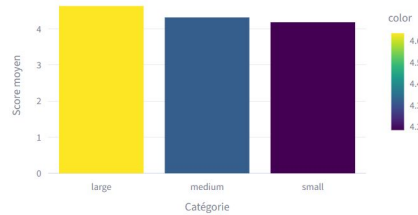
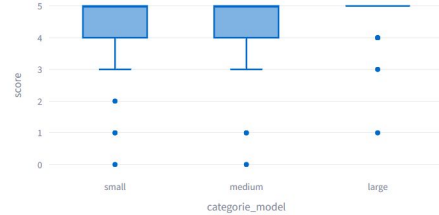


ComparAI: General finding

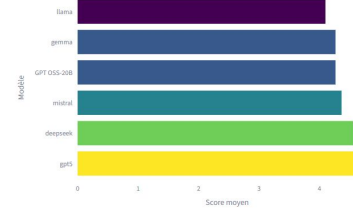
Score Moyen par Catégorie



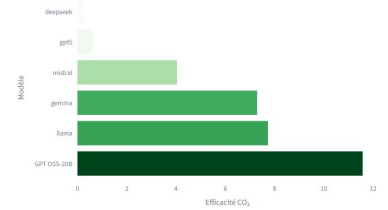
Distribution des Scores par Catégorie



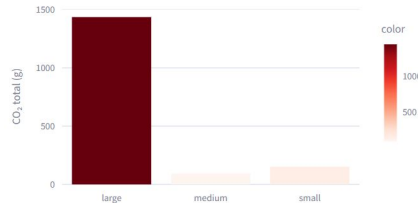
Top 10 - Meilleurs Scores



Top 10 - Efficacité CO₂ (Score/g)



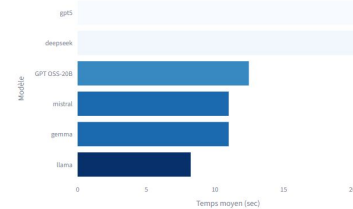
Émissions CO₂ Totales par Catégorie



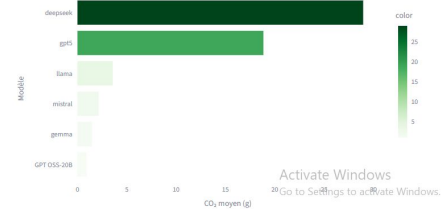
Temps de Réponse Moyen par Catégorie



Top 10 - Modèles les Plus Rapides



Top 10 - Plus Faible Empreinte Carbone



- Performance vs. Efficiency: Larger models provide slightly better average scores but at the cost of much higher carbon emissions and slower speeds.
- Top Models: gpt5 and Gpt OSS-20B offer an appealing compromise—high scores, fastest inference times, and lowest carbon footprints despite their CO₂ efficiency per score being lower.
- Environmental Impact: Deploying large models widely can have a heavy carbon impact, suggesting that, unless peak accuracy is indispensable, medium or small models may offer a better efficiency-to-performance ratio.

ComparAI: Model by case

Best by score

model	categorie model	categorie question	score	energy (wh)	CO2(g)
gemma 3n -4b	small	easy factual	5	1.03	0.64
GPT5	large	reasoning & quantitaiv e	4.2	20.97	12.62
GPT5	large	programm ing & debuggin g	5	26	16
Deepseek	large	Harder knowledg e & reasoning	5	60	36
GPT5	large	advanced	5	36	20

Best by trade-off

model	categorie model	categorie question	score	energy (wh)	CO2(g)
GPT OSS-20B	medium	easy factual	4.9	0.96	0.60
mistral	medium	reasoning & quantitaiv e	4	3.78	2.36
GPT OSS-20B	medium	programm ing & debuggin g	4.8	1.21	0.75
gemma 3n -4b	small	Harder knowledg e & reasoning	3.8	6	3.46
GPT OSS-20B	mediuin	advanced	4.6	1.06	0.65

Note: In the programming & debugging question categorie we notice that the CO2 for “medium model - 1.21 g” is less to “small model - 7.10 g” but the electricity consumption of “medium model” is inferior to “small model” so we can hypothesis that medium model are more train in this categorie than small one so their output token is less

Final recommendation

With all analysis we can conclude that using a medium model is the best trade-off for most of the case if you care about electricity consumption and the quantity of CO2 liberated.

My preferred model is “GPT OSS-20B”

github: <https://github.com/gyxcit/greenai-tp1>

streamlit dashboard: <https://gyxcit-greenai-tp1-app-tffke0.streamlit.app/>

spreadsheet:

https://docs.google.com/spreadsheets/d/1_v9mUKppzWSKCKMGYIYK-hUQYwcdhp3-cBZVgO3ZlYw/edit?gid=319276651#gid=319276651