

Estimating the environmental impact of agentic-AI workflows

Djoser SIMEU¹
Danilo CARASTAN DOS SANTOS¹
Laurent LEFÈVRE²
Denis TRYSTRAM¹

¹Université Grenoble Alpes, Grenoble INP, Inria, LIG, France

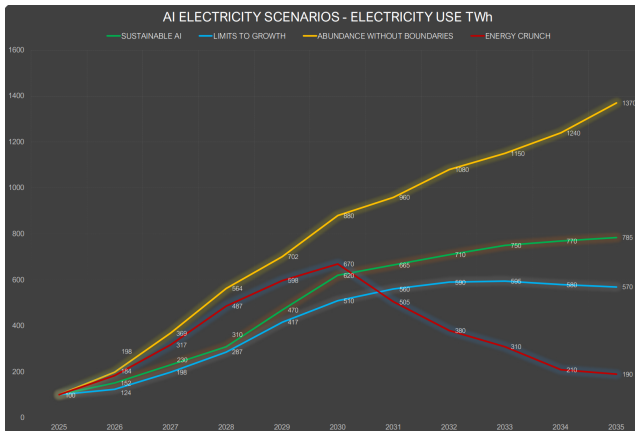
²École Normale Supérieure de Lyon, Inria, LIP, France

9 mars 2026



Evolution of the energy consumption of the AI sector

2025 AI consumption : 100 TWh

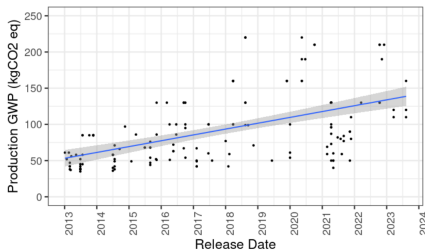


| Scenario | 2030 | 2035 |
|------------------------------|-------|--------|
| Sustainable AI | +520% | +685% |
| Limits to Growth | +410% | +470% |
| Abundance without boundaries | +780% | +1270% |
| AI Energy Crisis | +570% | +90% |

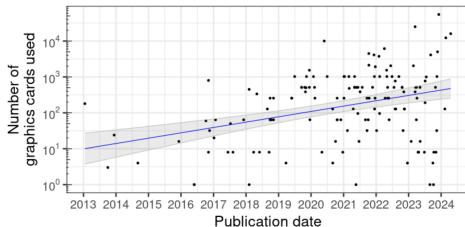
Source : Artificial Intelligence and Electricity, A System Dynamics Approach, Schneider Electric

Evolution of AI specialized hardware impacts

Evolution of the manufacturing environmental impact of NVIDIA Workstations, Morand et al. [2024]



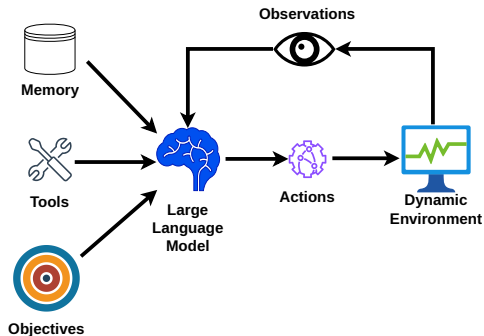
Evolution of the number of graphics cards used for training ML models, Morand et al. [2024]



What is Agentic-AI ?

Definition

Agentic AI includes the class of **autonomous AI** systems that undertake to finish a set of **complex tasks** that span over long periods of time **without human supervision**. Acharya et al. [2025]



Case study : Software development Agentic-AI workflows

Agentic pipeline : GitHub Issue solving

Clone Repo

Read Issue

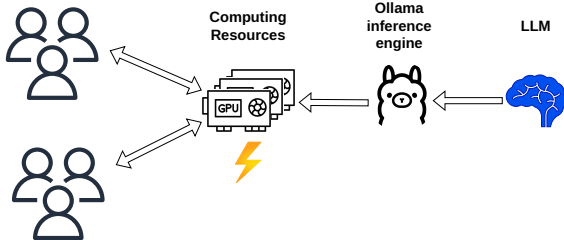
Edit Code

Test Code

Commit Code

Push Changes

N users



Research question

What is the relationship between the characteristics of Agentic-AI deployment and its estimated environmental impact ?

Experimental settings

Number of users considered

1, 10, 50, 100

Considered Github Issues

ADD

- A python code that add two variable ($a + b$)
- **Error** : The code use a wrong name for a variable (c instead of b)
- High level of detail in the issue

MATMUL

- A python code that multiply two matrices (AB)
- **Error** : the code use wrong indices for the multiplication ($C[i][j] += A[k][i] \times B[j][k]$ instead of $C[i][j] += A[i][k] \times B[k][j]$)
- Low level of detail in the issue

Computing resources : Grid'5000 HPC nodes

| Node | Nb GPUs | GPU | Memory | TDP |
|----------|---------|------|--------|-------|
| Hydra | 1 | H100 | 96 GB | 700W |
| Chuc | 4 | A100 | 160 GB | 1000W |
| Grouille | 2 | A100 | 80 GB | 500W |

Ollama LLMs

- Gemma3 12B
- Qwen3 14B
- Gemma3 27B
- Qwen3 30B

Estimating the environmental impact of Agentic-AI inference process

Method, Berthelot et al. [2024]

$$I = \underbrace{a_{GPU}(t) \times F_{GPU}}_{\text{Manufacture}} + \underbrace{C_{i,GPU} \times EGM_g \times PUE}_{\text{Usage}}$$

- $C_{i,GPU}$: electricity consumption of performing the inference i on the hardware GPU in kWh .
- EGM_g : electricity grid mix impact in geographic area g .
- PUE : power usage effectiveness of the hardware used.
- $a_{GPU}(t)$: time-based allocation of the hardware GPU during t time units.
- F_{GPU} : manufacture footprint of the hardware GPU in $kg CO_2 eq$, Falk et al. [2025].

Results : Interpretation models

Quality of User Satisfaction (QoUS) scores

$$QoUS = 10 \times R_{Success} \times R_{response\ time}$$

$$R_{response\ time} = \begin{cases} 1 & \text{if } t \leq 60 \\ \frac{1}{1 + \log_{10}\left(\frac{t}{60}\right)} & \text{if } t > 60 \end{cases}$$

Environmental scores

$$I_{index} = \begin{cases} 10 & \text{if } I \leq I_{base} \\ 10 \times \frac{I_{base}}{I} & \text{if } I > I_{base} \end{cases}$$

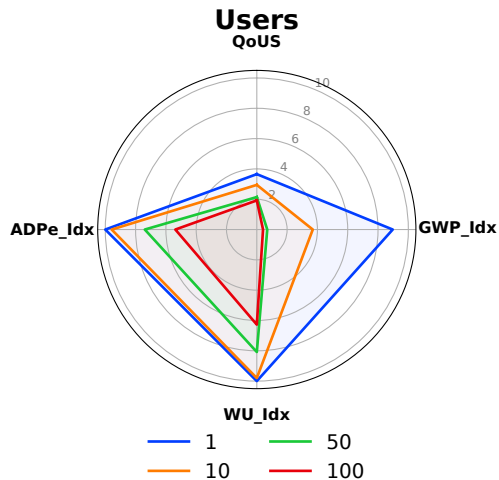
where :

I_{base} = 1 minute of streaming with a resolution of 720p impacts \rightarrow 0.69 g CO₂ eq (GWP), 3.7L eq (WU) and 0.11mg Sb eq (ADPe).

A GWP index of 2 over 10 means that the corresponding agentic system emits 5× more CO₂ than 1 minute of streaming in 720p

Source : <https://base-empreinte.ademe.fr/>

Results : Impacts analysis



Number of users

Quality of User Satisfaction (QoUS) scores

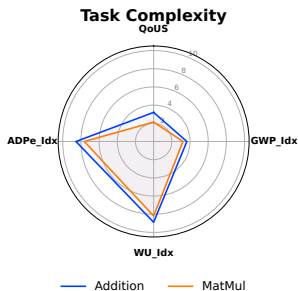
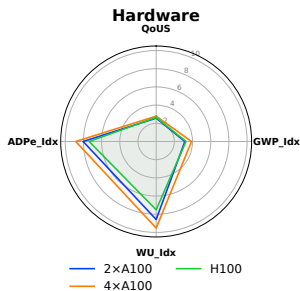
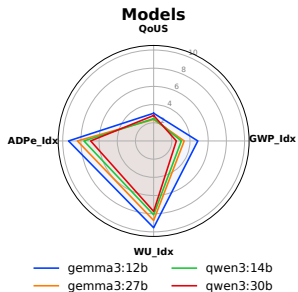
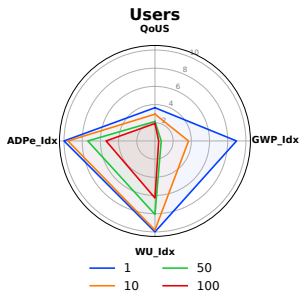
$$QoUS = 10 \times R_{Success} \times R_{response\ time}$$

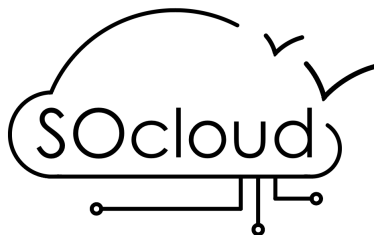
$$R_{response\ time} = \begin{cases} 1 & \text{if } t \leq 60 \\ \frac{1}{1 + \log_{10}\left(\frac{t}{60}\right)} & \text{if } t > 60 \end{cases}$$

Environmental scores

$$I_{index} = \begin{cases} 10 & \text{if } I \leq I_{base} \\ 10 \times \frac{I_{base}}{I} & \text{if } I > I_{base} \end{cases}$$

Results : Impacts analysis





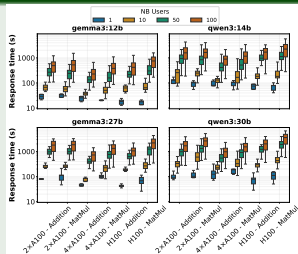
SOcloud Project

- Sober and Open cloud infrastructures
- Going beyond efficiency
- Providing sobriety evaluation metrics - Cloud AI workloads

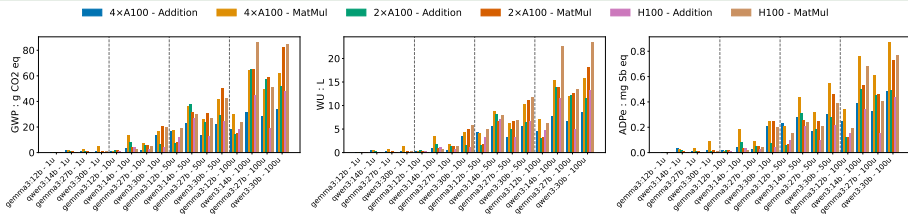
Appendix : Results - Global analysis

Performance Analysis : Success rate & Response time

| Model | Average ADD (%) | Average MATMUL (%) |
|--------------------|-----------------|--------------------|
| gemma3 :12b | 58.6 | 17.6 |
| qwen3 :14b | 37.6 | 41.5 |
| gemma3 :27b | 41.7 | 33.5 |
| qwen3 :30b | 50.8 | 49.3 |



Environmental impacts estimations



Appendix : Issues Description

(a) GitHub Issue: Incorrect addition result (ADD)

Title: Incorrect addition result

Description:

The function `add()` returns an incorrect result.
Instead of adding `a + c`, the code still uses `b`.

Expected behavior:

The function should return `a + c`.

Example:

Input: `a=2, c=3`

Output: 5

Code to Fix

```
int add(int a, int c) {  
    return a + b; // BUG: b should be c  
}
```

(b) GitHub Issue: Incorrect MatMul Multiplication Result (MAT-MUL)

Bug: Incorrect MatMul Multiplication Result

Description

The function `'multiply_matrices'` in `app.py` produces incorrect numerical results for valid input matrices.

Code to fix

```
def multiply_matrices(A, B):  
    m = len(A)  
    p = len(B[0])  
    C = [[0] * p for _ in range(m)]  
  
    for i in range(m):  
        for j in range(p):  
            total = 0  
            for k in range(len(B)):  
                total += A[k][i] * B[j][k]  
            C[i][j] = total  
    return C
```

Bibliography I

- Deepak Bhaskar Acharya, Karthigeyan Kuppan, and B Divya. Agentic ai : Autonomous intelligence for complex goals—a comprehensive survey. *IEEe Access*, 13 :18912–18936, 2025.
- Adrien Berthelot, Eddy Caron, Mathilde Jay, and Laurent Lefèvre. Estimating the environmental impact of generative-ai services using an lca-based methodology. *Procedia CIRP*, 122 :707–712, 2024.
- Sophia Falk, David Ekchajzer, Thibault Pirson, Etienne Lees-Perasso, Augustin Wattiez, Lisa Biber-Freudenberger, Sasha Luccioni, and Aimee van Wynsberghe. More than carbon : Cradle-to-grave environmental impacts of genai training on the nvidia a100 gpu. *arXiv preprint arXiv :2509.00093*, 2025.
- Clément Morand, Anne-Laure Ligozat, and Aurélie Névéol. How green can ai be ? a study of trends in machine learning environmental impacts. *arXiv preprint arXiv :2412.17376*, 2024.