Large language model co-pilot for transparent and trusted life cycle assessment comparisons

Nathan Preuss
Systems Engineering
Cornell University
Ithaca, New York 14853, USA

Fengqi You
Roxanne E. and Michael J. Zak Professor in Energy Systems Engineering
Cornell University
Ithaca, New York 14853, USA
fengqi.you@cornell.edu

Abstract

Intercomparing life cycle assessments (LCA), a common type of sustainability and climate model, is difficult due to basic differences in fundamental assumptions, especially in the goal and scope definition stage. This complicates decision-making and the selection of climate-smart policies, as it becomes difficult to compare optimal products and processes between different studies. To aid policymakers and LCA practitioners alike, we plan to leverage large language models (LLM) to build a database containing documented assumptions for LCAs across the agricultural sector, with a case study on livestock management. The articles for this database are identified in a systematic literature search, then processed to extract relevant assumptions about the goal and scope definition of the LCA and inserted into a vector database. We then leverage this database to develop an AI co-pilot by augmenting LLMs with retrieval augmented generation to be used by stakeholders and LCA practitioners alike. This co-pilot will accrue two major benefits: 1) enhance the decision-making process through facilitating comparisons among LCAs to enable policymakers to adopt data-driven climate policies and 2) encourage the use of common assumptions by LCA practitioners. Ultimately, we hope to create a foundational model for LCA tasks that can plug-in with existing open source LCA software and tools.

1 Introduction

Life cycle assessments (LCA) are a pre-eminent and commonly used model to evaluate the environmental and climate impacts of products and processes, as well as providing the basis for climate-informed decision-making (e.g. [1]). While the stages and contents of LCAs are standardized [2], different studies on the same process or product may share the same functional unit but include different processes in the system boundary. This inconsistency complicates the use of LCAs as a decision-support tool for formulating climate-smart policies and increases the time necessary to compare the results from the LCA studies. In the worst case, the decision maker might make an incorrect choice based on the high-level findings, or might take significant time trying to understand how the different assumptions affect the LCA results.

To address these problems, large language models (LLMs) can be utilized across all four stages of the LCA, from assisting practitioners in identifying the goal and scope of the LCA, establishing

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2024.

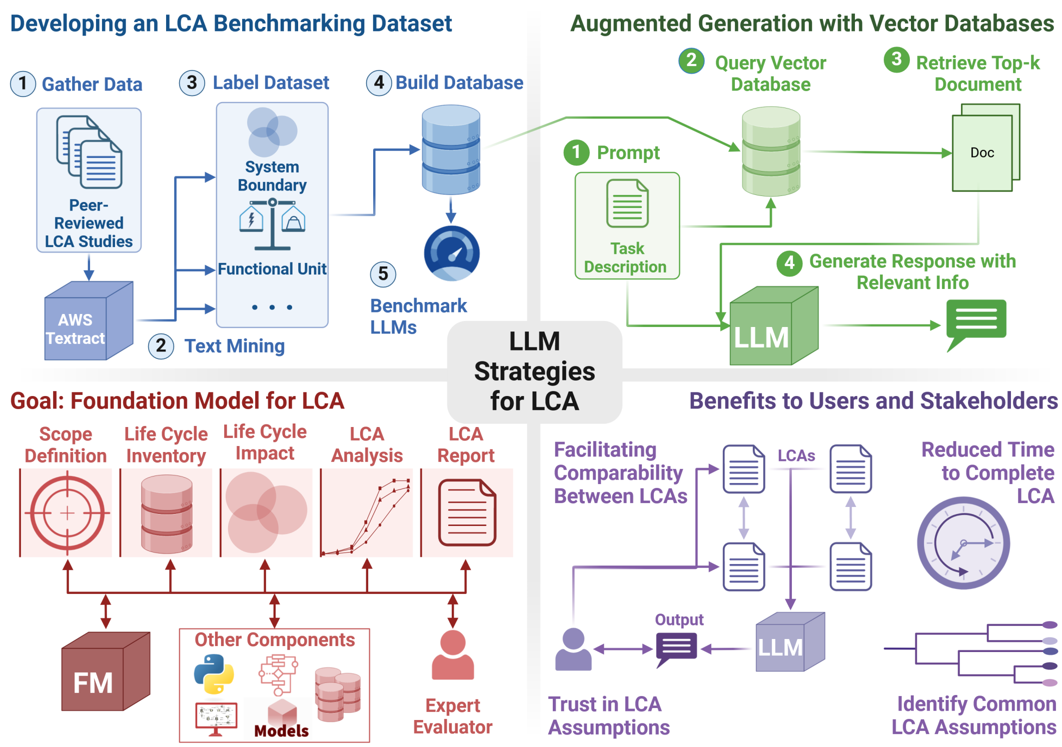

Figure 1: Overview of the key strategies and benefits to improve the usability of life cycle assessments (LCA) as the basis for climate-smart decision making through large language models (LLM), including developing an LCA benchmarking dataset and implementing retrieval augmented generation, with the ultimate goal of developing a foundation model (FM) for LCA.

provenance for emissions factors, and visualizing and interpreting results [3]. LLMs are emerging artificial intelligence methods that act as sophisticated tools to mimic human writing [4], and have achieved human-expert level accuracy across a variety of natural language tasks [5, 6, 7]. As a result, LLMs are already being used to manage life cycle inventory data, such as the TianGong AI project [8] and AutoPCF [9]. Moreover, fine-tuned BERT models are used for impact factor matching [10] and supply chain emissions estimation [11]. Up to this point, the use of language models has been confined to impact factor matching and other life cycle inventory collection tasks. In this study, we use LLMs to analyze the goal and scope definition assumptions to achieve two goals: increase the consistency of LCA studies by giving practitioners tools to utilize consistent assumption for certains systems, as well as enabling policy-makers to closely interrogate and analyze LCA results and assumptions, aiding the adoption of climate-smart policies and increasing the transparency of the decision-making process. This can streamline the process of conducting and interpreting LCAs, increasing comparability and transparency while increasing their usefulness as a support tool for climate-smart decision making.

Ultimately, we look to create a foundational model for LCAs, as outlined in Figure 1, which could potentially make sustainability assessments more accessible and encourage wider adoption in agriculture and other LCA-intensive sectors. In this project, we conduct a systematic literature review to extract LCA studies on livestock production systems in the agricultural sector. These studies are then processed to form a high-quality dataset that is processed and stored in a vector database, which is referenced through retrieval augmented generation (RAG) to create an LLM co-pilot for both decision makers and LCA practitioners.

2 Proposed work

To build a high-quality LCA-based dataset and a proof-of-concept LLM co-pilot using vector databases and RAG, we conduct a systematic literature review of life cycle assessments in livestock agricultural systems. Manuscripts were identified in academic journal databases, then screened to ensure that the articles are in english, have the full text available, and contain the correct systems and LCA steps. The text is extracted from these articles, chunked, and indexed in a vector database for RAG implementation. We choose agriculture as a sector due to the demonstrated success of applying LLM co-pilots as agronomist assistants [12, 13, 14], and we focus specifically on livestock systems due to the excessive variety of functional units and system boundary assumptions [15]. This is meant as a case study – such a dataset could be extended to other sectors in the future, such as transportation and manufacturing, as well as to other tasks throughout the LCA. Indeed, we invite other LCA practitioners, stakeholders, and LLM model developers to contribute to the dataset by making it and associated benchmarking prompts open source at the conclusion of the project.

To facilitate the development of a co-pilot for LCA practitioners, we plan to divide the goal and scope definition into a set of representative tasks, for example, the functional unit definition, the system boundary, and the choice of life cycle impact assessment method. By processing the collected LCAs to extract these attributes, we can significantly enhance the efficiency and effectiveness of LLMs on knowledge base question answering [16]. To improve the usability of the dataset by the state-of-the-art LLMs, the data will be labeled and annotated through collaboration with LCA experts before being entered into the dataset. Through an analysis of the dataset, common practices will be identified for LCA assumptions in the key sectors, facilitating consistent assumptions and the comparability between future and existing LCAs. Meanwhile, a focus on representative tasks will enable us to benchmark pre-trained state-of-the-art LLMs on learned representations of how assumptions made in the goal and scope definition are utilized in acquisition of LCI data, choice of impact assessment method, and the interpretation of LCA results within and between LCAs.

Using LLMs is not without its limitations. Because LLMs are plagued by "hallucinations", or model-generated errors [17], both LCA practitioners and stakeholders may doubt the efficacy of the LLM co-pilot to identify consistent assumptions, which reduces the trustworthiness of the approach. While hallucinations can arise from various causes, they are likely due to limited domain knowledge or reliance on outdated information. To address this limitation, integrating RAG into LLMs emerges as a viable and cost-efficient solution, with the production of higher quality output and a reduction in the incidence of model hallucinations [18], especially in the field of agriculture [19]. RAG enhances LLMs by providing in-context retrieval of domain-specific documents, thereby enabling the user to quickly validate the LCA task against relevant, related documents [20], increasing transparency by establishing provenance for results. This strategy ensures LLMs produce enriched outputs in specialized areas, bypassing the need for extensive LLM training, while ensuring the utilized sources in the vector database are referenced for verification [21, 16] by the human utilizing the co-pilot. Because the vector database can be updated at any time, RAG enables the LLM to draw on high-quality and up-to-date information, bypassing two additional limitations of traditional LLMs. Drawing from the vector database, we will augment a state of the art LLM with RAG to develop an LLM co-pilot. Subsequently, this infuses outputs from the co-pilot with current and specialized knowledge, enabling users to interact with this tool in natural language and uncover provenance for common assumptions, improving the comparability of LCAs and the selection of climate-smart policies. Ultimately, we hope that this model would be able to plug in with existing LCA software, such as the open-source program OpenLCA, and interact with existing and newly created co-pilots with different specialties, such as identifying emission impact factors to improve the accuracy and scalability of LCAs.

2.1 Climate impact

Returning to the example in the introduction, LLMs enhanced with RAG could identify the key differences in assumptions made by the two studies, enabling the decision-maker to quickly and easily identify the tradeoffs between the two production methods and choose the optimal process to reduce the carbon footprint of their product. Moreover, the co-pilot could reduce the time and resources necessary to conduct thorough LCAs, making sustainability assessments more accessible and encouraging wider adoption, reducing the time necessary to identify and implement climate-smart policies across a range of services and sectors.

Acknowledgments and Disclosure of Funding

Funding in direct support of this work: This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE – 2139899. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

All figures are created with BioRender.

References

[1] Haoyue Liang and Fengqi You. Reshoring silicon photovoltaics manufacturing contributes to decarbonization and climate change mitigation. *Nature Communications*, 14(1):1274, March 2023. Publisher: Nature Publishing Group.

[2] Matthias Finkbeiner, Atsushi Inaba, Reginald Tan, et al. The New International Standards for Life Cycle Assessment: ISO 14040 and ISO 14044. *The International Journal of Life Cycle Assessment*, 11(2):80–85, March 2006.

[3] Nathan Preuss, Abdulelah S. Alshehri, and Fengqi You. Large language models for life cycle assessments: Opportunities, challenges, and risks. *Journal of Cleaner Production*, 466:142824, August 2024.

[4] Humza Naveed, Asad Ullah Khan, Shi Qiu, et al. A Comprehensive Overview of Large Language Models, November 2023. arXiv:2307.06435 [cs].

[5] Gemini Team et al. Gemini: A Family of Highly Capable Multimodal Models, December 2023. arXiv:2312.11805 [cs].

[6] OpenAI. GPT-4 Technical Report, March 2023. arXiv:2303.08774 [cs].

[7] Hugo Touvron, Louis Martin, Kevin Stone, et al. Llama 2: Open Foundation and Fine-Tuned Chat Models, July 2023. arXiv:2307.09288 [cs].

[8] TianGong. AI for Sustainability, 2023.

[9] Biao Luo, Jinjie Liu, Zhu Deng, et al. AutoPCF: A Novel Automatic Product Carbon Footprint Estimation Framework Based on Large Language Models. *Proceedings of the AAAI Symposium Series*, 2(1):102–106, 2023. Number: 1.

[10] Bharathan Balaji, Venkata Sai Gargeya Vunnava, Nina Domingo, et al. Flamingo: Environmental Impact Factor Matching for Life Cycle Assessment with Zero-shot Machine Learning. *ACM Journal on Computing and Sustainable Societies*, 1(2):1–23, December 2023.

[11] Ayush Jain, Manikandan Padmanaban, Jagabondhu Hazra, et al. Supply chain emission estimation using large language models, August 2023. arXiv:2308.01741 [cs].

[12] Bruno Silva, Leonardo Nunes, Roberto Estevão, et al. GPT-4 as an Agronomist Assistant? Answering Agriculture Exams Using Large Language Models, October 2023. arXiv:2310.06225 [cs].

[13] Bayer. Bayer demonstrates digital technologies as a key enabler for regenerative agriculture, November 2023.

[14] UIUC. UIUC.chat, 2024.

[15] Elisa Maria Mano Esteves, Ana Maria Naranjo Herrera, Victor Paulo Peçanha Esteves, and Cláudia do Rosário Vaz Morgado. Life cycle assessment of manure biogas production: A review. *Journal of Cleaner Production*, 219:411–423, May 2019.

[16] Zhi Jing, Yongye Su, Yikun Han, et al. When Large Language Models Meet Vector Databases: A Survey, February 2024. arXiv:2402.01763 [cs].

[17] Yue Zhang, Yafu Li, Leyang Cui, et al. Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models, September 2023. arXiv:2309.01219 [cs].

[18] Patrick Lewis, Ethan Perez, Aleksandra Piktus, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, April 2021. arXiv:2005.11401 [cs].

[19] Angels Balaguer, Vinamra Benara, Renato Luiz de Freitas Cunha, et al. RAG vs Fine-tuning: Pipelines, Tradeoffs, and a Case Study on Agriculture, January 2024. arXiv:2401.08406 [cs].

[20] Yikun Han, Chunjiang Liu, and Pengfei Wang. A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge, October 2023. arXiv:2310.11703 [cs].

[21] Yi Zhang, Zhongyang Yu, Wanqi Jiang, et al. Long-Term Memory for Large Language Models Through Topic-Based Vector Database. In *2023 International Conference on Asian Language Processing (IALP)*, pages 258–264, November 2023. ISSN: 2159-1970.