Why ChatGPT may destroy the planet

(slight exaggeration)

Goals of this presentation:

- Where does the environmental impact of generative AIs come from?
- How to measure the energy consumption of LLMs?
- How to reduce the environmental impact of AI as a developer?
- How do I reduce my impact as a user?

Question time !

What is the environmental impact of using ChatGPT?
(e.g. sending request)

First, a little bit about me

I’m Tristan Coignion, and I’m defending my PhD thesis on LLM’s energy consumption next week!

You can also find me on my website
https://saauan.github.io

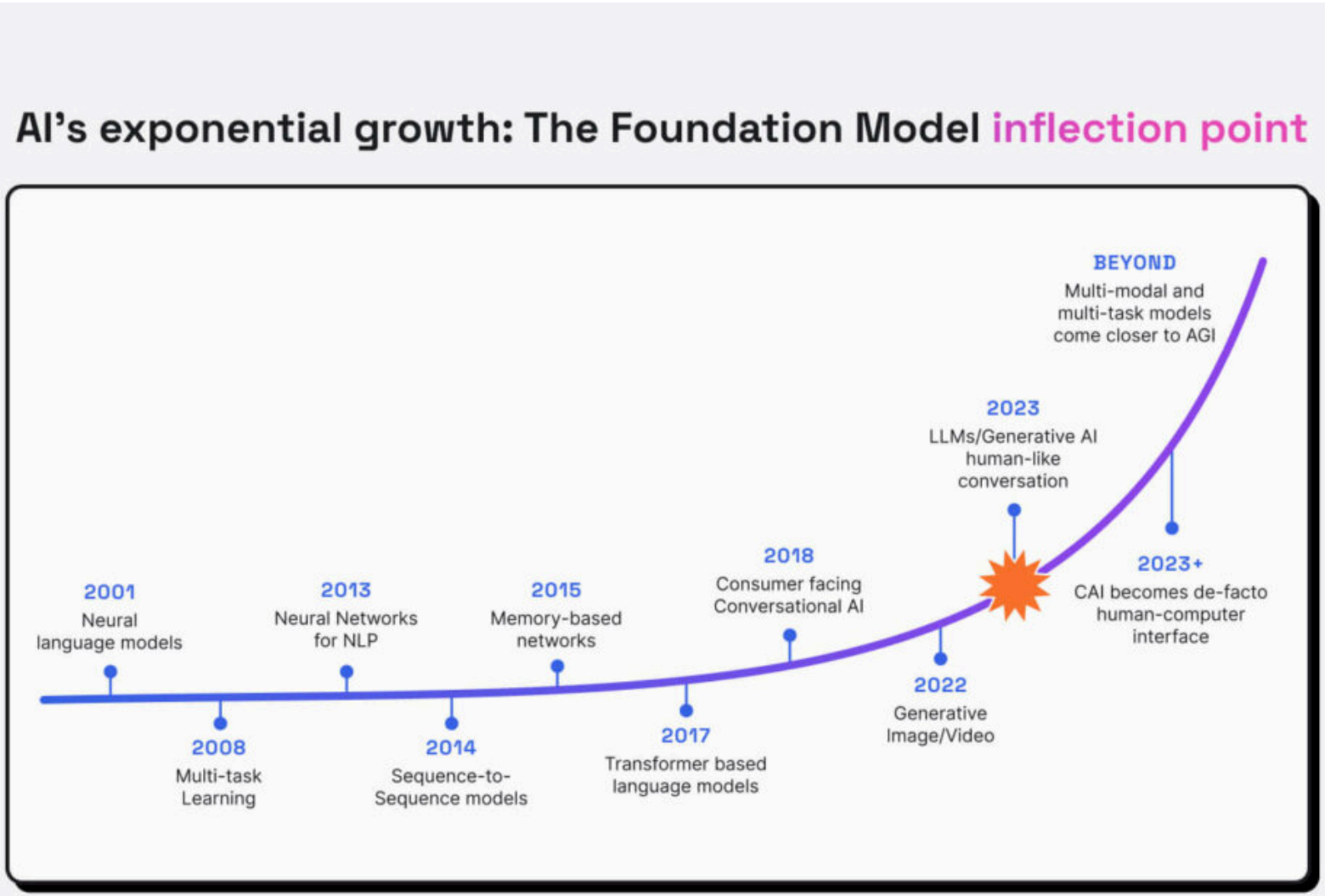
Please, interrupt me any time if you have a question or a remark

Note that while most of this course is based on empirical research and facts, some parts are explicitly based my own opinions and beliefs.

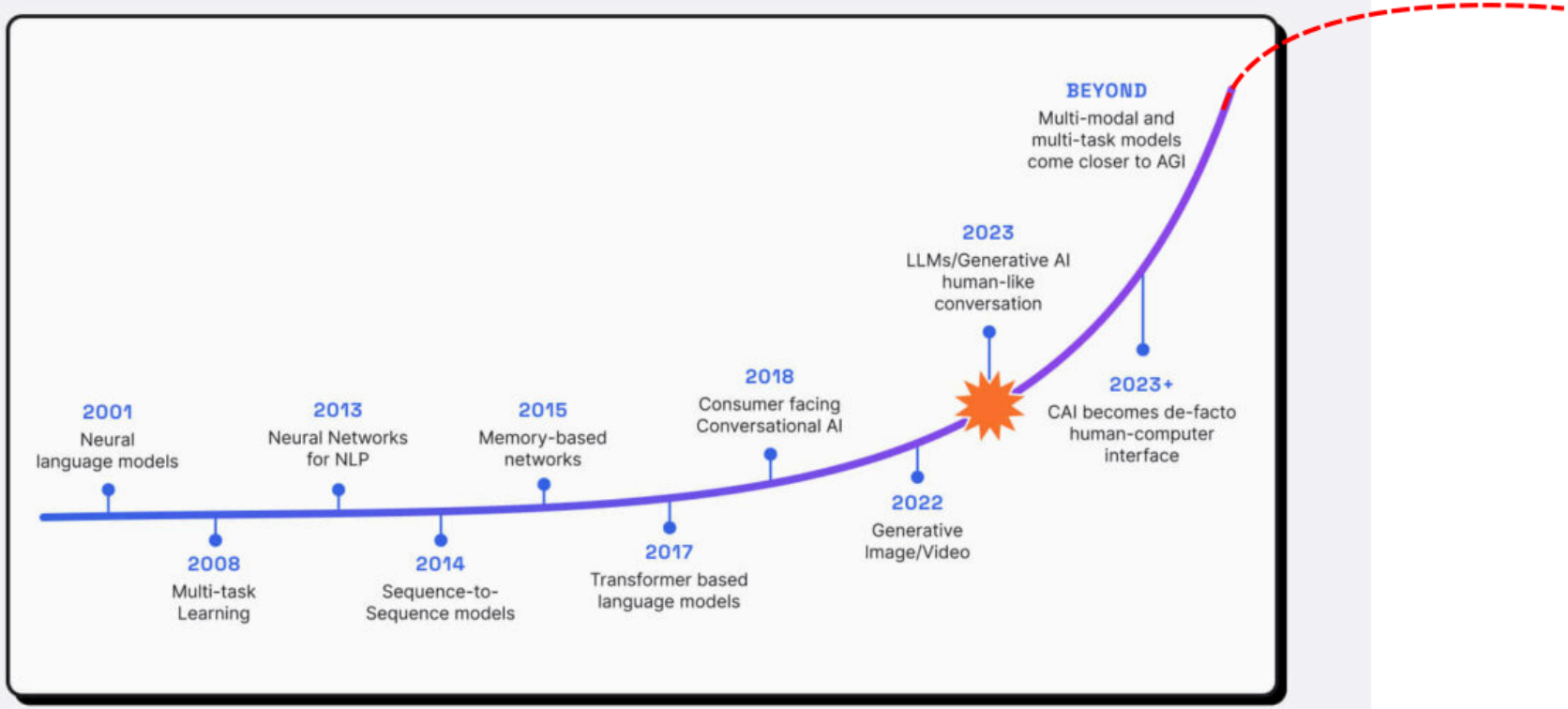
Reminders: The environmental impact of software


THIS IS NOT FINE.
WILD FIRES
HEAT WAVES
MELTING GLACIERS

The boom of generative AI



AI's exponential growth: The Foundation Model inflection point



LLMs

A family of AI that are based on the Transformer architecture and can read and produce text (but now they also take other modes of input)

Anatomy of the environmental impact of LLMs (mostly applicable to generative AI as a whole)

Where do the impacts come from?

- Energy
- Water
- Hardware

Training & Fine-tuning

A lot of GPU usage for thousands of hours.

e.g. BLOOM (176B) trained in France is estimated to have emitted 24.7 tonnes of CO2 (with only the energy) and 50 tonnes of CO2 if we include the equipment needed to train.

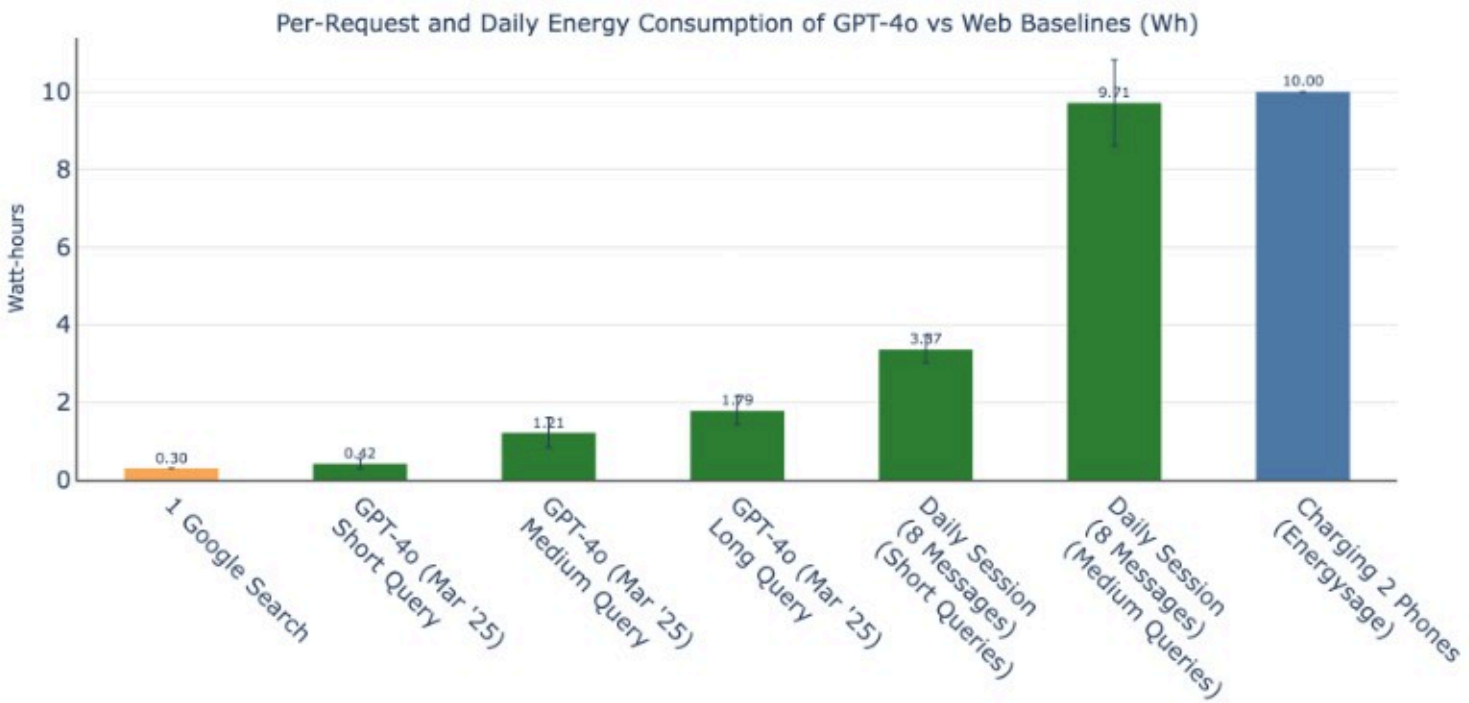
[1] *Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model*. Alexandra Sasha Luccioni, Sylvain Viguier, Anne-Laure Ligozat

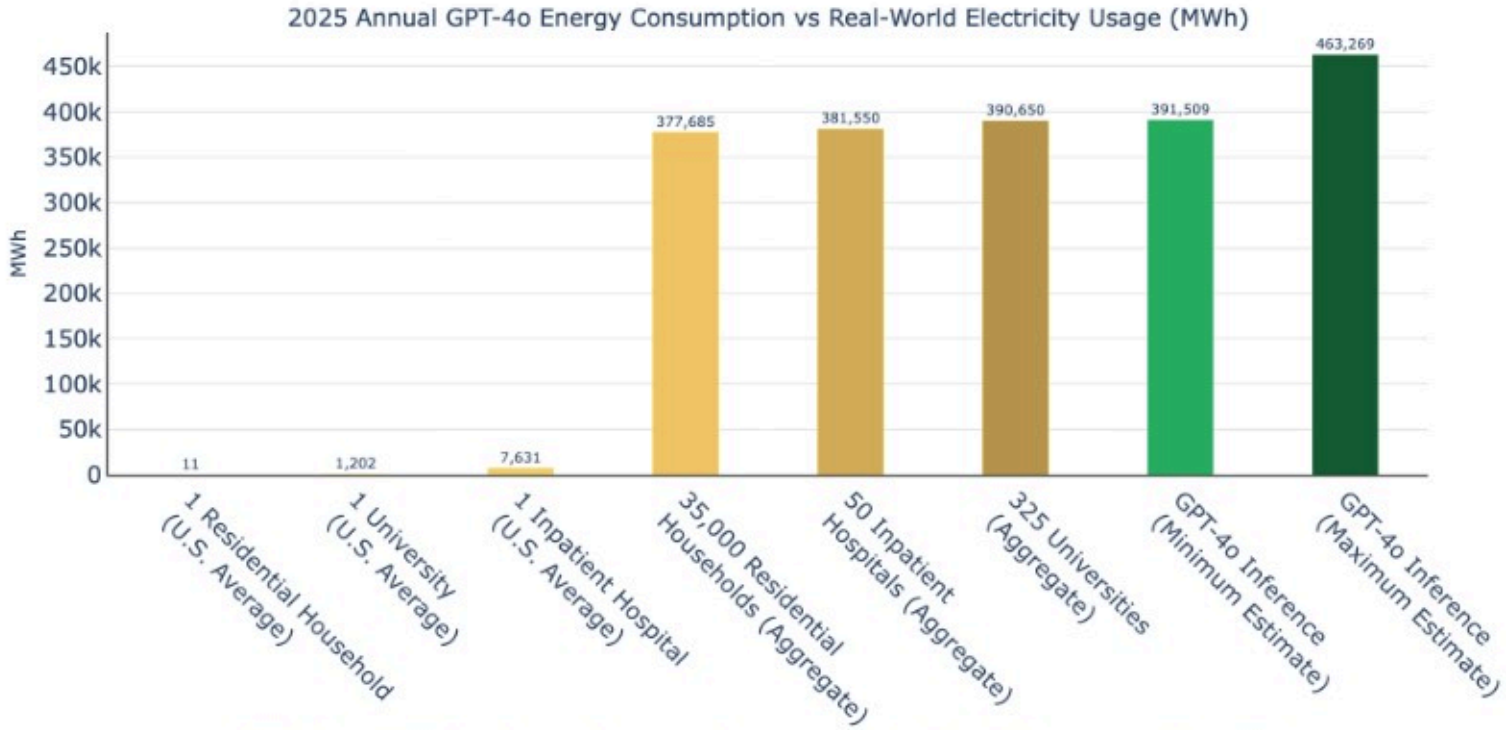
Inference

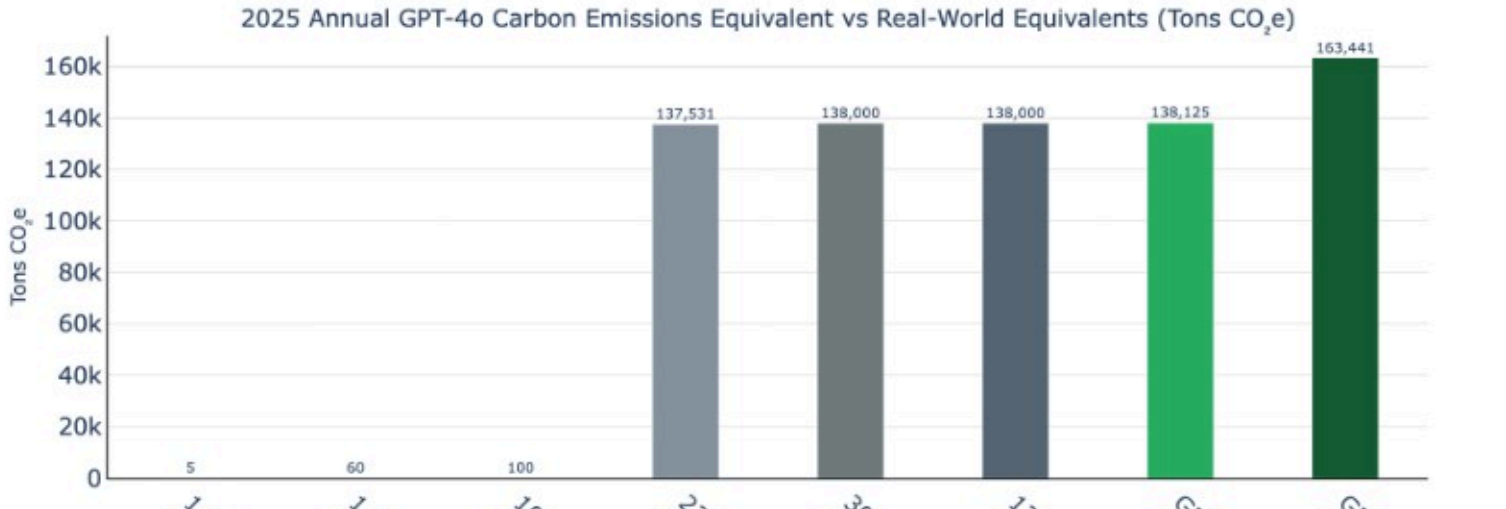
One inference <<< One training

But millions of inference each day. In the end a good part of the impact comes from the inference.

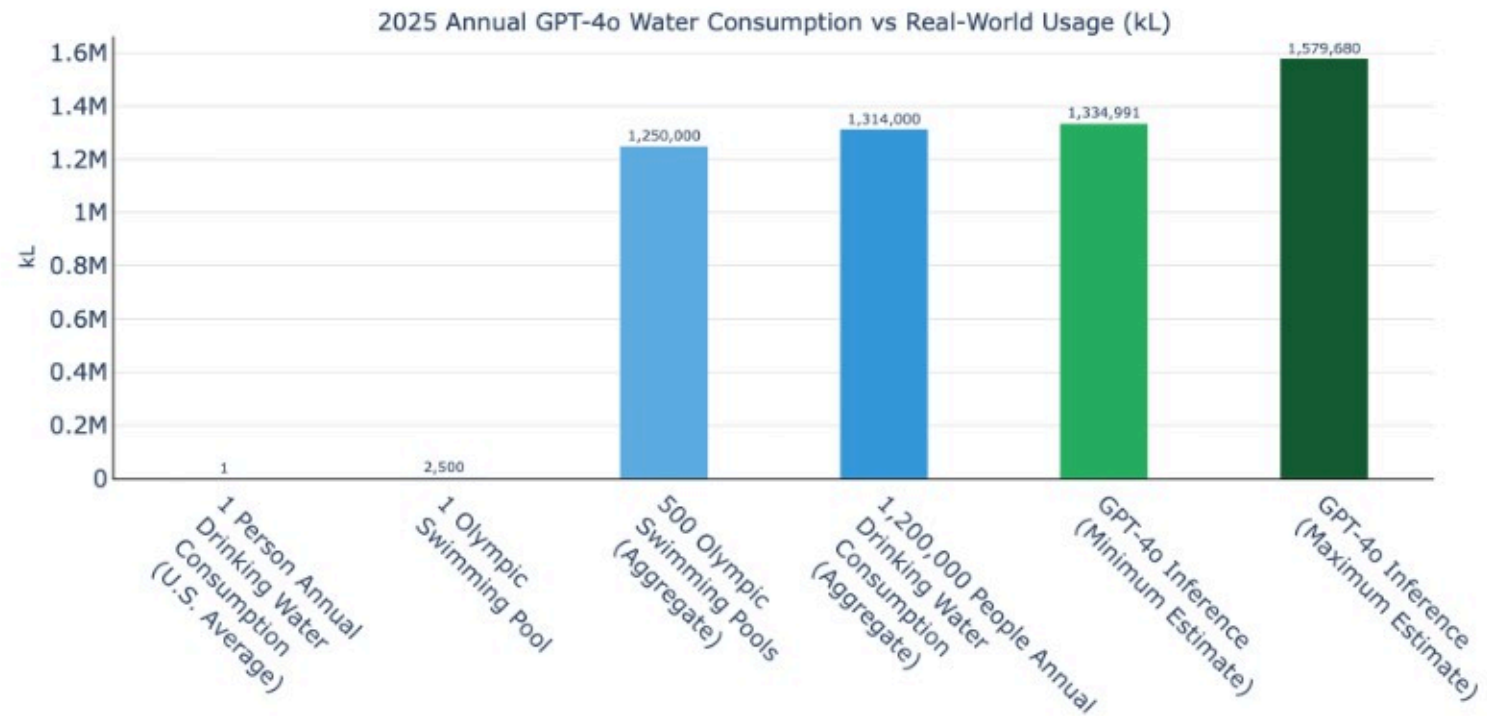
One request to o3 model: 33Wh. One request to GPT-4o: 0.42Wh


Per-Request and Daily Energy Consumption of GPT-4o vs Web Baselines (Wh)
Watt-hours
0
2
4
6
8
10
0.30
0.42
1.21
1.79
3.37
9.71
10.00
1 Google Search
GPT-4o (Mar '25) Short Query
GPT-4o (Mar '25) Medium Query
GPT-4o (Mar '25) Long Query
Daily Session (8 Messages) (Short Queries)
Daily Session (8 Messages) (Medium Queries)
Charging 2 Phones (Energysage)


2025 Annual GPT-4o Energy Consumption vs Real-World Electricity Usage (MWh)
MWh
450k
400k
350k
300k
250k
200k
150k
100k
50k
0
11
1,202
7,631
377,685
381,550
390,650
391,509
463,269
1 Residential Household (U.S. Average)
1 University (U.S. Average)
1 Inpatient Hospital (U.S. Average)
35,000 Residential Households (Aggregate)
50 Inpatient Hospitals (Aggregate)
325 Universities (Aggregate)
GPT-4o Inference (Minimum Estimate)
GPT-4o Inference (Maximum Estimate)


2025 Annual GPT-4o Carbon Emissions Equivalent vs Real-World Equivalents (Tons CO_2e)
Tons CO_2e
160k
140k
120k
100k
80k
60k
40k
20k
0
5
60
100
137,531
138,000
138,000
138,125
163,441

Don’t forget about water !



[2] *How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference*

How these costs add up

As of today, the usage of AI is made *on top* of the other usages of technologies. So it is an additional impact.

e.g. image generation, summarization, code generation

How to measure the environmental impact of LLMs?

Measuring the energy consumption

- Physical wattmeters
- Virtual wattmeters
  - CodeCarbon
  - `perf` + `nvidia-smi`

Measuring the other impacts

Harder than energy, but can be derived from energy usage and hardware usage.

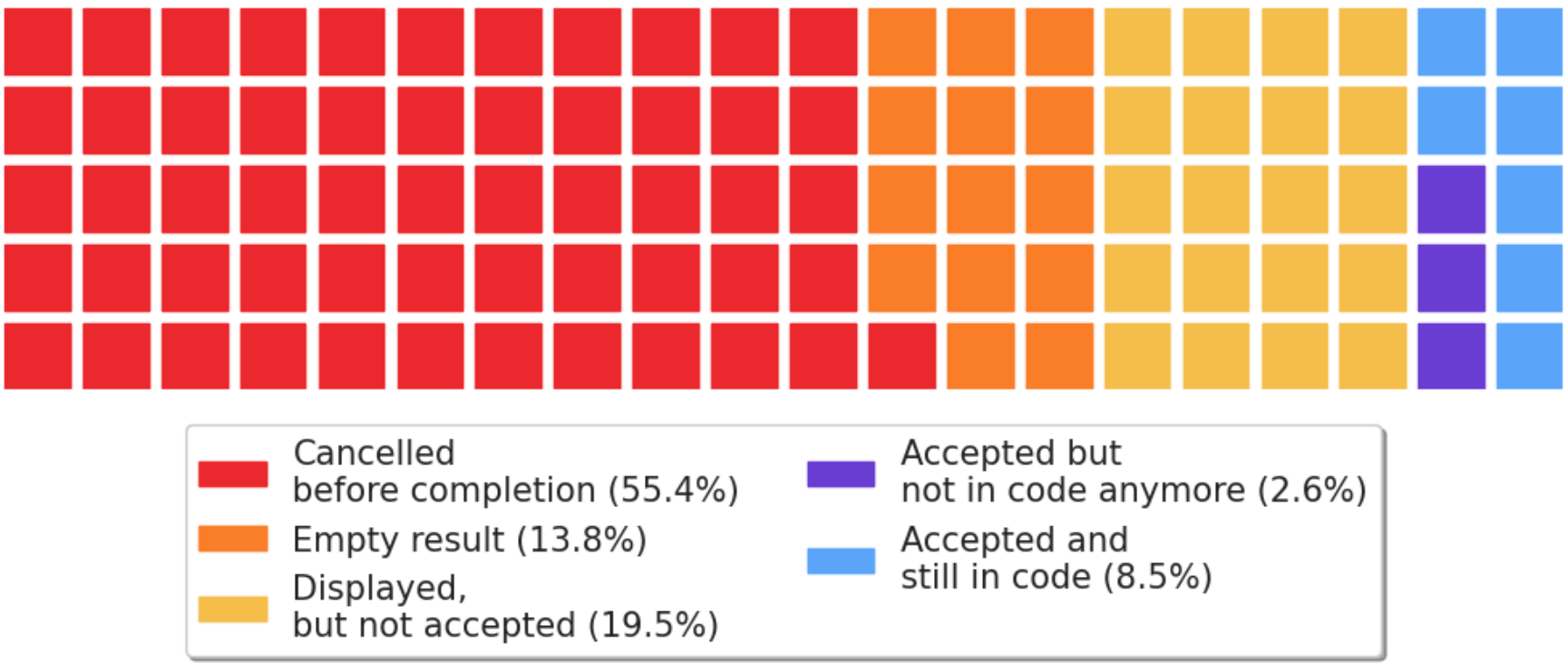
=> Measure your energy consumption!

Empirical findings from my own research

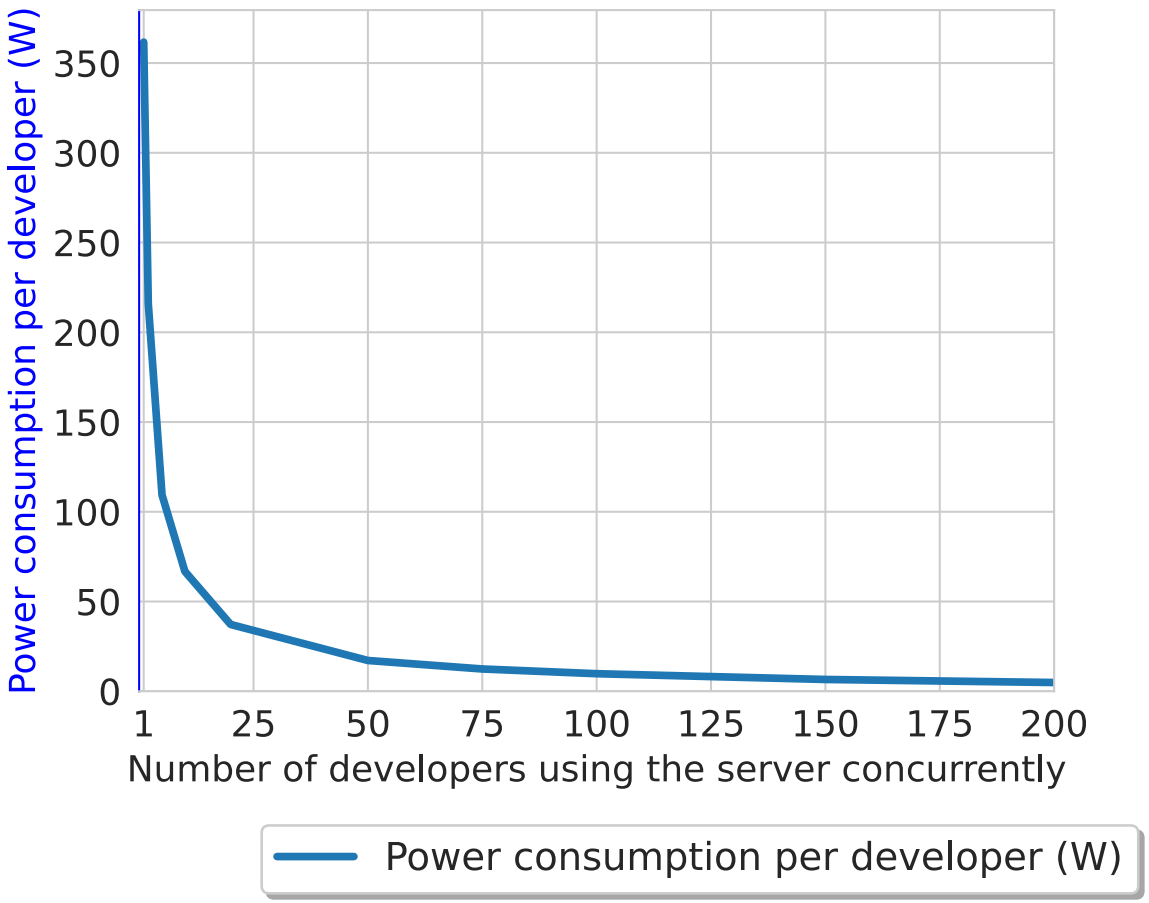
(shameless plug)

How much energy do *you* use when using a code assistant like GitHub Copilot?

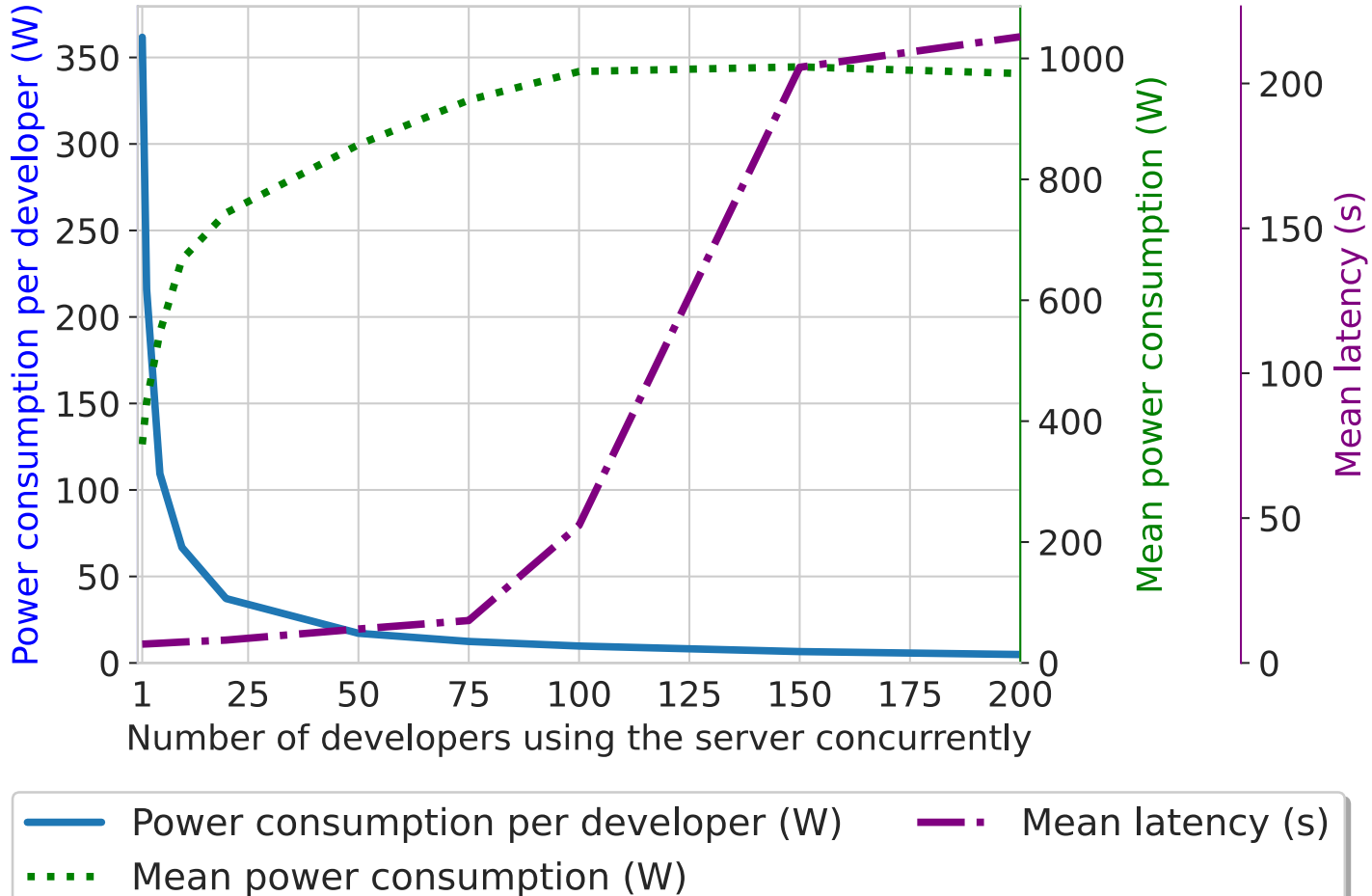
90% of the generations made by GitHub Copilot are useless.



As the number of concurrent developers increase, the energy consumed by each developer decreases…



But, after too many developers are added, the server becomes **saturated** as it cannot handle all the requests



So, how much energy does one actually consume when using GitHub Copilot?

So, how much energy does one actually consume when using GitHub Copilot?

- With a plausible configuration, saturation point was reached at 50 developers
- The server consumed on average 1038W
- So, each developer **consumed about 20W** by using GitHub Copilot

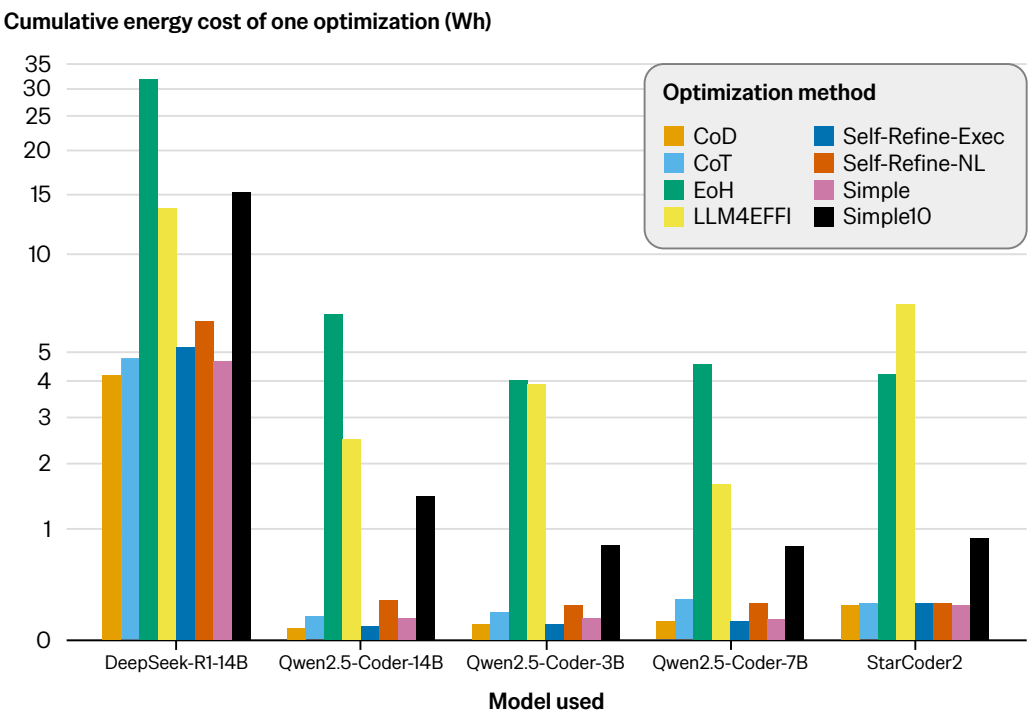
20W ~= 💡💡💡

But… 28.7 millions of developers, and 41% of them are using code assistants (according to the StackOverflow 2024 survey)

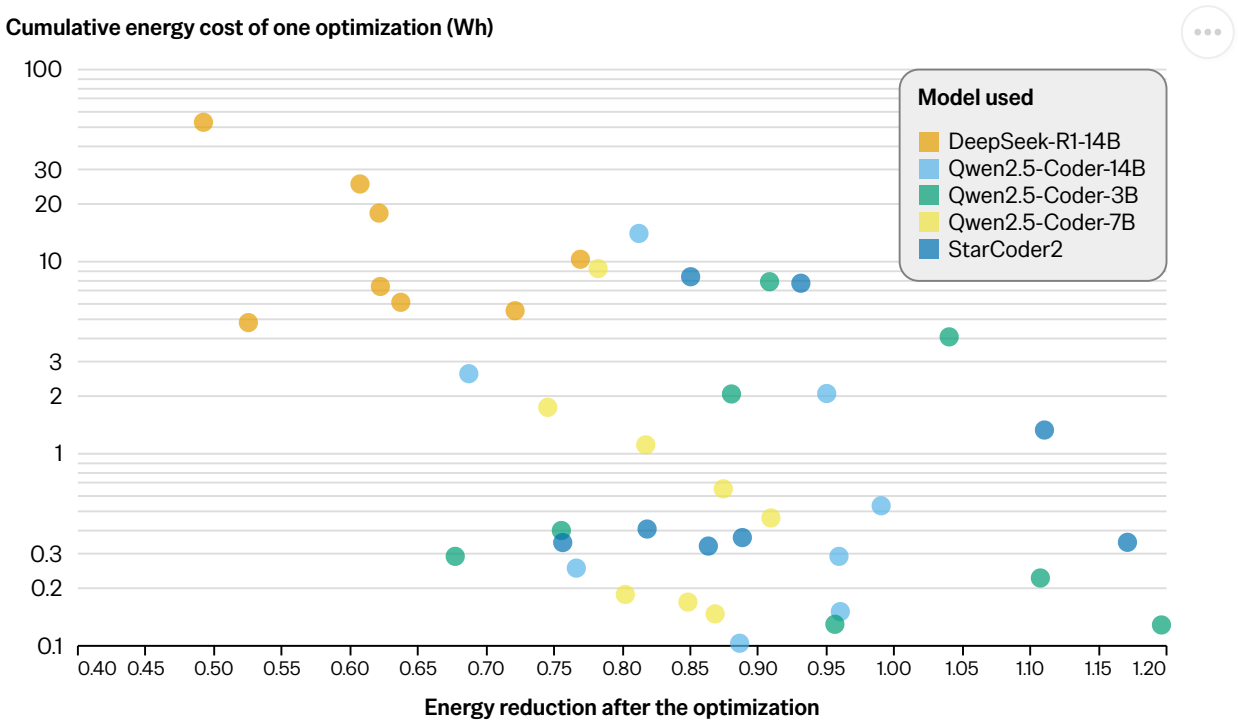
11.77 millions users x 20W = 235 MW

Can you use LLMs to reduce the environmental impact of software?

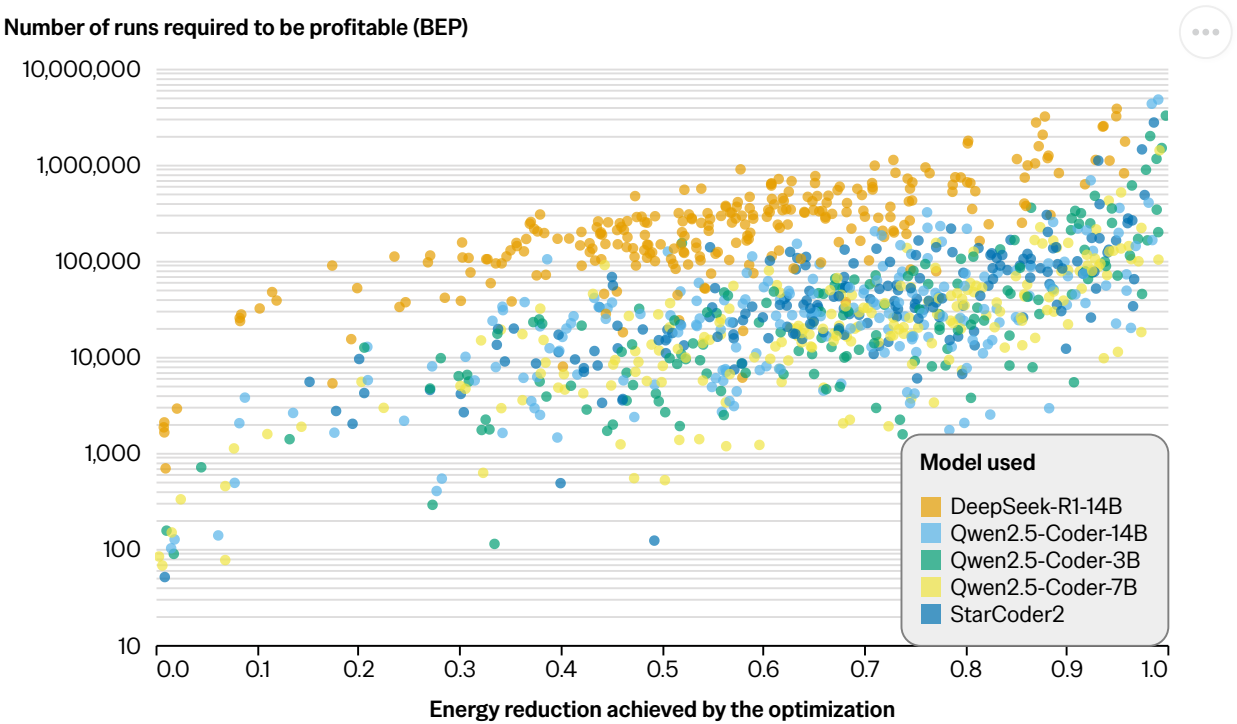
The cost of optimizing is highly dependent on the model and method used.



Configurations that produce better optimization also consume more energy



Most optimizations require a lot of executions of the optimized program to be profitable.



We should be carefull about when and where we use LLM-based code optimization.

When efficiency makes things worse

Jevon’s Paradox

Can AI really help the environment ?

- Efficiency gains lower costs → encourage more usage.
- Same paradox as with fuel-efficient cars.
- Without governance, rebound cancels most energy savings.

[3] From Efficiency Gains to Rebound Effects? The Problem of Jevons' Paradox in AI's Polarized Environmental Debate. Luccioni, Strubell and Crawford

How to be more sustainable in a world where AI is everywhere

(as developers and users)

Do you really need AI ? An introduction to sobriety

The most sustainable AI is often the one you don’t use

You don’t really need to have (generative) AI if you are just a sparkling water company.

Do you really need a 100 B parameters model ?

A 5B parameters model can usually do what you want it to do 95% as good as a 350B parameters model, for 10% of the cost.

Use specialized models!

Deployment

Measure your impact and adapt your usage

Only by knowing how much energy you consume can you find the levers to pull to reduce your impact.

e.g. Integrate energy & carbon metrics into CI/CD

More tips than you can dream of

https://ecoresponsable.numerique.gouv.fr/publications/general-ecoconception/

How to be more sustainable in a world where AI is everywhere

(as users)

TW: Self-harm, mental health issues

Be more mindful about your impact when using AI

- AI as a search engine
- AI before thinking
- AI “friend” - Also serious mental health risks: See AI Psychosis, or chatbot suicide.

[4] https://www.forbes.com/sites/bryanrobinson/2025/11/02/ai-psychosis-at-work-mental-health-experts-express-concerns/

Sobriety and comfort

Can you accept trading your 99% levels of comfort for 95% of comfort, in exchange for lower costs and like, half the environmental impact?

It’s not about not using AI

It’s about not using it *everywhere* in our life and work, if we don’t want to destroy our mental health and our planet.

Thank you for listening !!

(And I hope, thank you for the discussions)

Any questions?