# Quantifying the Environmental Impacts of LLMs in DH through Carbon Reporting

Sarah Lang*

2026-02-03

*DHd AG GKI Webinar*, `https://agki-dh.github.io/`

*Max Planck Institute for the History of Science (Berlin)

# Is LLM Usage Environmentally Problematic?

## Is (personal/academic) LLM usage environmentally problematic?

- AI-based search is more energy-intensive than traditional search.
- Many search engines now include AI by default.
- Some argue personal AI usage has negligible environmental impact (Masley 2025).
- We need environmental reporting but it also requires points of comparison.
- **Main Argument:** Carbon reporting should be standardized in (DH) research practices.


## References

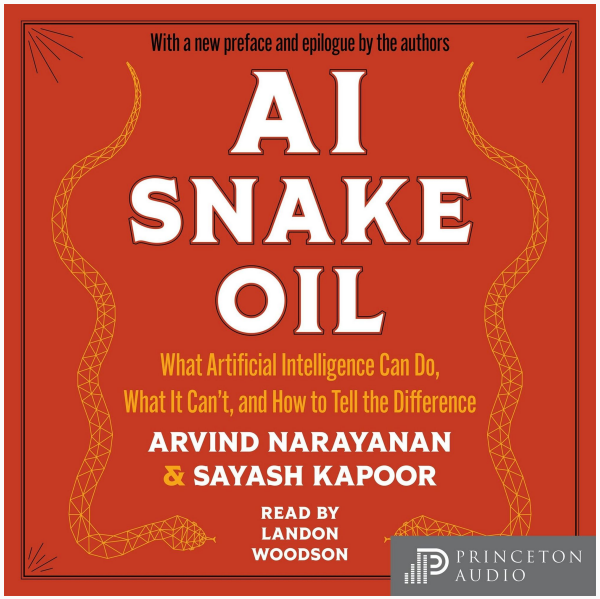
*https://www.nytimes.com/2025/09/26/opinion/ai-quartz-mining-hurricane-helene.html*

*https://www.youtube.com/watch?v=OJWcZLx45pw*

# Environmental Impacts in Digital Humanities

## LLMs and the Environment

- Early interest in environmental impacts via DHd's Greening DH and the *DH and the Climate Crisis* manifesto (Baillot et al. 2021).
- Sustainability in archiving already discussed since 2019 (Pendergrass et al. 2019).
- Surge in LLM use and public interest since 2023 with key papers referencing environmental aspects (e.g. *Stochastic Parrots* paper = Bender et al. 2021).
- But the perceived LLM revolution is mainly in usability and perception, not tech itself (Narayanan and Kapoor 2024).



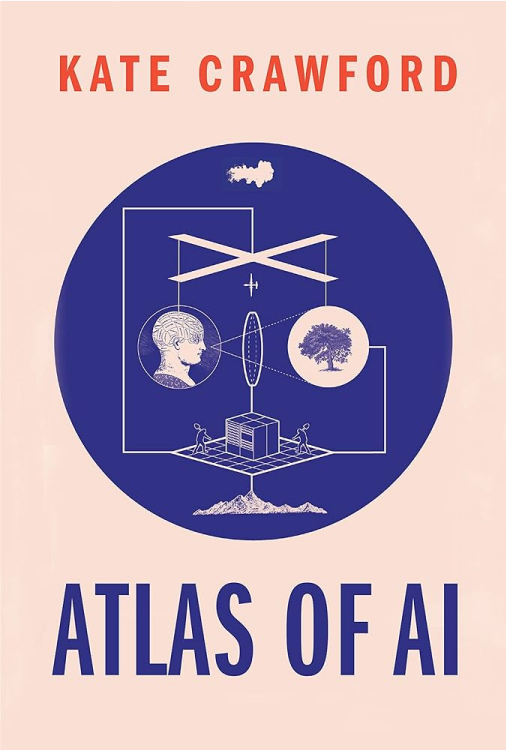

## Learn more at

*https://dhd-greening.github.io/*

# Critical Perspectives: Kate Crawford's *Atlas of AI*

## AI Criticism: Earth, Water, Energy

- **Kate Crawford:** advocate advancing critical perspectives on AI (*ImageNet Roulette*, *Anatomy of an AI System*, ...).
- **Recent Work:** NY Times piece mentions Spruce Pine (NC, US) mining and hurricane Helene impacts → essential extraction site for rare earths (High Purity Quartz) in AI supply chain.




# Water and Energy: Guesstimates and Uncertainty

## AI, Water and the Challenges of Reporting

- **Hank Green** explains complexities of AI water usage.
- Multiple estimates are all kind of valid, yet differ widely.
- The term **“guesstimate”** reflects the uncertainty in impact metrics.
- Query impacts are difficult to track due to many hidden variables, even if all required information were publicly available.


## Watch on Youtube

*https://www.youtube.com/watch?v=H_c6MWk7PQc*

# Carbon Reporting

## Key Points

- Carbon reporting is gaining attention in some LLM contexts.
- The BLOOM model is built with this in mind (Luccioni, Viguier, and Ligozat 2023).
- **DH Community:** Mixed or lacking response in Computational Humanities (CHR) conference – rejected CHR paper, reviewers indifferent.
- Does advocating for carbon reporting not matter in DH?

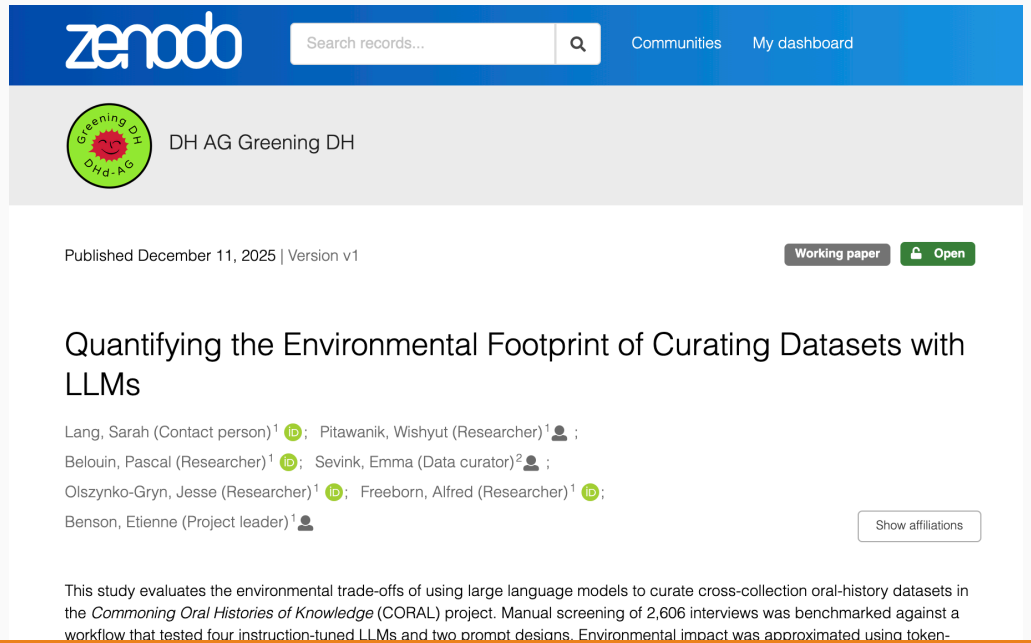

## Learn more:

Lang, Sarah, Wishyut Pitawanik, Pascal Belouin, Emma Sevink, Jesse Olszynko-Gryn, Alfred Freeborn, and Etienne Benson. "Quantifying the Environmental Footprint of Curating Datasets with LLMs". Zenodo, December 11, 2025. *https://doi.org/10.5281/zenodo.17902822*. (Zenodo DHd AG Greening DH Community)

# Carbon Reporting: Technical and Regulatory Considerations

## It's challenging...

- U.S.-companies-based tools are not compliant with German privacy regulations. The local alternative provider GWDG seems to be working on a solution.
- Our paper attempts to provide practical guidance and derived metrics for local use. The methodology is (somewhat naively) based on input/output tokens and location of computation.
- Remember this is a guesstimate due to partial and indirect metrics! Be wary of spurious precision! Modeling theory teaches us that quantifying isn't inherently meaningful. Putting a number on it does not necessarily make it more objective!



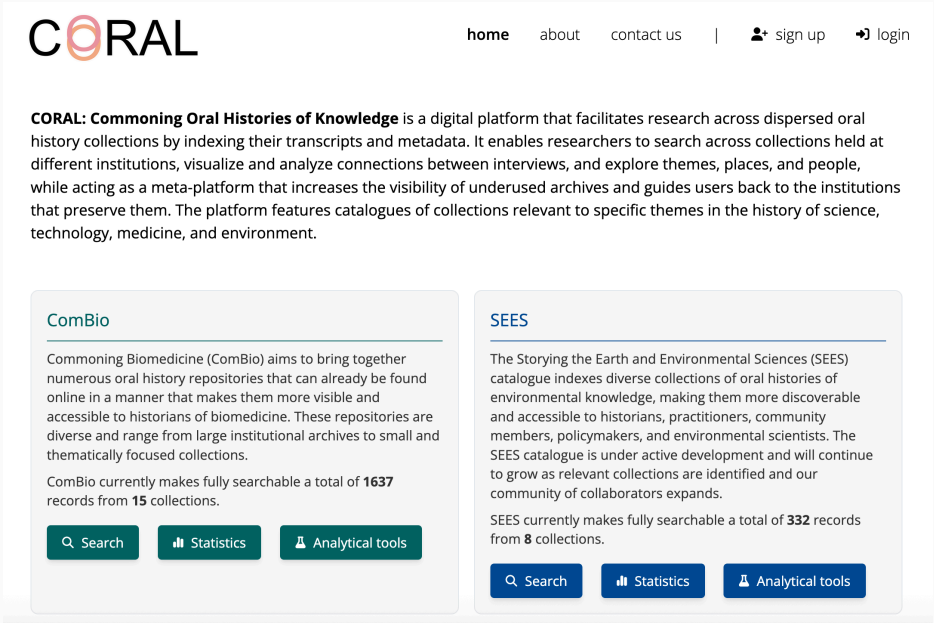

## Reference

Lang, Sarah, Wishyut Pitawanik, Pascal Belouin, Emma Sevink, Jesse Olszynko-Gryn, Alfred Freeborn, and Etienne Benson. "Quantifying the Environmental Footprint of Curating Datasets with LLMs". Zenodo, December 11, 2025. *https://doi.org/10.5281/zenodo.17902822*. (Zenodo DHd AG Greening DH Community)

# The CORAL Project: Carbon Reporting in Practice

## CORAL

- *Commoning Oral Histories of Knowledge* project.
- LLM usage for curating thematic sub-collections.
- Keyword matching fails for complex environmental terms.
- LLMs employed for pre-sorting interviews by relevance.
- Human-in-the-loop workflow for final selection and validation.
- Comparison of human effort vs. LLM results.



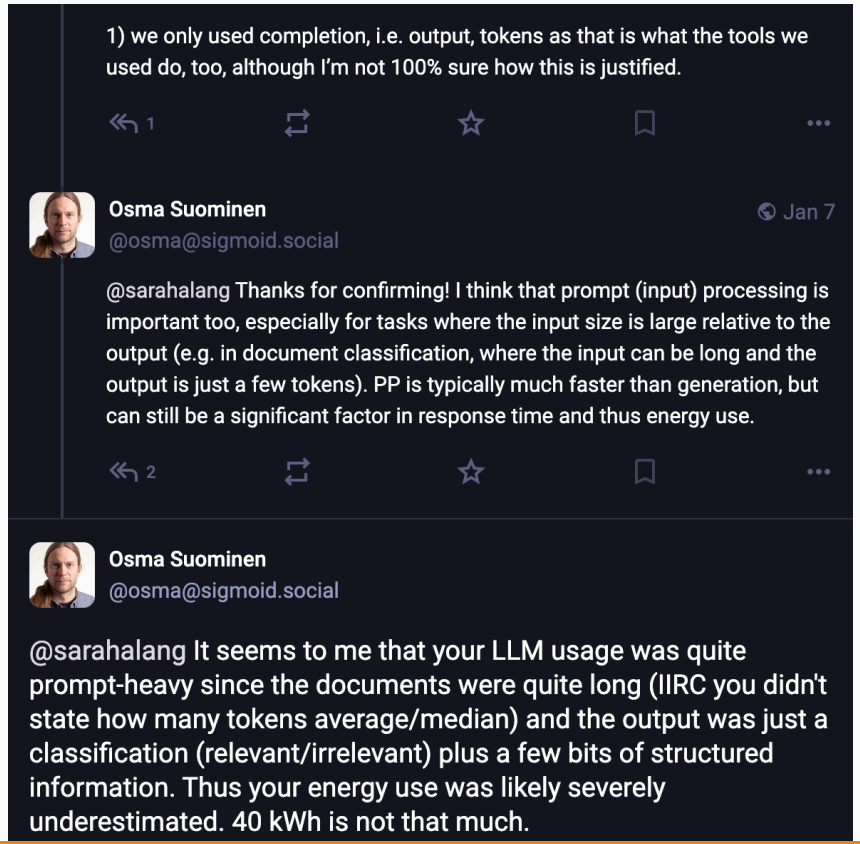

## See the platform

*https://coral.mpiwg-berlin.mpg.de/*

# Conclusions and Reactions

## A bright future?

- Surprise at limited interest in CHR but engagement from this audience and on Mastodon is appreciated.
- Discussion continued on Mastodon. **Criticism:** Estimates likely underestimate actual environmental impact. Suggestion to consider both input and output tokens for better accuracy.

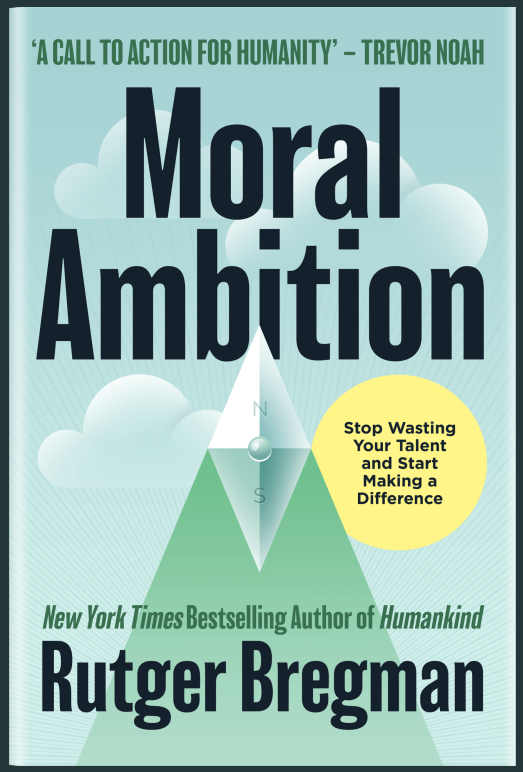



## Reference

*https://fedihum.org/@osma@sigmoid.social/115854498900953842*

Carbon Reporting helps us raise awareness &

What gets measured gets managed. (P. Drucker)




But are more awareness and visibility enough?

# References i


[1] Anne Baillot et al. ***Digital Humanities and the Climate Crisis: A Manifesto.*** *https://dhc-barnard.github.io/dhclimate/*. Online document. 2021.

[2] Emily M. Bender et al. **"On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?"** In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '21. Virtual Event, Canada: Association for Computing Machinery, 2021, pp. 610–623. ISBN: 9781450383097. DOI: *10.1145/3442188.3445922*. URL: *https://doi.org/10.1145/3442188.3445922*.

[3] Alexandra Sasha Luccioni, Sylvain Viguier, and Anne-Laure Ligozat. **"Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model".** In: *Journal of Machine Learning Research* 24 (2023), pp. 1–15. URL: *https://www.jmlr.org/papers/volume24/23-0069/23-0069.pdf*.

[4] Andy Masley. ***Why using ChatGPT is not bad for the environment – a cheat sheet.*** *https://andymasley.substack.com/p/a-cheat-sheet-for-conversations-about*. Substack blog post. Apr. 2025.

# References ii


[5] Arvind Narayanan and Sayash Kapoor. ***AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference.*** Princeton, NJ: Princeton University Press, 2024.

[6] Keith Pendergrass et al. **"Toward Environmentally Sustainable Digital Preservation".** In: *The American Archivist* 82.1 (June 2019), pp. 165–206. DOI: *10.17723/0360-9081-82.1.165*. URL: *https://doi.org/10.17723/0360-9081-82.1.165*.