Differentially Private Federated Learning for High-Accuracy Carbon Footprint Prediction

Hanna Jarlaczynska[1], Vijay K. Narasimhan[2], Tingting Ou[3]
[1]Unit8 SA, Cracow, Poland | [2]EMD Electronics, San Jose, CA, USA | [3]Columbia University, NY, USA

NeurIPS 2025 | Tackling Climate Change with ML Workshop

Introduction

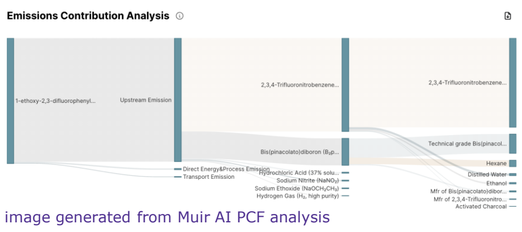
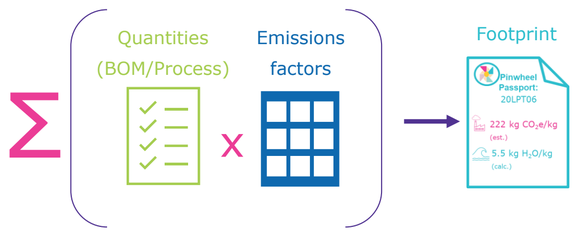


image generated from Muir AI PCF analysis

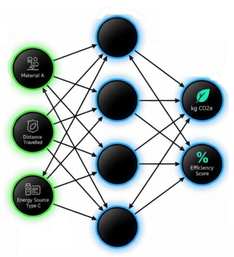
Accurate carbon footprint prediction is essential for sustainable industrial practices and regulatory compliance. To compute the carbon footprint of products, we need to know the footprints of the individual components and steps used in manufacturing.



The calculation involves multiplying each component's quantity by a specific emission factor and then summing over all components. Often, emission factors aren't known by suppliers of specific components, so we use standard values from emissions databases (essentially look-up tables).

Even the most extensive databases today have only thousands of unique chemistries and materials out of the hundreds of millions of chemistries and materials published on to date.

Part of the problem is that detailed carbon footprint calculations include proprietary materials and process data. This significantly limits data sharing. Thus, databases are scaling, but not quickly enough to meet the needs of regulatory reporting and sustainable decision-making.



Inputs: Quantities
Weights: Factors
Outputs: Footprints

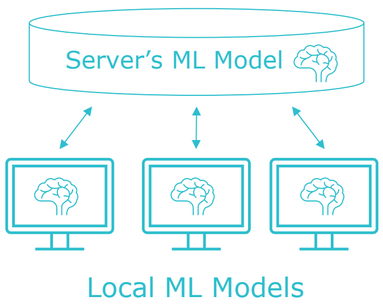
We recast the footprint calculation as a deep learning problem, allowing us to use privacy-preserving techniques to improve the accuracy of calculations.

Layered Privacy Preserving Framework

Our framework, called **Differentially Private Federated Learning**, enables collaborative model training across multiple participants without requiring any exchange of raw data. **Federated Learning** allows local models to be trained independently and only share model updates, not proprietary datasets. **Differential Privacy** ensures that even the shared model updates cannot reveal information about any individual dataset, by introducing mathematically calibrated noise. **Secure Transmission** further protects information during aggregation by transmitting only model differences instead of full parameters.

Federated learning

Aggregate the models , not the data from different clients.

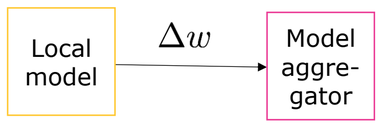


Differential privacy

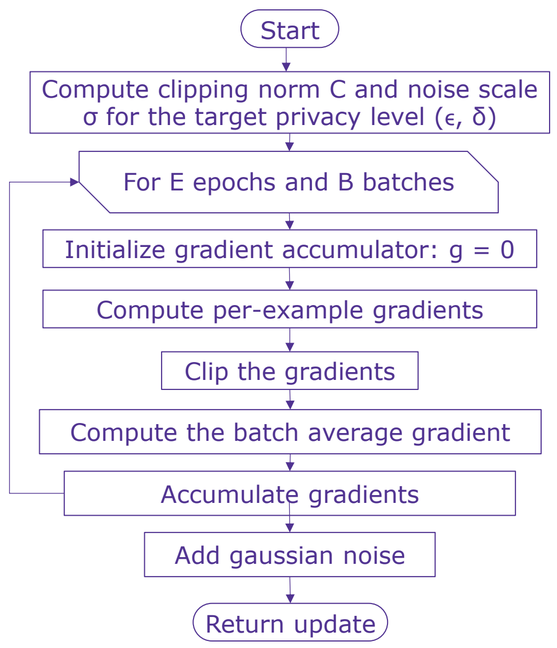
Add a little noise to the model updates to hide the true updates in the FL framework

Secure transmission

Transmit the differences between the models, instead of the full model.



Key algorithm for the local model update



Test Data and Tools

We evaluated our framework using **real-world Life Cycle Assessment (LCA)** data from the **TianGong** open-source database, containing over 10,000 unit processes. Each process was matched with corresponding **carbon emission factors** from **openLCA Nexus**, ensuring consistent input dimensions (n = 557). The curated dataset comprised **508 processes** involving **97 materials**, with carbon footprints computed as the weighted sum of input quantities and emission factors.

The data were split into **80% for federated training** (three clients) and **20% for testing**. Experiments were conducted in **Google Colab** using **Python 3.10** and **PyTorch 2.5.1**. Each client trained an identical **three-layer multilayer perceptron (hidden dimension = 500)** using the **Opacus** library to enforce differential privacy through clipping and noise addition. For proof-of-concept validation, we performed a single federated round (**T = 1**) to demonstrate the effectiveness of the proposed DPFL approach.

Results

DPFL achieves high accuracy while preserving privacy. Aggregating noisy local models improves performance compared to individual clients, reaching **R^2 = 0.96 at ε = 15**, within **5% of the non-private baseline**.

ϵ	$R^2(1)$	$R^2(2)$	$R^2(3)$	$R^2(agg)$	$R^2(baseline)$
1.5	0.9775	0.8150	0.4637	0.7709	0.9039
3	0.6969	0.7240	0.8734	0.8976	0.9852
15	0.9952	0.9464	0.9917	0.9608	0.9947
30	0.9990	0.9853	0.9910	0.9789	0.9954

Conclusions and Impact

Accurate and transparent Life Cycle Impact Assessment (LCIA) is essential for sustainable design and compliance with emerging climate regulations. Yet, existing databases capture only a fraction of the **204 million known materials**, leading to inconsistent impact values that can differ by up to **600%** across sources.

Our **Differentially Private Federated Learning (DPFL)** framework enables organizations to collaboratively build carbon footprint prediction models **without sharing proprietary data**.

This approach provides a **scalable and privacy-preserving foundation for Scope 3 emissions reporting**, bridging data silos across industries. By combining accuracy, security, and collaboration, DPFL advances global sustainability through **trustworthy, data-driven environmental insight**.

Selected References

- Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, CCS '16, page 308–318, New York, NY, USA, 2016. Association for Computing Machinery.
- Mikaela Algren, Wendy Fisher, and Amy E. Landis. Chapter 8 - machine learning in life cycle assessment in Jennifer Dunn and Prasanna Balaprakash, editors, Data Science Applied to Sustainability Analysis, pages 167–190. Elsevier, 2021.
- Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics, pages 1273–1282. PMLR, 2017.
- Xiaoju Chen, H. Scott Matthews, and W. Michael Griffin. Uncertainty caused by life cycle impact assessment methods: Case studies in process-based lci databases. Resources, Conservation and Recycling, 172:105678, 2021.
- P. Kairouz *et al.*, Advances and open problems in federated learning. Found. Trends Mach. Learn., 14:1–210, 2019.
- Ashkan Yousefpour, Igor Shilov, Alexandre Sablayrolles, Davide Testuggine, Karthik Prasad, Mani Malek, John Nguyen, Sayan Ghosh, Akash Bharadwaj, Jessica Zhao, Graham Cormode, and Ilya Mironov. Opacus: User-friendly differential privacy library in pytorch. CoRR, abs/2109.12298, 2021.

The Electronics business of Merck KGaA, Darmstadt, Germany operates as EMD Electronics in the U.S. and in Canada.
Intermolecular is a part of the EMD Electronics business of Merck KGaA, Darmstadt, Germany.

www.emdgroup.com