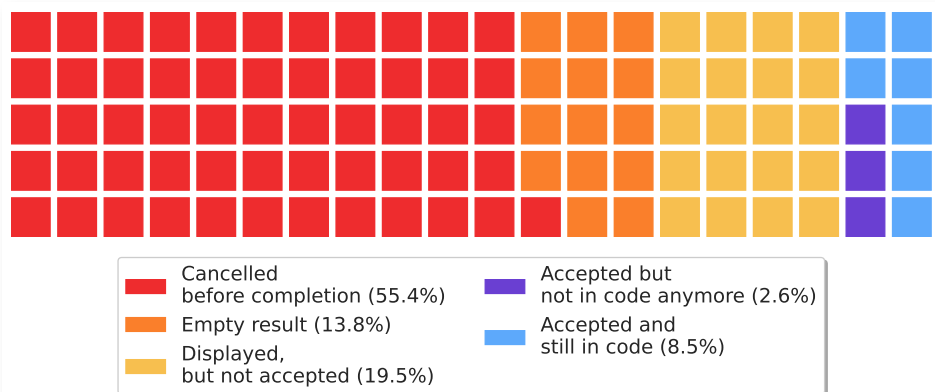
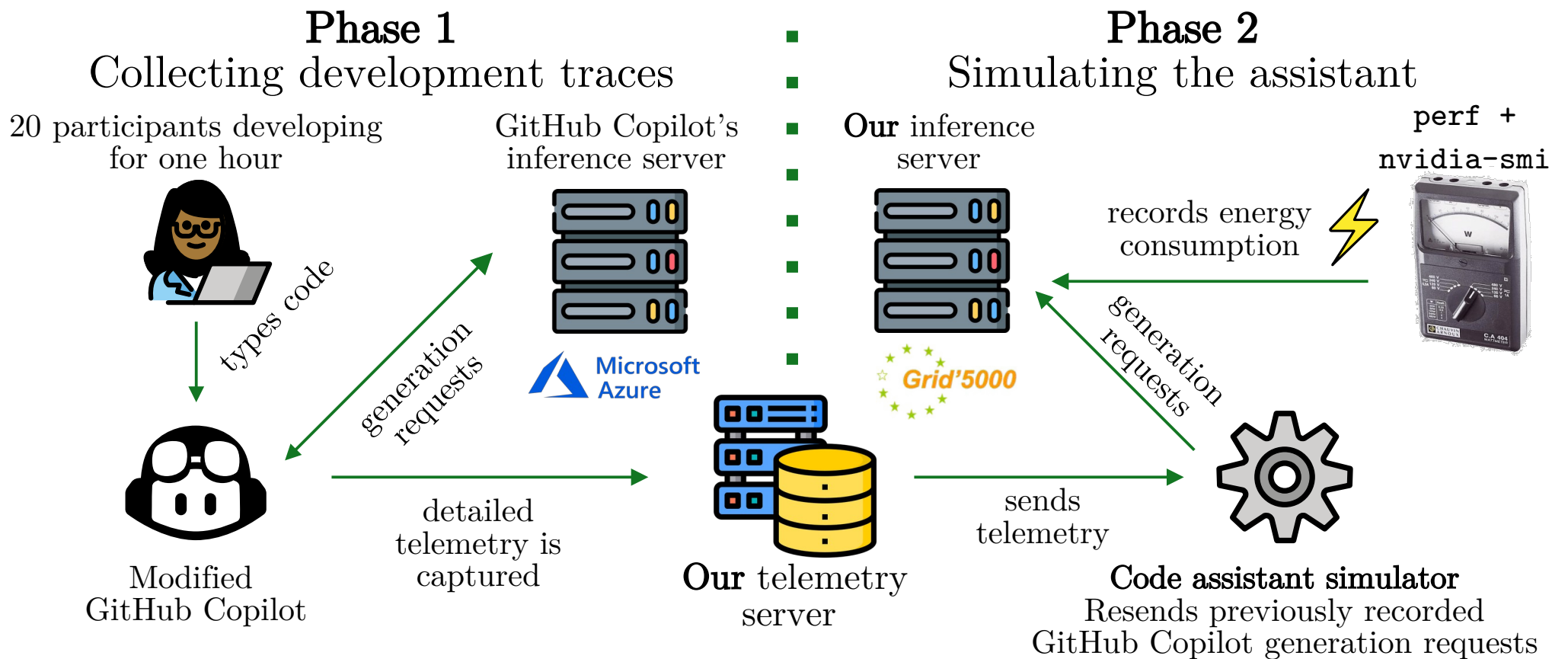




How to measure the energy consumption of code assistants?

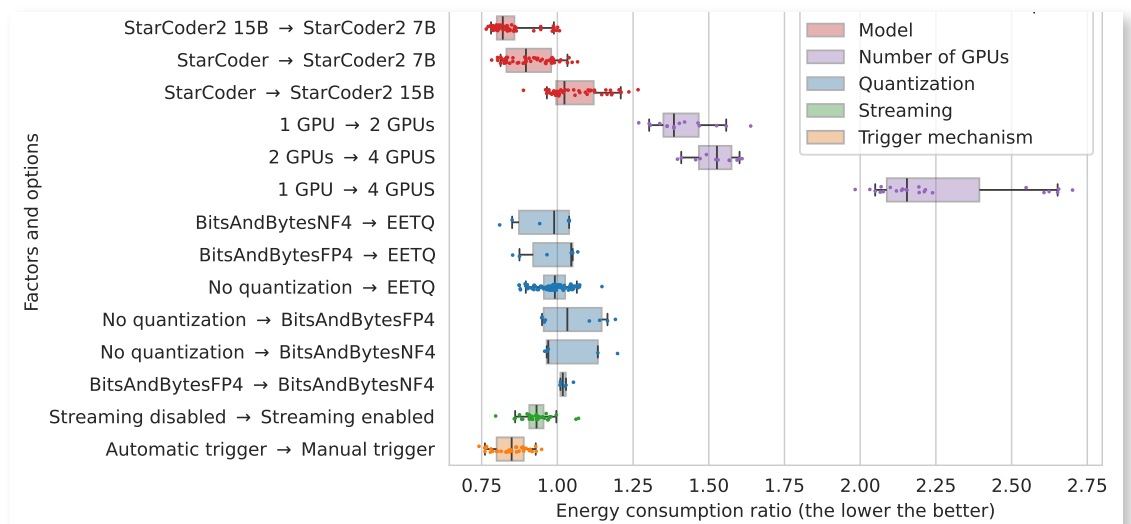
How much energy is consumed when you use code assistants like GitHub Copilot?

How to reduce the energy consumption of code assistants?



End state of generation requests sent by GitHub Copilot

A lot of energy is wasted!



Energy impact ratio from switching from one option to another.

Number of developers using a code assistant hosted on one server	5	20	Maximum until server saturation (e.g. GitHub Copilot)
Average server power	249.3 – 633.1 W	363.2 – 898.8 W	858.9 – 1038 W
Average energy per developer	49.9 – 126.6 W	18.2 – 44.9 W	11.4 – 20.8 W
Energy per 1000 generation requests	12.0 – 30.7 Wh	0.9 – 2.2 Wh	0.1 – 0.4 Wh
CO2 emission per hour per developer	2.8 – 7.1 g	1.0 – 2.6 g	0.6 – 1.2 g
Average latency	6.4 – 1.2 s	7.7 – 1.7 s	16.5 – 2.3 s

Energy usage of a code assistant. Numbers on the left and right represent respectively the most energy-efficient and the fastest configurations

x 11.77 millions of users of code assistants

= 235 MW

= 1/4 of a



What **you** can do about it if you are a...

provider

- Use more efficient models
- Maximize the usage of inference servers
- Allow for easy manual triggering of suggestions

user

- **Don't use AI by default**
- Use AI more mindfully
- Disable automatic suggestions
- Be conscious about your impact

researcher

- Measure the impact of AI
- **Talk about it around you**

Find our article online!

