

What does AI emissions data look like in 2026?



Julia Lozano
GreenOps Specialist



Ross Hunter
Data Scientist

Agenda

- The entropy of AI
- Data to fight it
- Live: the footprint of one Claude Code prompt
- Why it matters



The entropy of AI

- AI is a black box
- Enterprises can barely see what it costs - let alone what it emits
- Where is this running? On what hardware? At what PUE?
- Every provider exports usage differently- some give you more details than others.
- No shared standard for turning that into an emissions number. How to take a consistent approach to estimating emissions?
- How do we honestly tell concerned users their impact is being taken seriously?

GPX AI:

Data that can fight this entropy

AI Explorer

663 of 663 rows · meth-v5.0.0

Columns

Copy rows

JSON

CSV

Table

JSON

Energy, carbon and water per 1,000 tokens for every model the Greenpixie AI Explorer covers.

MODEL

Search models...

PROVIDERS

☐ AI21 Labs

☐ Alibaba

☐ Anthropic

☐ AWS

☐ Cohere

☐ DeepSeek

☐ Google

☐ Liquid AI

☐ Meta

☐ Microsoft

☐ MiniMax

BATCH SIZE

Medium

How densely requests are batched on the serving hardware — higher batching spreads the energy across more tokens.

CODE

cURL

Python

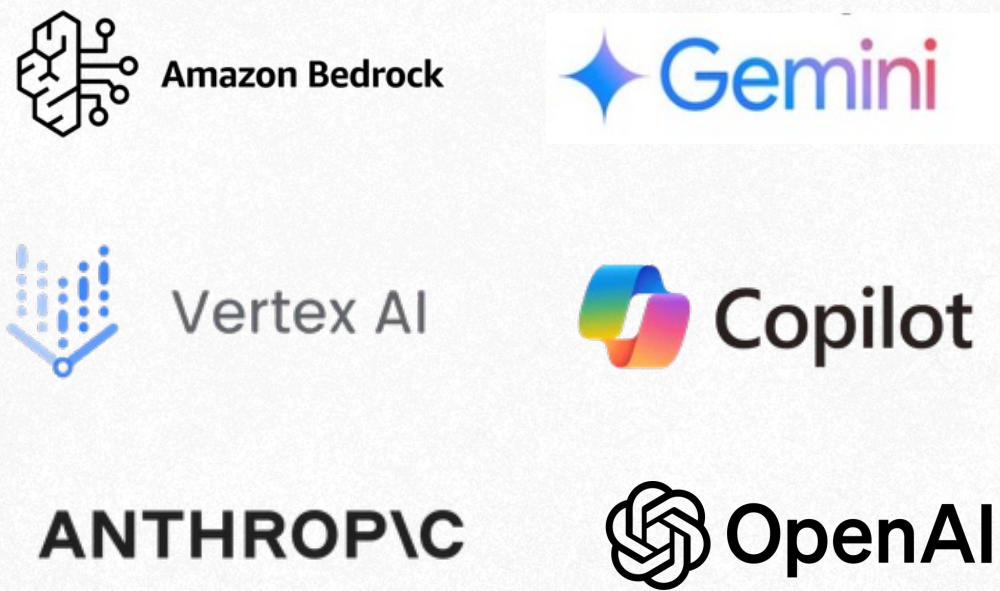
Copy

```
curl -G "https://api.greenpixie.ai/v1/models" \
  --data-urlencode "effective_batch_size=100" \
  --data-urlencode "limit=100" \
  -H "Authorization: Bearer a..."
```

This request returns every model — the provider and model-search filters narrow the results in your browser, not in the API call.

Provider	Model	Token Type	Effective B...	Energy	Carbon	Water	
☐	AI21 Labs	Jamba 1.5 Large	cached_input	medium	4.553e-6	2.241e-9	2.625e-5
☐	AI21 Labs	Jamba 1.5 Large	input	medium	4.553e-5	2.241e-8	0.00026247
☐	AI21 Labs	Jamba 1.5 Large	output	medium	0.00096007	4.713e-7	0.0055341
☐	AI21 Labs	Jamba 1.5 Mini	cached_input	medium	6.076e-7	2.952e-10	3.502e-6
☐	AI21 Labs	Jamba 1.5 Mini	input	medium	6.076e-6	2.952e-9	3.502e-5
☐	AI21 Labs	Jamba 1.5 Mini	output	medium	0.00013106	6.332e-8	0.00075547
☐	Alibaba	Qwen3 235B 2...	cached_input	medium	1.899e-6	9.592e-10	1.095e-5
☐	Alibaba	Qwen3 235B 2...	input	medium	1.899e-5	9.592e-9	0.00010946
☐	Alibaba	Qwen3 235B 2...	output	medium	0.00046765	2.323e-7	0.0026957
☐	Alibaba	Qwen3 235B A...	cached_input	medium	1.899e-6	9.592e-10	1.095e-5
☐	Alibaba	Qwen3 235B A...	input	medium	1.899e-5	9.592e-9	0.00010946
☐	Alibaba	Qwen3 235B A...	output	medium	0.00046765	2.323e-7	0.0026957
☐	Alibaba	Qwen3 235B A...	cached_input	medium	1.899e-6	9.592e-10	1.095e-5
☐	Alibaba	Qwen3 235B A...	input	medium	1.899e-5	9.592e-9	0.00010946
☐	Alibaba	Qwen3 235B A...	output	medium	0.00046765	2.323e-7	0.0026957
☐	Alibaba	Qwen3 32B	cached_input	medium	1.030e-6	4.957e-10	5.940e-6
☐	Alibaba	Qwen3 32B	input	medium	1.030e-5	4.957e-9	5.940e-5
☐	Alibaba	Qwen3 32B	output	medium	0.00016834	8.115e-8	0.00097035
☐	Alibaba	Qwen3 Coder ...	cached_input	medium	2.833e-7	1.417e-10	1.633e-6
☐	Alibaba	Qwen3 Coder ...	input	medium	2.833e-6	1.417e-9	1.633e-5
☐	Alibaba	Qwen3 Coder ...	output	medium	7.048e-5	3.463e-8	0.00040626
☐	Alibaba	Qwen3 Coder ...	cached_input	medium	3.312e-6	1.631e-9	1.909e-5
☐	Alibaba	Qwen3 Coder ...	input	medium	3.312e-5	1.631e-8	0.00019089
☐	Alibaba	Qwen3 Coder ...	output	medium	0.00086322	4.205e-7	0.00019089
☐	Alibaba	Qwen3 Coder ...	cached_input	medium	3.312e-6	1.631e-9	1.909e-5

→ We built this core data set to be applied to any and all of the different usage reports from across AI providers, SaaS tool or agent.



AI energy consumption is a moving target, requiring constant research and development.

That's our job!



Joshua Horswill
PhD



Ross Hunter
PhD



Matt Clifford
PhD



James Hall



Rokšana Malinowska-
Traczyk

[Read our methodology for supporting AI models](#)



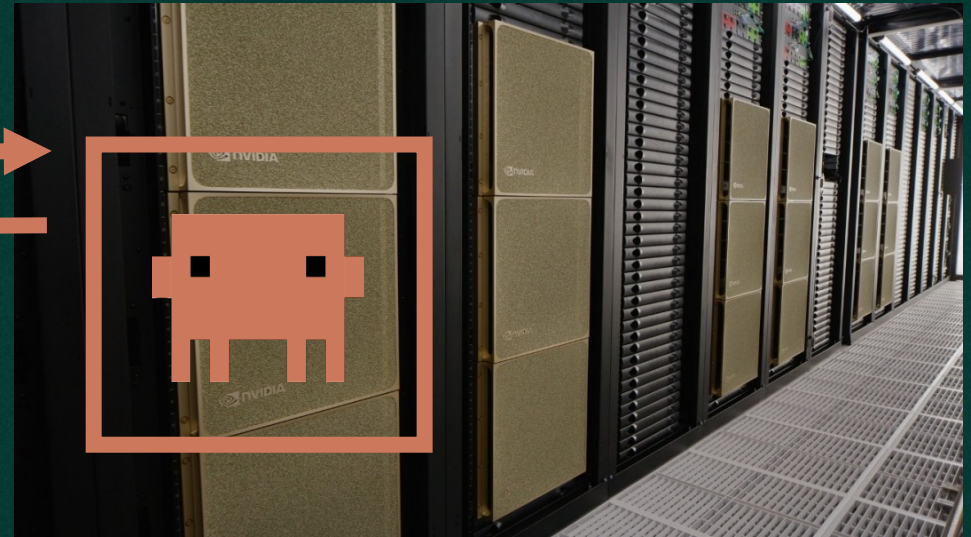
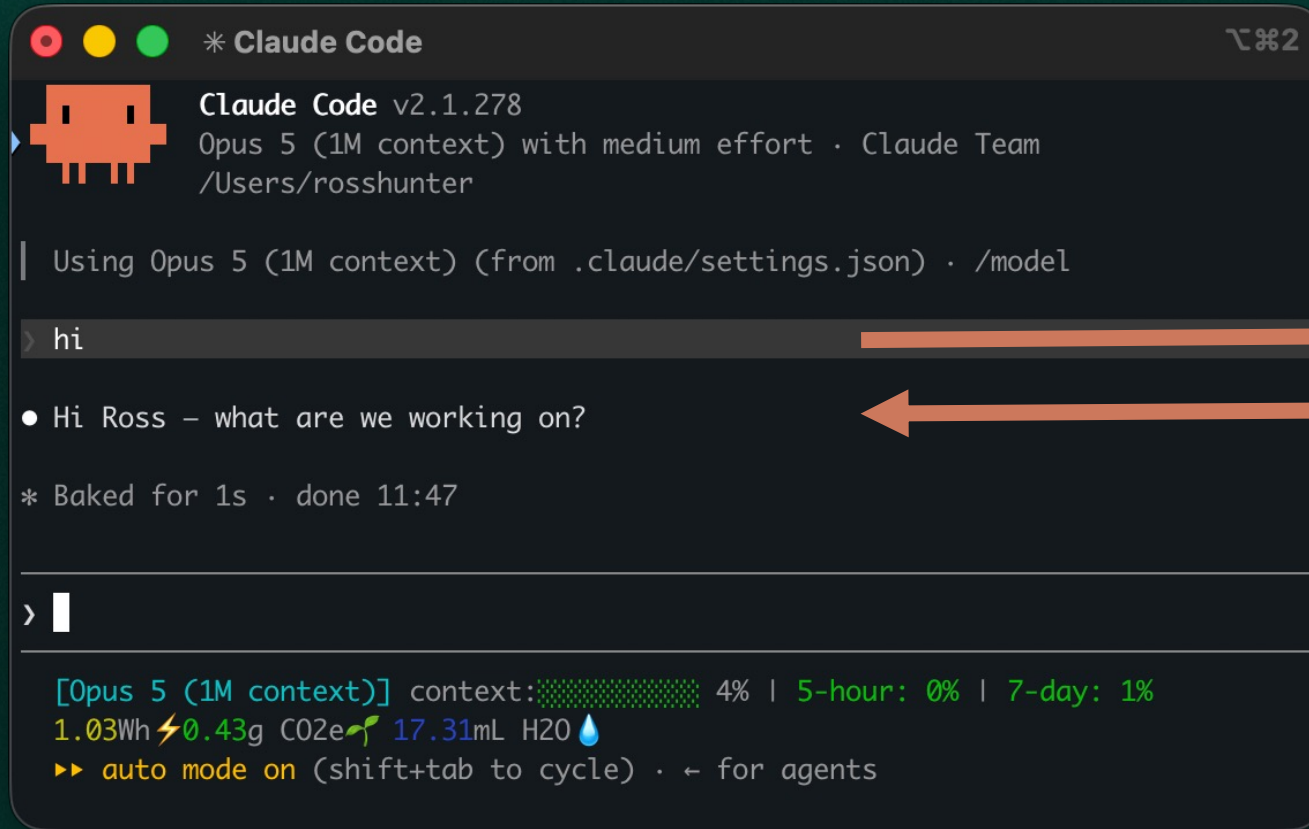


An engineer's example with

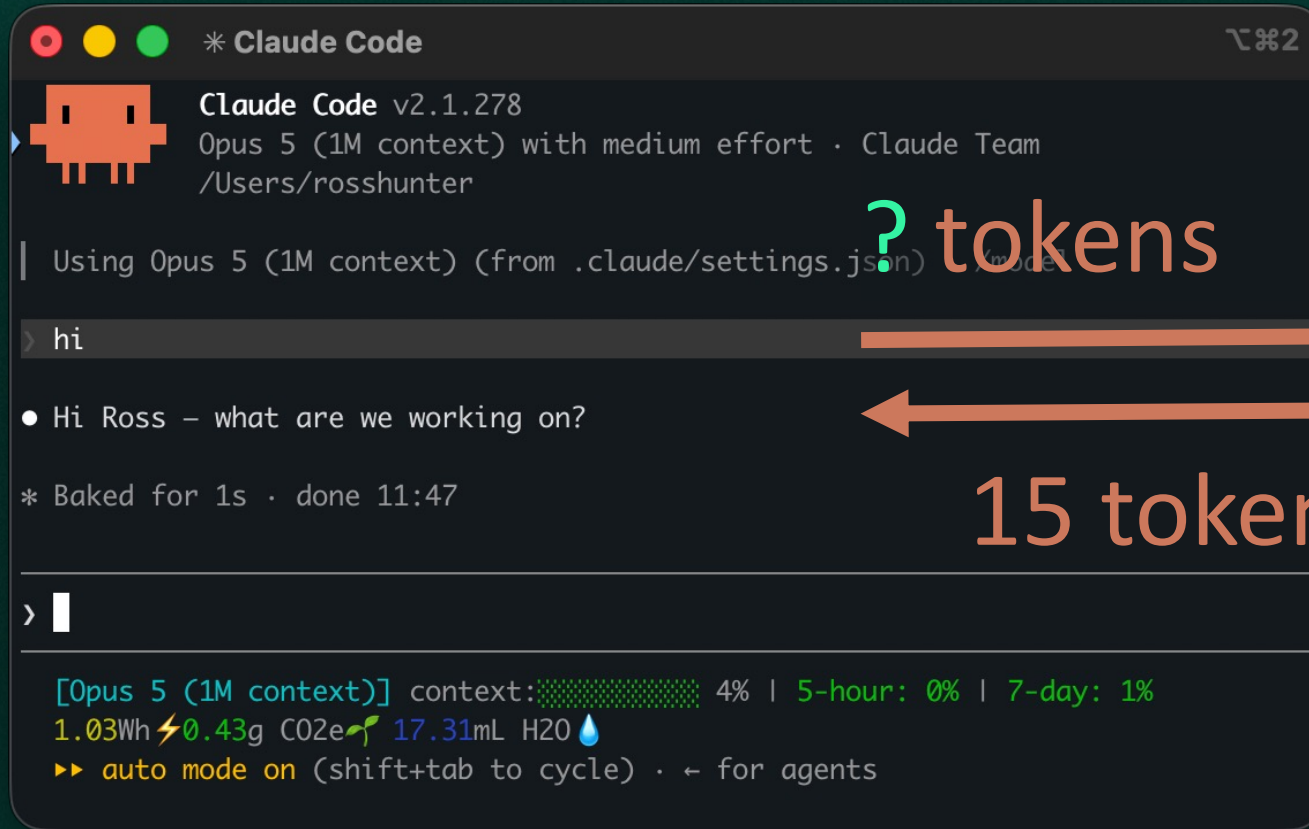




Measuring Claude 's emissions impact

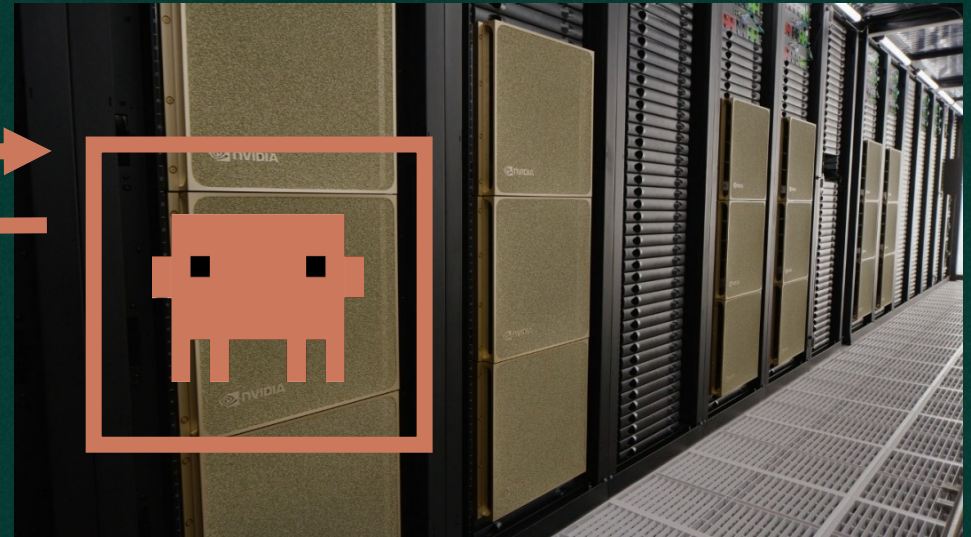


~500x more power hungry than a laptop



? tokens

15 tokens

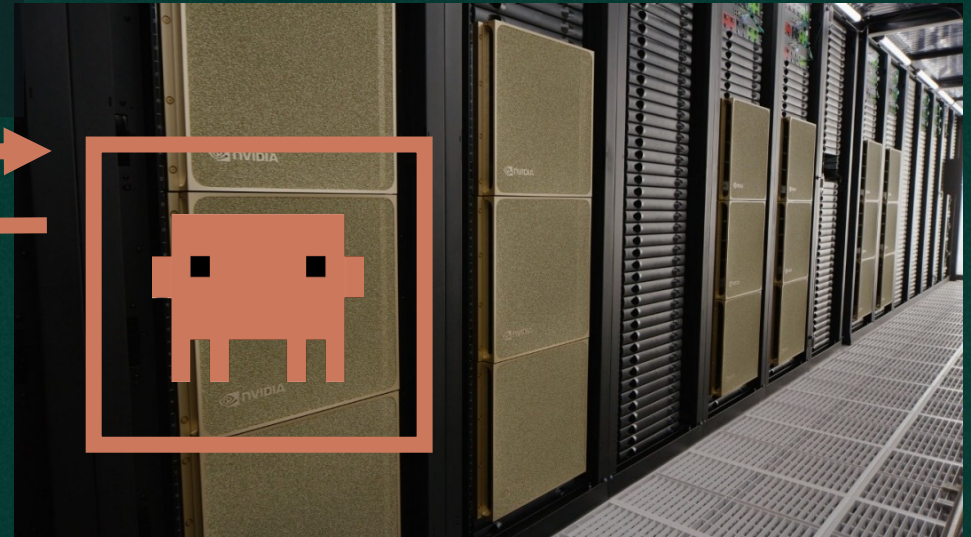


How many input tokens?

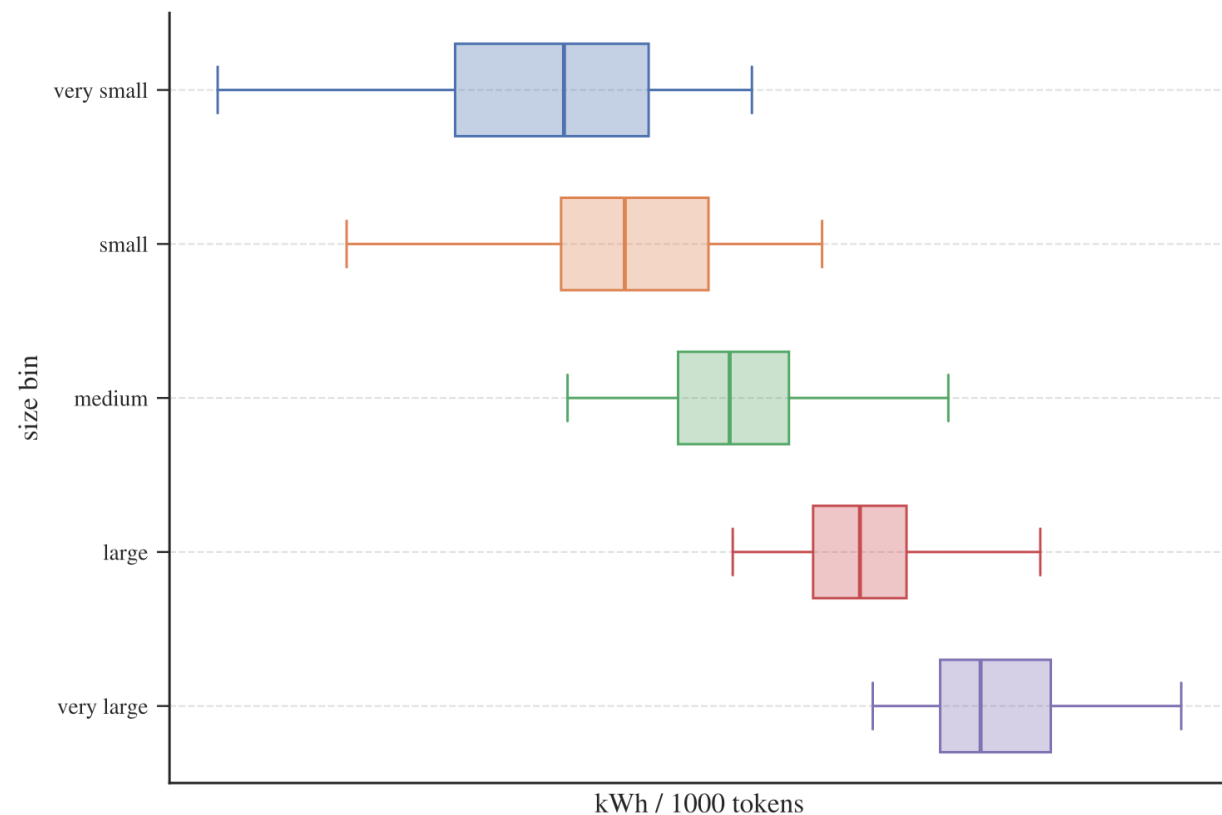
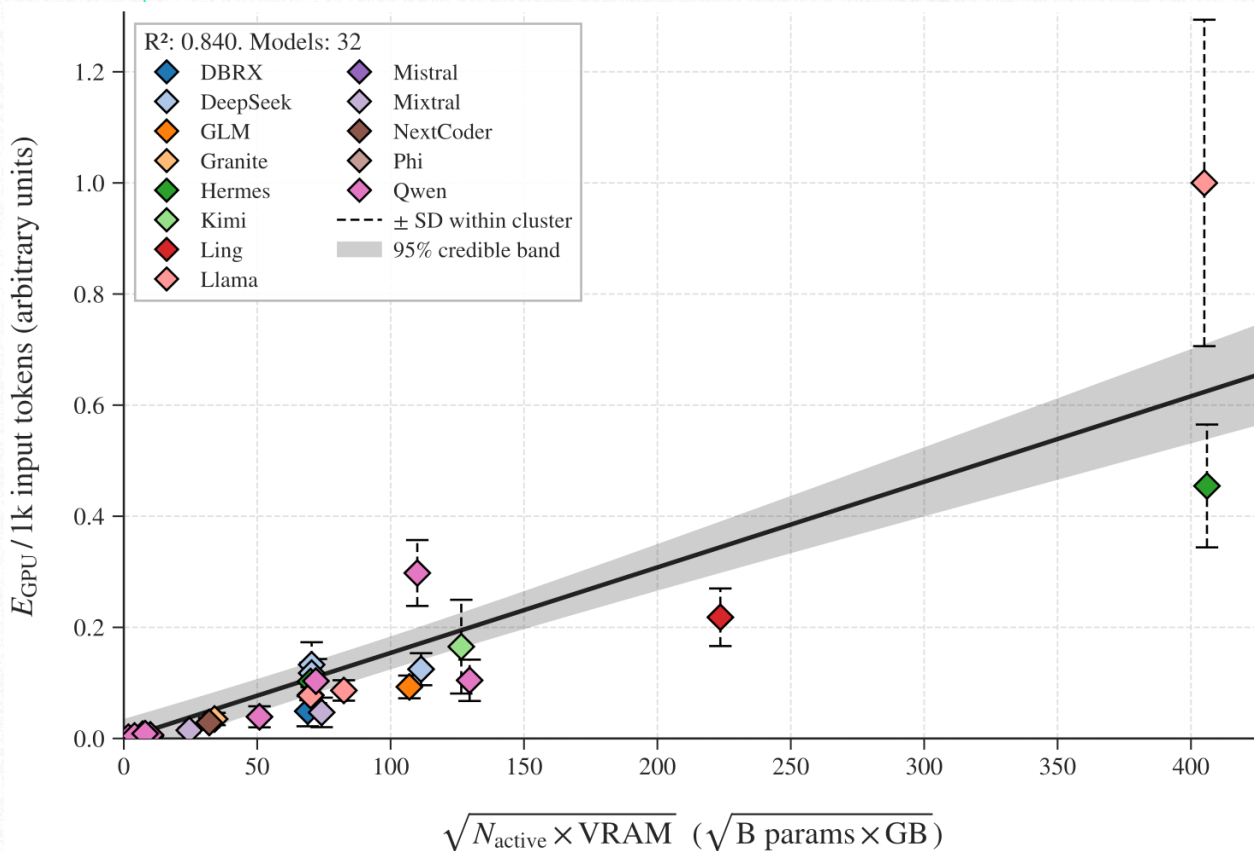


40K tokens

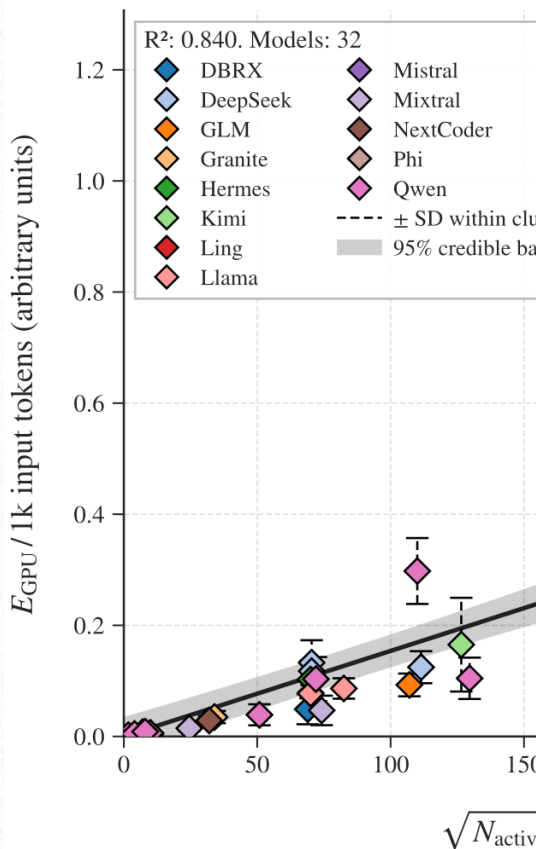
15 tokens



... lots of hidden token usage = hidden cost & hidden emissions



Experimental results = E/token for 250 commercial models & cloud SKUs



Greenpixie's AI Token Methodology

Assessing the Energy, Water and CO₂-eq Impact of AI Tokens for Open and Closed Weight Models

Joshua Horswill, Ross Hunter, Matt Clifford, James Hall

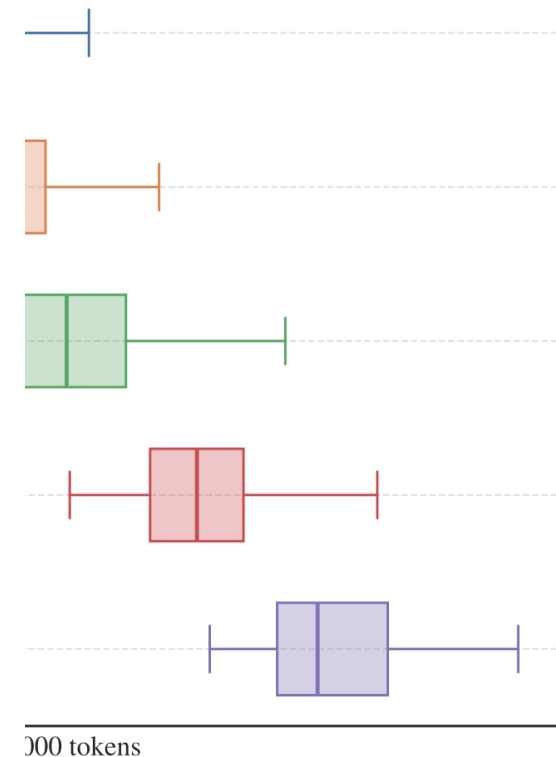
Greenpixie Ltd., London, UK.



© 2026 Greenpixie Ltd. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence (CC BY-NC-ND 4.0), <https://creativecommons.org/licenses/by-nc-nd/4.0/>. Commercial use is not permitted.

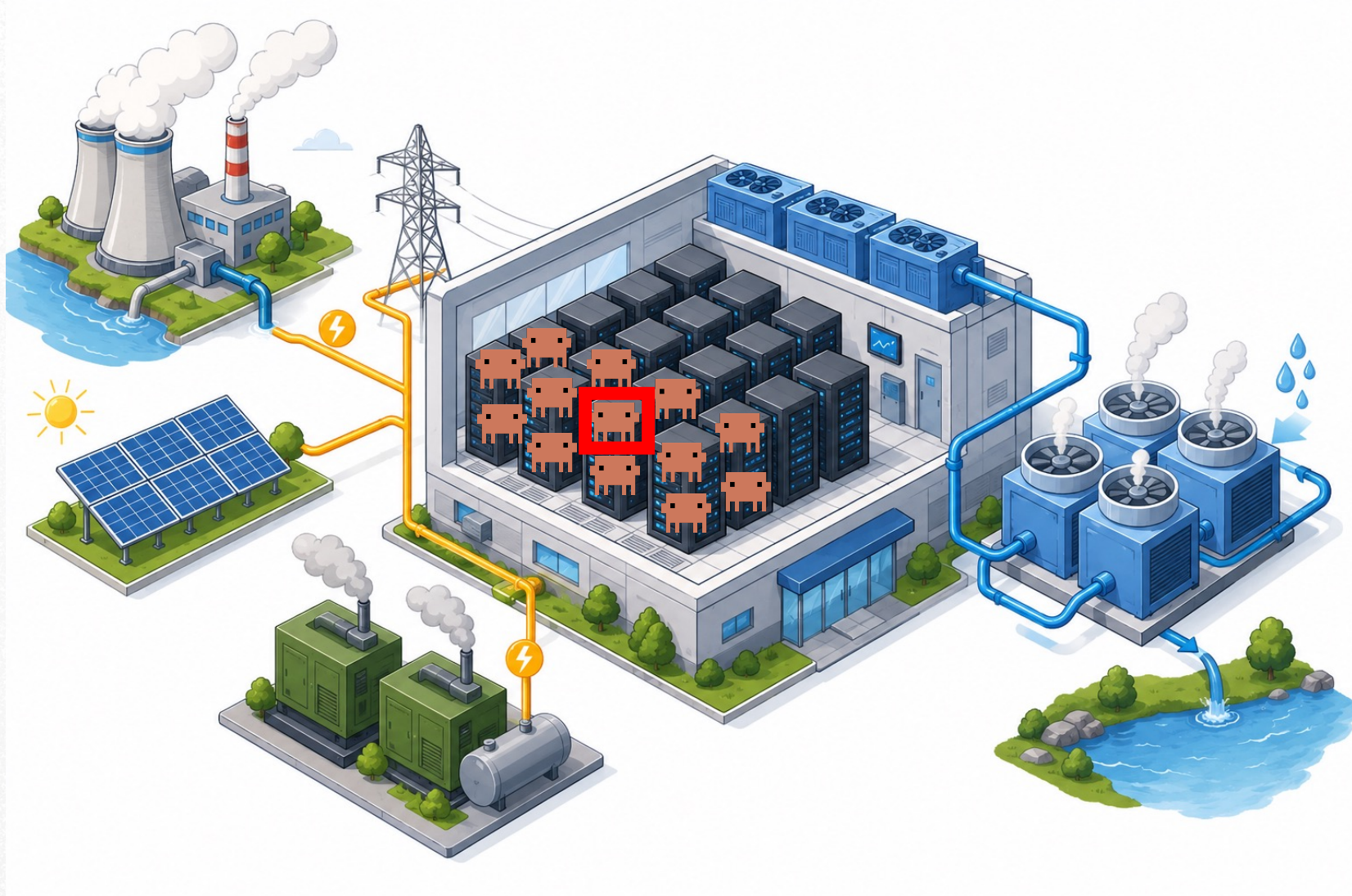
Abstract

We describe a methodology for estimating the per-token energy cost of cloud-hosted large language model (LLM) inference which separates between input tokens (prefill) and output tokens (decode). Graphics processing unit (GPU) energy is measured during inference benchmarking on a wide range of text-based tasks, open weight models and workloads. The remaining server energy contribution from non-GPU hardware is estimated from the inference wall time. Bayesian linear regression is used to model the relationship between LLM size, request traffic, hardware, and energy per token. Proprietary frontier LLMs are binned into size buckets based on naming conventions and performance priors to estimate a range of electricity values per bin, from which a representative average and uncertainty can be derived. We also describe how these energy measurements can be used to estimate the carbon-dioxide equivalent (CO₂-eq, usage and embodied) emissions and water consumed per



Available soon on arXiv

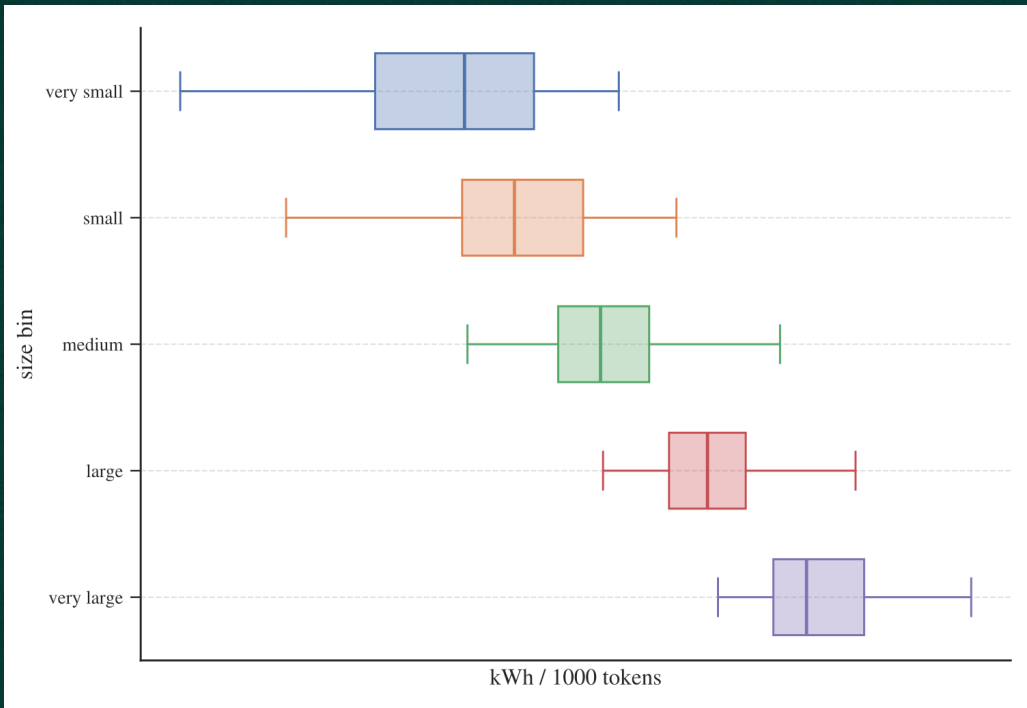
Claude and Consumption



Claude and Consumption



How can you use this?



What you can control:

- Effort level (/effort),
- Work type (/workflows),
- Model size (/model).

GPX AI data translates into energy, carbon & water savings.

2.11Wh ⚡ 0.87g CO2e 🌱 35.38mL H2O 💧

▶▶ auto mode on (shift+tab to cycle) · ← for agents

Back to our prompt



Claude Sonnet 5

10.52g



GREENPIXIE · GREEN IO LONDON 2026

Know the electricity, carbon and water behind every **AI** call

Model choice, region and inference settings change the environmental footprint of an LLM workload by an order of magnitude. This prototype estimates that footprint the way Greenpixie estimates cloud infrastructure impact — bottom-up, transparent about assumptions, and built to sit alongside cost data, not replace it.

[Try the calculator](#)[How it's calculated](#)**3**

Claude models
compared

5

Deployment regions
modelled

7

Leveraged you can tune

EXAMPLE — MID-SIZE SONNET WORKLOAD

Electricity	47.8 kWh
Water	43 L
CO ₂ e	11 kg CO ₂ e
Estimated electricity cost	US\$9

CALCULATOR

Model your own AI workload

Adjust the model, region and inference settings below. Results update instantly and are calculated entirely in your browser using the mock methodology described further down this page.

Claude Haiku 4.5

10.52g



Powered by Greenpixie Intelligence

AI Energy & Sustainability Calculator

Understand the electricity, water, and carbon footprint of your Claude AI inference workloads. Make informed decisions about model selection, deployment regions, and optimisation strategies.

Start Calculating →

Calculate Your Impact

Configuration

Model

Claude 3.5 Haiku

Region

US West (Oregon)

Monthly Input Tokens

100000000

100,000,000 tokens

Monthly Impact & Cost

⚡ Electricity

182,314 kWh

💧 Water

218,777 L

🌍 CO₂e

15,496.7 kg

Claude Opus 5



GREEN IO LONDON · DEMO BUILD

Put an energy number on every AI token you serve.

Your AI programme has a cost per million tokens. It does not yet have an electricity, carbon or water number — so sustainability sits outside the decisions your engineers actually make. This calculator models all four on one basis, at the granularity your FinOps data already uses.

Model a workload

How the model works

DATA COVERAGE

MODELLED

~10m

cloud SKUs with an energy, carbon and water rate

~110k

new SKUs absorbed per month

ISO 14064

independently verified methodology

3

hyperscalers on one comparable basis

Greenpixie models energy consumption from the hardware up, then derives carbon and water. It does not meter servers directly — every figure here carries a stated uncertainty band.

— CALCULATOR

Model a month of inference

Enter the shape of your traffic. Every figure updates as you type — nothing leaves the browser, and no external service is called.

Workload

MONTHLY VOLUMES

Start from:

Customer support copilot

Internal coding agent

Electricity

941 kWh

Includes facility overhead at PUE 1.16

Carbon

186 kg CO₂e

±36% · location-based, operational

What did it cost?



x4

82.3g CO₂e

Claude
Opus 5
(Medium Effort)



x4

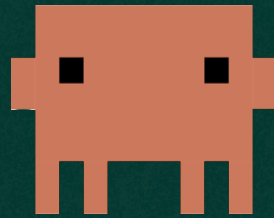
19.3g CO₂e

Claude
Sonnet 5



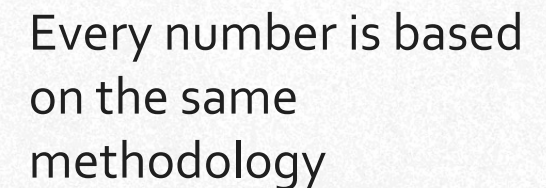
5.4g CO₂e

Claude
Haiku 4.5



Why does this all matter?

How do organisations use our AI data



Every number is based on the same methodology

AI Emissions Management Dashboard

CONNECTED AI PROVIDERS

Anthropic

Google Gemini

Microsoft 365 Copilot

OVERVIEW

The last three months side by side, and six months of the estate broken down however you group it.

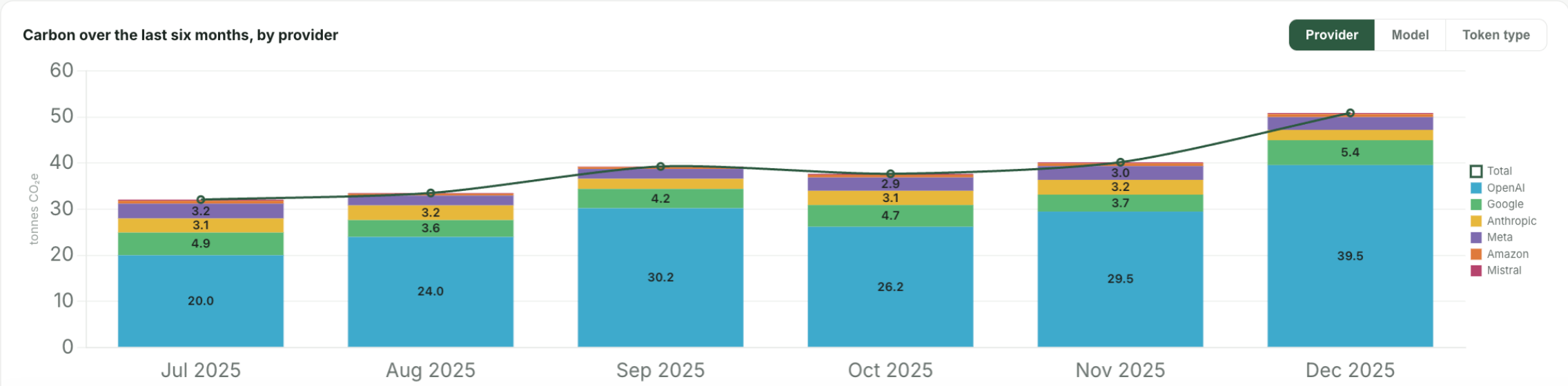
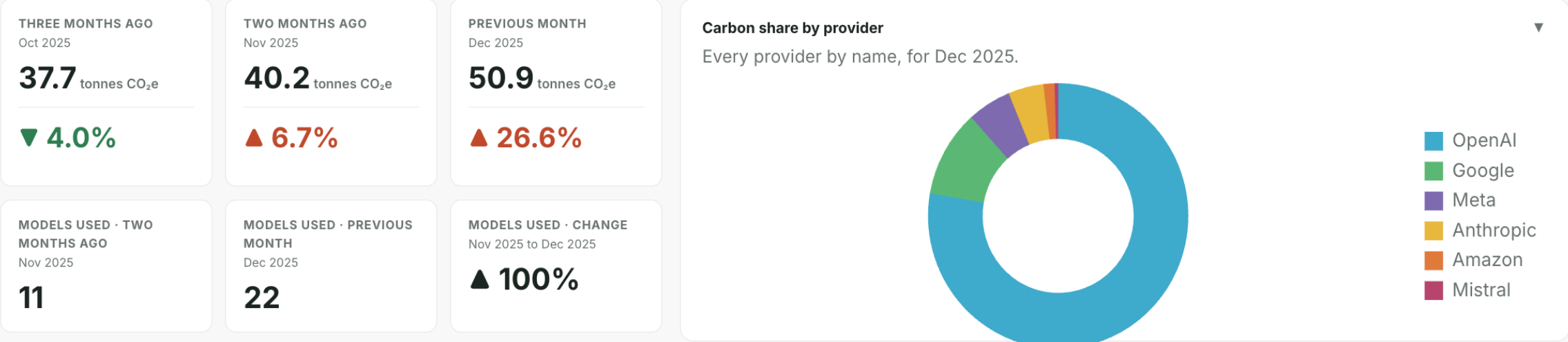
Carbon

Energy

Water

Cost

Tokens





Check out app.greenpixie.com to look at AI emissions data

Greenpixie Data Explorer

BETA

Enhance your technology strategy with sustainability data for cloud and AI.

Start exploring the data for free



Cloud Data Explorer

Browse our comprehensive emissions factors across AWS, Azure and GCP.

Service	Instance ty...	Energy (k...	Carbon (tC...	Water (L)
 EC2	x1.32xlarge	0.91373	0.000633...	3.7781

Explore cloud data →



AI Data Explorer

See the energy and emissions behind every major AI model and provider.

Provider	Model	Energy (k...	Carbon (tC...	Water (L)
 OpenAI	GPT-5.6	0.0015287	6.471e-7	0.0075476

Learn more →

Try our free tools and see the data in context

Unlock "Double-Value" with FinOps Mode

Quantify the carbon savings of a FinOps recommendation and shift-left with our Compute Modeller.



Compare AI emissions

See how the models stack up - every major provider covered by our methodology.



Check our Region Scorecard

See how every cloud region in AWS, Azure and GCP ranks on carbon, water stress, cost and availability.



eu-west3 

Drop by the Greenpixie booth
for more info on GPX AI



← Join us for an evening of Dinner and Drinks!



Tomorrow 18:30



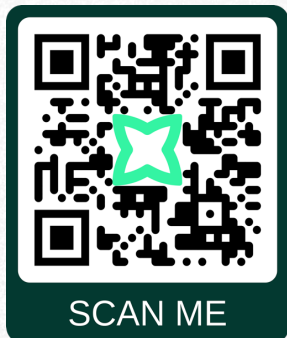
Caravan City



Join our AI Working Group kicking off mid October (3 sessions)

We have gathered different organisations such as OVO Energy, Cambridge University Press, CBA etc...to identify problem areas so we can apply our research to actually useful use cases.

👉 julia.lozano@greenpixie.com



Check out our GreenOps for AI course at the GreenOps Academy, our home-grown e-learning platform.

What did you think of the talk?

