AI in Life Cycle Assessment

Alexei Lapkin[1,2,3] Dogancan Karan,[2] Michael Zhou[3]

[1]*[Chemical Data Intelligence (CDI) Pte Ltd]* [2]*[Cambridge Centre for Advanced Research and Education in Singapore, CARES Ltd]* [3]*[Department of Chemical Engineering and Biotechnology, University of Cambridge]*

Correspondence to: [Alexei Lapkin] aal35@cam.ac.uk

1. Introduction

Life cycle assessment (LCA) is widely used for evaluating environmental impacts of existing and future products and processes. The methodology is based on rigorous enumeration of material and energy balances for the individual process steps, and material and energy flows through the whole value chain of a product (process), from its raw materials to manufacture, to use and end-of-life. The cost of developing materials and energy inventories is precluding LCA from being scaled to the evaluation of every product. For example, there are approx. 100k molecules in commercial use, whereas the largest dataset of inventories of molecules in LCA database ecoinvent contains only about 500 molecules. As a result there is an appetite to develop AI methods for predicting life cycle impacts.

2. AI for LCA Challenges

AI methods need data. As life cycle inventories are relatively limited, building statistical models or language models on existing life cycle inventories leads to low predictive accuracy. There are some published life cycle assessment studies. It is therefore possible to use language models to learn from the published LCAs to the extrapolate to new targets. However, it is a well-known issue (known within the LCA community), that published LCAs are often wrong. Furthermore, the bulk of LCA studies are never published, as these are done by industry for internal purposes. The challenge of using AI for LCA is the lack of data.

3. Automated Evaluation of Environmental Impacts

In our approach we focused on data generation, since the actual impacts calculation is a trivial task once an inventory is known.

We developed a knowledge graph approach to host data and first principles models, and further developed a ‘simulator’ environment that takes advantage of access to data.

In our approach, the information about chemistry and a suggested process is ingested from the user-provided data (PDFs) by our automated data-ingestion pipeline. We then correct errors in the data and supplement the data with missing information using the agentic environment (agents predicting properties, agents seeking information in databases, etc). This creates a problem ‘scheme’ which is then passed to our Simulator to generate the inventory data and later for post-processing to produce impacts.

In this presentation we will show several elements of the overall workflow, and will present the concept of transfer learning for developing deep learning models to predict life cycle impacts that will further generalise our approach.

3.1 Related work

The topic of AI for LCA is growing in importance and interest. Early studies attempted to create ML models that train a neural network on molecular and process features against the available impacts [1]. This leads to low accuracy. There has also been earlier attempts to automate modelling of inventories in LCA [2]. More recently effort was shifted to use of language models for LCA. Recent studies are summarised in two reviews [3,4].

Acknowledgments

This work is co-funded by Pharma Innovation Programme Singapore (PIPS) via a grant “Automated Evaluation of Environmental Impacts”. The authors are grateful to Boehringer Ingelheim for funding and providing experimental data. MZ is grateful to UKRI (EPSRC) for funding his PhD via SynTech CDT (EP/S024220/1).

References

[1] J. Kleinekorte, L. Kröger, K. Leonhard, A. Bardow, A Neural Network-Based Frame-work to Predict Process-Specific Environ-mental Impacts, in: A.A. Kiss, E. Zondervan, R. Lakerveld, L. Özkan (Eds.), Computer Aided Chemical Engineering, Elsevier,

2019: pp. 1447–1452. https://doi.org/10.1016/B978-0-12-818634-3.50242-3.

[2] V.K. Mittal, S.C. Bailin, M.A. Gonzalez, D.E. Meyer, W.M. Barrett, R.L. Smith, Toward Automated Inventory Modeling in Life Cycle Assessment: The Utility of Semantic Data Modeling to Predict Real-World Chemical Production, ACS Sustainable Chem. Eng. 6 (2018) 1961–1976. https://doi.org/10.1021/acssuschemeng.7b03379.

[3] N. Preuss, F. You, Automating Life Cycle Assessments through Artificial Intelligence Agents and Integrated Assessment Models, Environ. Sci. Technol. 60 (2026) 33–48. https://doi.org/10.1021/acs.est.5c14493.

[4] X.X. Romeiko, X. Zhang, Y. Pang, F. Gao, M. Xu, S. Lin, C. Babbitt, A review of machine learning applications in life cycle assessment studies, Science of The Total Environment 912 (2024) 168969. https://doi.org/10.1016/j.scitotenv.2023.168969.