Our contribution to a global environmental standard for AI

Company

Jul 22, 2025 | Mistral AI

At Mistral AI, our mission is to bring artificial intelligence in everyone's hands. For this purpose, we have consistently advocated for openness in AI, with a unique focus on empowering organizations that want to own their AI future.

Today, as AI becomes increasingly integrated into every layer of our economy, it is crucial for developers, policymakers, enterprises, governments and citizens to better understand the environmental footprint of this transformative technology. At Mistral AI, we believe that we share a collective responsibility with each actor of the value chain to address and mitigate the environmental impacts of our innovations.

Even though some recent initiatives have been taken, such as the Coalition for Sustainable AI, launched during the Paris AI Action Summit in February 2025, the work to achieve here remains important. Without more transparency, it will be impossible for public institutions, enterprises and even users to compare models, take informed purchasing decisions, fill enterprises' extra-financial obligations or reduce the impacts associated with their use of AI.

In this context, we have conducted a first-of-its-kind comprehensive study to quantify the environmental impacts of our LLMs. This report aims to provide a clear analysis of the environmental footprint of AI, contributing to set a new standard for our industry.

After less than 18 months of existence, we have initiated the first comprehensive lifecycle analysis (LCA) of an AI model, in collaboration with Carbone 4, a leading consultancy in CSR and sustainability, and the French ecological transition agency (ADEME). To ensure robustness, this study was also peer-reviewed by Resilio and Hubblo, two consultancies specializing in environmental audits in the digital industry.

Stage	#	Step	GHG Emissions	Water Consumption	Materials Consumption
INFRASTRUCTURE	1	**Model conception** Download and storage of training data, developers' laptops embodied impacts and power consumption	<1%	<1%	<1%
INFRASTRUCTURE	2	**Datacenter construction** Building and support equipment manufacturing	<1%	<1%	1.5%
INFRASTRUCTURE	3	**Hardware embodied impacts** Server manufacturing transportation and end-of-life	11%	5%	**61%**
COMPUTING	4	**Model training & inference** Power and water use of servers and support equipment	**85.5%**	**91%**	29%
COMPUTING	5	**Network traffic of tokens** Transfer of requests to inference clusters and responses back to users	<1%	<1%	<1%
USAGE	6	**End-user equipment** Embodied impacts and power consumption	3%	2%	7%
USAGE	7	**Downstream 'enabled' impacts** Indirect impacts that result from the product's use	N/A	N/A	N/A

The location of data centers is one of the main factors of a model's environmental impact.

At Mistral AI,
we are building our own data center in France, leveraging low-carbon electricity and a cool climate to reduce GHG emissions and water usage.

The footprint of a model is strongly correlated with its size. Choosing smaller or case-specific models helps mitigate the environmental impact.

At Mistral AI,
we offer our customers a broad range of model sizes, starting with our smallest model, Ministral 3B.

Inference is a key part of a model's life cycle. End users can also help reducing the impact.

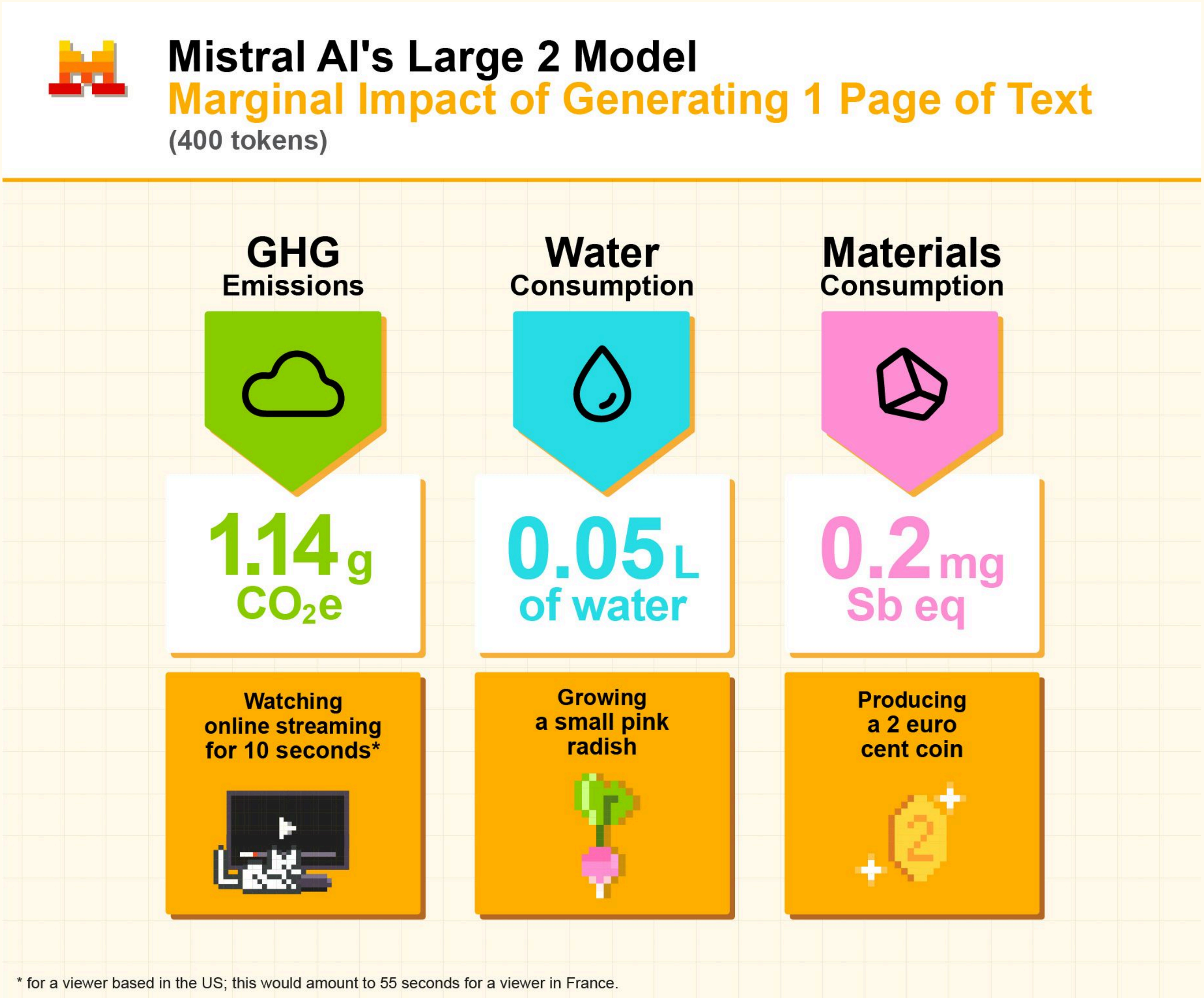
At Mistral AI,
we recommend writing precise prompts and asking for short, grouped answers whenever possible.

In addition to complying with the most rigorous standards*, the aim of this analysis was to quantify the environmental impacts of developing and using LLMs across three impact months of usage, Large 2 generated the following impacts:

  - 20,4 ktCO_2e,
  - 281 000 m3 of water consumed,
  - and 660 kg Sb eq (standard unit for resource depletion).

- the marginal impacts of inference, more precisely the use of our AI assistant Le Chat for a 400-token response - excluding users’ terminals:

  - 1.14 gCO_2e,
  - 45 mL of water,
  - and 0.16 mg of Sb eq.



use.

Learnings and limits of this study

Given the results of our study, we are convinced that the following three indicators are of great importance for users, AI developers and policy makers to fully understand and manage the environmental impacts of LLMs:

1. the absolute impacts of training a model,
2. the marginal impacts of inference,
3. and the ratio of total inference to total life-cycle impacts.

Indicators 1 and 2 could be mandatory figures to report in order to inform the public on impacts, while indicator 3 can act as an internal indicator with optional disclosure. The latter indicator is key to grasp a complete vision of lifecycle impacts, and to ensure that models' training phases are amortized, and not wasted.

Our study also shows a strong correlation between a model's size and its footprint. Benchmarks have shown impacts are roughly proportional to model size: a model 10 times bigger will generate impacts one order of magnitude larger than a smaller model for the same amount of generated tokens. This highlights the importance of choosing the right model for the right use case.

It is worth noting that this study is a first approximation, with the difficulty to make precise calculations in such an exercise when no standards exist for LLM environment accountability, and the absence of publicly available impact factors. For instance, a reliable life-cycle inventory of GPUs is yet to be made, as their embodied impacts had to be approximated but account for a significant portion of total impacts.

To comply with the GHG Protocol Product Standard, future audits made in the industry may follow this study's principles of using location-based approach for electricity emissions and including all significant upstream impacts— i.e., not only those from GPU electricity use, but also all other electricity consumptions (CPUs, cooling devices, etc.) and manufacturing of hardware.

A path toward a global environmental standard

These results point to two levers to reduce the environmental impact of LLMs.

- First, to improve transparency and comparability, AI companies ought to publish the environmental impacts of their models using standardized, internationally recognized frameworks. Where needed, specific standards for the AI sector could be developed to ensure consistency. This could enable the creation of a scoring system, helping buyers and users identify the least carbon-, water- and material-intensive models.
- Second, from the user side, encouraging the research for efficiency practices can make a significant difference:
  - developing AI literacy to help people use GenAI in the most optimal way,
  - choosing the model size that is best adapted to users' needs,
  - grouping queries to limit unnecessary computing,

For public institutions in particular, integrating model size and efficiency into procurement criteria could send a strong signal to the market.

Conclusion

Moving forward, we are committed to updating our environmental impact reports in the future and participating in discussions around the development of international industry standards. We will advocate for greater transparency across the entire AI value chain and work to help AI adopters make informed decisions about the solutions that best suit their needs. The results will later be available via ADEME's Base Empreinte database, setting a new standard for future reference for transparency in the AI sector.

By encouraging sufficiency and efficiency practices and publishing standardized environmental impact reports, we can collectively work towards aligning the AI sector with global climate goals. This study is a humble contribution towards a more accessible and sustainable future for AI.

—

** This study was carried out following the Frugal AI methodology developed by AFNOR and is compliant with international standards, including the Green House Gas (GHG) Protocol Product Standard and ISO 14040/44.*

** The environmental impacts were assessed using standard indicators common in Lifecycle Analyses (LCAs): greenhouse gas emissions measured by Global Warming Potential over 100

ADP quantifies the depletion of non-renewable resources (such as metals and minerals) by considering both the current extraction rates and the estimated reserves of each material. These values are standardized relative to Antimony's ADP, providing a uniform unit since Antimony is a scarce resource.

For example, extracting 1 kg of gold corresponds to an ADP of 2.35 kg Sb eq, whereas extracting 1 kg of copper corresponds to 0.000000161 kg Sb eq.