direct comparison between training and creating other popular products is completely legit.

The per-user comparison doesn't really change even under the most extreme interpretation. It raises the per-user cost of training GPT-4 from one CD to two. The reason the cost of creating a product used by so many people is so low compared to things like Coke and sneakers is that in comparison, **computing is very energy-efficient**, and thus it's very rarely the most promising thing to optimize or cut for the climate.

It's true that all training across all AI models is forecast to make up a significant fraction of US electricity use in the next few years, maybe rising to whole percentage points of US electricity usage. This is because a huge portion of total AI training will be concentrated in America, and because AI is entirely electrified, while most other industries aren't. This post was mainly about the training of individual AI models. In the next post I'll try to paint an overall picture of where training and AI more broadly fit into America and the world's overall energy and climate situation.

-
- [1](#) The [2025 Stanford AI Index Report](#), using Epoch AI's methodology, estimates GPT-4's training emissions at approximately 5,184 tonnes CO2. A [separate widely-cited analysis](#) by Kasper Groes Albin Ludvigsen in Towards Data Science, based on leaked hardware specifications (25,000 Nvidia A100 GPUs running for 90-100 days) estimates 51-62 GWh of electricity consumption and 12 456-

difference largely comes from assumptions about the electricity's carbon intensity.

I'm using the higher ~15,000 tonne estimate for two reasons:

- I want to use the very highest defensible number I can find to show that even then training doesn't look large.
- The lower estimates only cover the final successful training run and do not account for the full cost of developing the model. [Epoch AI estimates](#) that the ratio of total development compute to final training run compute ranges from 1.2x to 4x across frontier models, with a median of 2.2x.

Applying the median 2.2x experimentation multiplier to the Stanford estimate of 5,184 tonnes gives ~11,400 tonnes, and adding embodied hardware emissions could bring it to ~14,000 tonnes, close to the Ludvigsen estimate. So ~15,000 tonnes can be read either as the high-end estimate for the final training run alone (assuming a dirtier grid), or as a more realistic all-in estimate that includes experimentation and embodied emissions on a moderate grid. Either way, it represents what looks like a reasonable upper bound, and every comparison in this post becomes even more favorable to AI training if the true number is lower.

- [2](#) [Apple's Product Environmental Reports](#) give lifecycle emissions of 56-74 kg CO₂e per iPhone 16 depending on model and storage, with [approximately 80% from production](#) — roughly 45-59 kg CO₂e per device for manufacturing. [Apple shipped approximately 230 million iPhones in 2024](#). 230 million iPhones × 52 kg = approximately 12 million tonnes.
- [3](#) See footnote 2 ^
- [4](#) San Francisco County consumes roughly 5,100 GWh of electricity annually, producing approximately 890,000 metric tons of CO₂. Source: [FindEnergy](#).
- [5](#) The San Francisco-Sydney route is approximately 7,400 miles (11,900 km). A [Boeing 777-200 flying a comparable distance \(Chicago-Hong Kong, 7,821 miles\)](#) burns roughly 42,000 gallons of jet fuel. At approximately 6.5 pounds per gallon, that's about 124,000 kg of fuel. Each kilogram of jet fuel produces [approximately 3.16 kg of CO₂](#) when burned. This gives roughly 390 tonnes of CO₂ per flight for the entire aircraft. 15,000 tonnes ÷ 390 tonnes per flight ≈ 38 flights.
- [6](#) Last NeurIPS in San Diego had [24,500 people](#) attend in-person. This one is such a long trip that I'm assuming it would get ~half the attendance.

Discussion about this post

Comments Restacks



Write a comment...



Kenny Easwaran  Kenny Easwaran 17h

...

 Liked by Andy Masley

This is very helpful, since I'm updating a lecture on the resource use of AI for an AI literacy class I'm teaching right now!

A few minor points - in my academic field (philosophy) there are some people who are complaining about the carbon emissions associated with conference travel, and arguing that we should have more online conferences and fewer in-person conferences. (See, eg, the comments here: <https://dailynous.com/2026/01/15/apa-to-end-experiment-with-online-divisional-meetings/>) I expect there are some people making similar arguments on computer science.

One thing that this really drives home for me is that there's a difference between the general resource use of AI and specifically the electricity use. Unlike legos and CDs, which involve moving physical objects and making things out of refined petroleum, AI has its emissions almost entirely through electricity. It looks like there are individual data centers that use as much electricity as the entire city of San Diego! (For instance, the Amazon datacenter for Claude training in New Carlisle, IN: <https://epoch.ai/data/data-centers/> .) But the carbon emissions and water use of this datacenter are more comparable to the emissions and water use of South Bend, or perhaps even New Carlisle. The water use change could be a significant effect locally, but it's nowhere near the potential disruption of having to power a new San Diego in the middle of Indiana!

The issues of water and emissions will be handled by the same processes that are handling those issues for everything else in the world. But electricity use will be a significantly different type of issue to deal with, especially as we also electrify transportation, cooking, and climate control.

 REPLY

 SHARE



Energy x AI Energy x AI 12h

...

 Liked by Andy Masley

"GPT-4's footprint is roughly equal to the annual emissions of 1,550 US citizens" Was this intended by the author to seem like a large amount of emission? Are people really seeing this and thinking "AI bad"? It's tiny for a global product. Like, *obviously* you divide it by the number of users. Maybe I'm too optimistic about how people view numbers. (It's

worth)

On another note, I want to float something I'm tentatively calling the Masley paradox - the idea that training and using LLMs has negligible electricity associated with them compared to other activities on a per person basis, while data centre growth has such strong local implications for electricity costs (via increased network and generation requirements).

New data centres are largely going in specific locations where a) the regulatory and investment frameworks are favourable, and b) the internet cable connectivity is favourable, such as greater Sydney in my backyard, leading to a lot of local efforts to see it goes right.

 REPLY

 SHARE

1 reply by Andy Masley

4 more comments...