EMPOWERING SUSTAINABLE FINANCE: LEVERAGING LARGE LANGUAGE MODELS FOR CLIMATE-AWARE INVESTMENTS

Ayush Jain, Manikandan Padmanaban, Jagabondhu Hazra, Shantanu Godbole, Hendrik Hamann
IBM Research Labs

ABSTRACT

With the escalating urgency of climate change, it is becoming more imperative for businesses and organizations to align their objectives with sustainability goals. Financial institutions also face a critical mandate to fulfill the Sustainable Development Goals (SDGs), particularly goal 13, which targets the fight against climate change and its consequences. Mitigating the impacts of climate change requires a focus on reducing supply chain emissions, which constitute over 90% of total emission inventories. In the financial industry, supply chain emissions linked to lending and investments emerge as the primary source of emissions, posing challenges in tracking financed emissions due to the intricate process of collecting data from numerous suppliers across the supply chain.

To address these challenges, we propose an emission estimation framework utilizing a Large Language Model (LLM) to drastically accelerate the assessment of the emissions associated with lending and investment activities. This framework utilizes financial activities as a proxy for measuring financed emissions. Utilizing the LLM, we classify financial activities into seven asset classes following the Partnership for Carbon Accounting Financials (PCAF) standard. Additionally, we map investments to industry categories and employ spend-based emission factors (kg-CO2/$-spend) to calculate emissions associated with financial investments. In our study, we compare the performance of our proposed method with state-of-the-art text classification models like TF-IDF, word2Vec, and Zero-shot learning. The results demonstrate that the LLM-based approach not only surpasses traditional text mining techniques and performs on par with a subject matter expert (SME) but most importantly accelerates the assessment process.

1 INTRODUCTION

The United Nations Sustainable Development Goals (SDGs) outline a path towards a more sustainable future, comprising 17 goals and 169 targets agreed unanimously by all UN member countries in 2015. SDG 13 focuses on addressing climate change, advocating for the integration of climate measures into national policies, strategies, and plans Nations (2023). Despite these efforts, it has been challenging to meet the Paris Agreement’s goal of limiting temperature increase to 1.5°C above pre-industrial levels and achieving net-zero emissions by 2050. Projections suggest a potential surpassing of the critical 1.5°C threshold by 2035 unless significant actions are undertaken Nations (2023). The financial sector, a key player in the global economy, holds substantial influence over resource allocation. Financial institutions, through investments and lending, can significantly impact carbon emissions. Directing capital toward decarbonization aligns with a 1.5°C scenario, and institutions can actively mitigate their climate impact by disclosing the carbon footprint of their portfolios.

The GHG Protocol Corporate Standard Protocol et al. (2004) provides a systematic framework for quantifying an enterprise’s greenhouse gas (GHG) emissions, categorizing them into three scopes. Scope 1 (S1) pertains to direct emissions from owned or controlled sources, while Scope 2 (S2) involves indirect emissions from purchased energy generation. Scope 3 (S3) encompasses all indirect emissions within the reporting company’s value chain, including upstream and downstream emissions. The GHG Protocol Corporate Value Chain (Scope 3) Accounting and Reporting Standard Protocol (2011) further categorizes scope 3 emissions into 15 categories, with Scope 3 Category 15 (investments) addressing emissions from a reporting company's loans and investments, often a substantial part of financial institutions GHG emissions inventory. The Partnership for Carbon Accounting Financials (PCAF) has developed the Global GHG Accounting and Reporting Standard for the Financial Industry PCAF (2022), facilitating standardized measurement and reporting of financed emissions. This process is essential for adhering to disclosure regulations, providing transparency to investors, establishing net-zero goals, and navigating transition risks. However, the presence of limited data, along with other challenges outlined in Appendix A.2, poses a significant hurdle.

Foundation models, a recent breakthrough in artificial intelligence, involve large-scale self-supervised learning on very large datasets. Their adaptability allows for superior fine-tuning in various tasks, often surpassing the performance of traditional machine learning and deep learning models. This success has driven researchers to investigate how such foundation models, specifically applied to applications in climate and sustainability can be developed and deployed. For instance, Nugent et al. (2021) utilized a domain-specific language model pre-trained on business and financial news data to identify Environmental, Social, and Governance (ESG) topics. Similarly, Luccioni et al. (2020) employed a custom transformer-based NLP model to extract climate-relevant information from financial documents. Additionally, Corringham et al. (2021) investigated the use of pre-trained transformers to classify sentences in a dataset of climate action plans submitted to the United Nations after the 2015 Paris Agreement. To encourage the development of foundation models for Earth monitoring, Lacoste et al. (2021) proposed a benchmark comprising various downstream tasks related to climate change. Furthermore, Balaji et al. (2023) suggested estimating the carbon footprint of household products using zero-shot semantic text similarity. However, the application of foundation models in estimating supply chain emissions in the financial industry remains unexplored. Further investigation is needed to understand their potential contribution to reducing greenhouse gas emissions and achieving sustainable development goals, particularly in the financial sector.

In this paper, we propose a LLM based financed emission estimation framework by utilizing financial investment as a proxy for financed emissions. We leveraged LLM to accelerate the classification of the financial disclosures to 7 asset classes, further mapped the investment to 66 industry classes and leveraged industry average spend based emission factors (kg CO_2/\$) to calculate emission associated with the financial investments. We compared the performance of the proposed method with the state-of-art text classification models such as TF-IDF, word2Vec, and Zero shot learning.

2 METHODOLOGY

We introduce an novel framework A.3 for estimating Scope 3 emissions in the financial industry by leveraging large language models (LLM). The framework comprises four modules: data preparation, fine-tuning of LLM, classification, and emission computation. In the data preparation module, we thoughtfully created approximately 2600 examples of financial investment expenses for seven asset classes following PCAF guidance (see appendix A.1) and 21000 samples for 66 industry sector classes based on the US EEIO standard (refer to appendix A.4). These samples ensure adequate representation from each subclass. In the subsequent module, we selected pre-trained foundation models (BERT and RoBERTa) and fine-tuned them using labeled samples with a 70:20:10 train-validation-test split. After adapting the model, we assessed its performance with test samples, identifying classes with low performance for further fine-tuning with additional training samples. Once fine-tuning is complete, the fine-tuned foundation models are employed for inferring both asset and industry classes in unknown financial investment description data. In the last module, the financed emissions of a loan or investment in a company are computed by multiplying the attribution factor with the emissions of the respective borrower or investee company. The attribution factor is derived from the proportional share of investment in an investee company. The emission of the investee company is calculated by multiplying its revenue in each industry class with the spend-based emission factor (CO_2/\$) Ingwersen & Li (2022) of that industry class . For each asset class, emissions are computed according to the PCAF standard (see A.1).

3 EXPERIMENTATION

We explore different training strategies, including our proposed approach, to assess and compare their performance on the given problem.

3.1 USING ZERO-SHOT CLASSIFICATION

We conducted zero-shot learning using semantic text similarity, following the approach proposed in Balaji et al. (2023), to evaluate if financial investment ledger data classification could be performed without using domain-specific training data. This method utilizes pre-trained language models and sentence transformers, designed to encode textual information into fixed-length vector representations. Different open-source sentence-transformer models, such as `all-mpnet-base-v2`, `all-MiniLM-L12-v2`, and `all-MiniLM-L6-v2`, were experimented with. Cosine similarity between embeddings of input financial ledger text and asset or industry sector classes served as the measure of semantic similarity in all experimental settings.

3.2 SUPERVISED LEARNING USING CLASSICAL MODELS

TF-IDF (Term Frequency-Inverse Document Frequency) signifies term importance within a document based on its frequency in that document (TF) and scarcity across the entire dataset (IDF). We extracted features and computed TF-IDF values for each term in the training dataset. Subsequently, we vectorized input transaction text into numeric feature vectors based on TF-IDF values, utilizing these vectors to train our machine learning classifier model.

Word2Vec, introduced by Mikolov et al. (2013), is an widely adopted algorithm for generating word embeddings. We utilized Word2Vec to craft feature representations for financial transaction description data by calculating the mean of word embeddings for all words in a specific transaction text. We trained classical ML models, including random forest classifier, SVM classifier, and MLP classifier, for both TF-IDF and Word2Vec-based feature vectors representation.

3.3 SUPERVISED FINE-TUNING

We conducted extensive experiments involving the refinement of encoder-based Large Language Models (LLMs) to classify transaction data into distinct commodity categories. Pre-trained models, such as `bert-base-uncased` and `roberta-base`, available on the Huggingface library, were fine-tuned. These models are pre-trained using masked language modeling (MLM) and next sentence prediction (NSP) objectives, capturing contextual nuances and bidirectional relationships between words. Adjustments in sequence length parameters and learning rates were explored, and model evaluation relied on the checkpoint with the lowest validation loss.

4 RESULTS AND DISCUSSIONS

4.1 EVALUATION METHOD

We evaluate the outlined approaches using test data consisting of 2610 samples of financial investment transaction texts. The metric employed for comparison is the weighted F1 score.

Performance of zero-shot method Table 1 displays the results from zero-shot approaches. Notably, F1 scores range from 64-71% for asset class classification and 20-22% for industry class classification. This aligns with expectations, given the fewer (7) asset classes compared to the numerous (66) industry classes. Among the sentence transformers, `all-MiniLM-L12-v2` outperforms others, consistent with documentation details. Overall, the text similarity approach exhibits low performance, underscoring the importance of a supervised approach.

Table 1: Zero-shot classification

Model	F1(Asset)	F1(Industry)
all-mpnet-base-v2	66.70%	21.67%
all-MiniLM-L6-v2	64.65%	20.49%
all-MiniLM-L12-v2	71.04%	21.48%

Table 2: Supervised learning- classical model

	F1 (Asset)	F1 (Industry)
TF-IDF	96%	21%
Word2Vec	95%	37%

Performance of Supervised Learning using classical models Table 2 evaluates classical models with TF-IDF and Word2Vec vectorization using a random forest classifier. Both methods perform similarly for asset classification, but Word2Vec outperforms TF-IDF in the 66 industry sector classification. This is attributed to Word2Vec's ability to capture semantic relationships and provide contextual insights, surpassing TF-IDF's limitations.

Performance of Supervised Fine Tuning Table 3 presents the performance of `roberta-base` and `bert-base-uncased` for different *max_length* settings. The models performance remains unaffected by varying *max_length*, indicating the negligible impact on model performance. Reduced learning rates during fine-tuning result in more gradual convergence and improved validation loss as shown in table 4. This approach outperforms traditional machine learning models, showcasing the efficacy of pre-training LLMs on extensive text corpora.

Table 3: Supervised fine-tuning (α=5e-5)

Model	max _length	F1 (Asset)	F1 (Industry)
roberta-base	64	98.17%	83.12%
	128	98.18%	83.28%
	256	98.2%	83.22%
	512	98.4%	83.52%
bert-base	64	97.6%	82.16%
	128	97.18%	82.38%
	256	97.22%	82.32%
	512	97.84%	82.44%

Table 4: Learning rate (α=5e-6), max_length=512

Model	F1 (Asset)	F1 (Industry)
roberta-base	98.89%	83.87%
bert-base	97.98%	82.84%

Table 5: 50% smaller training data

Model	F1 (Assets)	F1 (Industry)
TF-IDF	64%	17%
Word2Vec	66%	34%
roberta-base	91.29%	77.29%
bert-base	90.18%	75.18%

4.2 Ablation Study

Subset of training data Using 50% of the training and validation data in supervised learning (sections 3.2 and 3.3) leads to reduced F1 scores (Table 5). This suggests that a smaller dataset may hinder effective learning, limiting the model's ability to capture diverse patterns in financial ledger descriptions. Larger training datasets are recommended to enhance model performance.

4.3 Estimation and analysis of Scope 3 emission of financial industry

After mapping financial ledger descriptions to PCAF asset classes and EEIO summary commodity industry/sector classes, we calculate financed emissions using the asset-specific formulation (A.1) and EEIO emission factors (A.4). Appendix A.5 displays sample financial transactions, their classified asset and industry classes, and emission distribution. Notably, the correlation between expenses and emissions is non-trivial and counter-intuitive, providing detailed insights for financial organizations to profile financed emission distribution and make informed decisions on future investments.

5 Conclusion

This paper introduces a fine-tuned large language model (LLM) framework for enabling and accelerating accurate estimation of financed emissions. In experiments, non-foundation models perform similarly to fine-tuned foundation models in asset class classification, while the latter excel in industry class classification. The supervised fine-tuned foundation model outperforms zero-shot classification. The domain-adapted foundation model matches the domain-expert performance in industry class classification. Estimating financed emissions from readily available financial investment records mitigates the complexity and dependency on extensive data collection.It empowers financial organizations for precise estimation, enabling informed decisions in line with Sustainable Development Goal 13.

References

Bharathan Balaji, Venkata Sai Gargeya Vunnava, Geoffrey Guest, and Jared Kramer. Caml: Carbon footprinting of household products with zero-shot semantic text similarity. In *Proceedings of the ACM Web Conference 2023*, pp. 4004–4014, 2023.

Tom Corringham, Daniel Spokoyny, Eric Xiao, Christopher Cha, Colin Lemarchand, Mandeep Syal, Ethan Olson, and Alexander Gershunov. Bert classification of paris agreement climate action plans. Tackling Climate Change with Machine Learning Workshop, held in conjunction with ICML, 2021, 2021.

Wesley W Ingwersen and Mo Li. Supply chain ghg emission factors for us commodities and industries v1.1. 2022. doi: 10.23719/1524524.

Alexandre Lacoste, Evan David Sherwin, Hannah Kerner, Hamed Alemohammad, Björn Lütjens, Jeremy Irvin, David Dao, Alex Chang, Mehmet Gunturkun, Alexandre Drouin, Pau Rodriguez, and David Vazquez. Toward foundation models for earth monitoring: Proposal for a climate change benchmark, 2021.

Alexandra Luccioni, Emily Baylor, and Nicolas Duchene. Analyzing sustainability reports using natural language processing, 2020.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

United Nations. The sustainable development goals report, 2023. URL https://unstats.un.org/sdgs/report/2023/The-Sustainable-Development-Goals-Report-2023.pdf.

Tim Nugent, Nicole Stelea, and Jochen L Leidner. Detecting environmental, social and governance (esg) topics using domain-specific language models and data augmentation. In *International Conference on Flexible Query Answering Systems*, pp. 157–169. Springer, 2021.

PCAF. The global ghg accounting and reporting standard part a: Financed emissions. second edition., 2022. URL https://carbonaccountingfinancials.com/files/downloads/PCAF-Global-GHG-Standard.pdf.

Greenhouse Gas Protocol. Ghg protocol corporate value chain (scope 3) accounting and reporting standard, 2011.

Greenhouse Gas Protocol, Greenhouse Gas Protocol Initiative, et al. A corporate accounting and reporting standard. *World Resources Institute and World Business Council for Sustainable Development*, 2004.

Y. Yang, W. W. Ingwersen, T. R. Hawkins, M. Srocka, and D. E. Meyer. Useeio: A new and transparent united states environmentally-extended input-output model. *Journal of cleaner production*, 158:308–318, 2017.

A Appendix

A.1 Global GHG Accounting and Reporting Standard for the Financial Industry

The Partnership for Carbon Accounting Financials (PCAF) is a global coalition of more than 130 financial institutions that work together to develop and implement a harmonized approach to assess and disclose the greenhouse gas (GHG) emissions associated with their loans and investments. The PCAF was formed in 2015 by major European banks and investors, and has since expanded to include members from North America, Latin America, Africa and Asia-Pacific

Global GHG Accounting and Reporting Standard for the Financial Industry PCAF (2022) was developed by The Partnership for Carbon Accounting Financials (PCAF) to meet the growing industry demand for a global, standardized approach to measure and report financed emissions. The Standard has been reviewed by the GHG Protocol and is in conformance with the requirements set forth in the

Corporate Value Chain (Scope 3) Accounting and Reporting Standard, for Category 15 investment activities. The standard provides detailed methodological guidance to measure and disclose GHG emissions associated with seven asset classes:

1. **Listed equity and corporate bonds:** These are securities issued by companies that are traded on stock exchanges or over-the-counter markets. The PCAF standard provides methods to calculate the emissions of the companies based on their ownership or debt share.

$$\text{Financed emission} = \sum_c \frac{\text{Outstanding amount}_c}{\text{Enterprise value including cash}_c} \times \text{Company emission}_c$$

$$\text{For bonds to private companies} = \sum_c \frac{\text{Outstanding amount}_c}{\text{Total equity} + \text{debt}_c} \times \text{Company emissions}_c$$

2. **Business loans and unlisted equity:** These are loans or equity investments made to private companies that are not listed on public markets. The PCAF standard provides methods to estimate the emissions of the companies based on their turnover, sector, or physical activity data.

$$\text{Financed emission} = \sum_c \frac{\text{Outstanding amount}_c}{\text{Total equity} + \text{debt}_c} \times \text{Company emissions}_c$$

where, for unlisted equity, the outstanding amount is calculated as follows:

$$\text{Outstanding amount} = \frac{\text{Number of shares of financial institution}_c}{\text{Number of totat shares}_c} \times \text{Total equity}_c$$

3. **Project finance:** This is a type of financing that is used for specific projects, such as renewable energy, infrastructure, or mining. The PCAF standard provides methods to measure the emissions of the projects based on their output, capacity, or technology.

$$\text{Financed emission} = \sum_p \frac{\text{Outstanding amount}_p}{\text{Enterprise value including cash}_p} \times \text{Project emissions}_p$$

4. **Commercial real estate:** This is a type of property that is used for business purposes, such as offices, retail, or industrial buildings. The PCAF standard provides methods to assess the emissions of the buildings based on their energy use, floor area, or location.

$$\text{Financed emission} = \sum_b \frac{\text{Outstanding amount}_b}{\text{Property value at origination}_b} \times \text{Building emissions}_b$$

5. **Mortgages:** These are loans that are secured by residential properties, such as houses or apartments. The PCAF Standard provides methods to estimate the emissions of the properties based on their energy use, floor area, or location.

$$\text{Financed emission} = \sum_b \frac{\text{Outstanding amount}_b}{\text{Property value at origination}_b} \times \text{Building emissions}_b$$

6. **Motor vehicle loans:** These are loans that are used to purchase vehicles, such as cars, trucks, or motorcycles. The PCAF Standard provides methods to calculate the emissions of the vehicles based on their fuel type, fuel efficiency, or distance traveled.

$$\text{Financed emission} = \sum_v \frac{\text{Outstanding amount}_v}{\text{Total value at origination}_v} \times \text{Vehicle emissions}_v$$

7. **Sovereign debt:** These include sovereign bonds and sovereign loans of all maturities issued in domestic or foreign currencies.

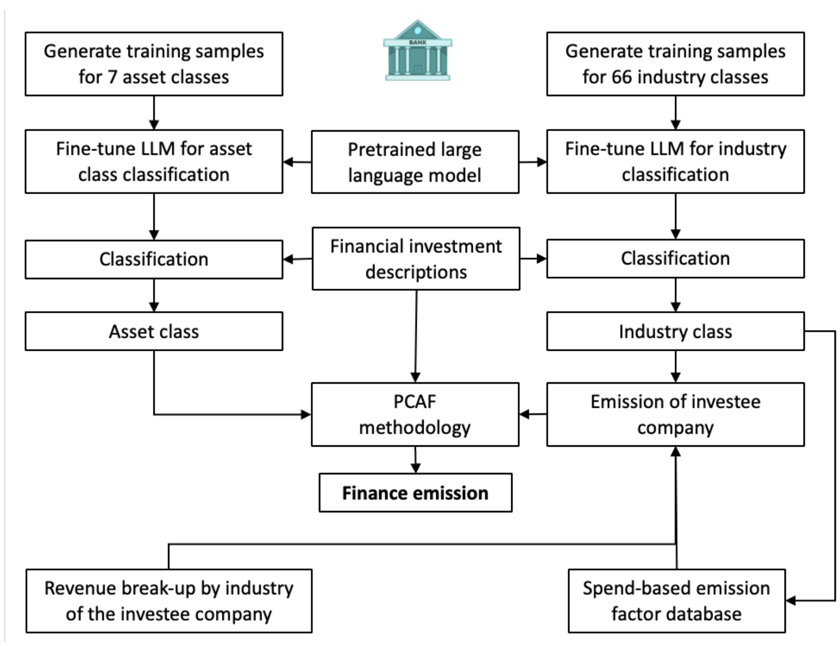
$$\text{Financed emission} = \sum_v \frac{\text{Exposure to sovereign bond}}{\text{Purchase power parity - adjusted GDP}} \times \text{Sovereign emissions}$$

A.2 CHALLENGES WITH ESTIMATING SUPPLY CHAIN EMISSIONS OF FINANCIAL INDUSTRY

Financed emissions are the greenhouse gas emissions that are associated with the activities and projects that are financed by financial institutions, such as loans, investments, insurance and underwriting. Some of the challenges associated with them are:

- Data Availability and Quality: Access to comprehensive and reliable data regarding the carbon footprint of invested companies and projects can be limited. Incomplete or inaccurate data can lead to misleading estimations of emissions.
- Data Complexity: Analyzing emissions from diverse sectors, projects, and investments requires understanding complex industrial processes, energy consumption patterns, and emissions factors.
- Indirect Emissions: Calculating emissions beyond a financial institution's direct control (Scope 3 emissions) requires considering the broader supply chains and operations of invested companies, which can involve diverse methodologies and data sources.
- Variability: The carbon intensity of investments can vary over time due to changes in business practices, technology adoption, and regulatory shifts, making consistent and accurate tracking difficult.
- Lack of Personnel Resources: In order to deal with the above issues, financial institutions need a good management, leadership support as well as expertise in this area, which in itself is not a simple task.

A.3 FRAMEWORK FOR ESTIMATING FINANCIAL EMISSION



A.4 US EEIO DATASET

The US Environmentally-Extended Input-Output (US EEIO) model Yang et al. (2017) functions as a life cycle assessment (LCA) tool, tracing the economic and environmental dynamics of goods and services within the United States. This approach involves a comprehensive dataset and methodology that merges economic input-output analysis with ecological data. Its purpose is to quantify the ecological impacts connected to economic endeavors. The US EEIO model is both open-source and freely accessible, downloadable from the official website of the US Environmental Protection Agency.

Under the US EEIO framework, products are classified into over 380 groups based on shared environmental characteristics, referred to as industry sector classes. These classes are strategically aligned with the North American Industry Classification System (NAICS) and Bureau of Economic Analysis (BEA) codes. This ensures coherence and facilitates integration with existing economic and ecological datasets. Complementing the detailed industry sector classifications, US EEIO also furnishes emission factors for 66 condensed industry classes. These emission factors play a pivotal role in estimating the environmental ramifications of expenditure data. Calculated for each industry sector class, these factors shed light on the ecological implications tied to diverse economic activities.

A.5 SAMPLE FINANCED EMISSION

Sl no.	Ledger entries	Investment	Attribution	Asset	Industry	Emission factor(kg.CO_2/USD)	Final emission
1	ABC has created a well-balanced investment portfolio, directing a total of $20 million into a combination of US SEC, encompassing investment-grade corporate bonds within category D, and listed equity securities within category E	$20 million USD	15%	Listed equity and corporate bonds	Funds, trusts, and other financial vehicles	0.182	546 t.CO_2
2	Balancing social impact, Corporate Finance Ventures allocated $4.5 million for unlisted equity and granted a $3.2 million business loan to a community-based initiative in US	$7.7 million USD	30%	Business loans and unlisted equity	Social assistance	0.135	311.8 t.CO_2
3	Retail Real Estate Partners invested $12 million in the development of a new shopping center designed with experiential retail concepts.	$12 million USD	40%	Commercial real estate	Other retail	0.147	705.6 t.CO_2
4	A $310,000 mortgage loan was provided for the purchase of a duplex, allowing the buyer to live in one unit and rent out the other.	$310K USD	70%	Mortgages	Rental and leasing services and lessors of intangible assets	0.09	19.53 t.CO_2
5	Retail Ventures Inc. secured a $250,000 business auto loan to upgrade their fleet of company cars, with a 5-year term and a variable interest rate based on the prime rate.	$250K USD	80%	Motor vehicle loans	Motor vehicle and parts dealers	0.174	34.8 t.CO_2
6	Gas Power Development obtained a $105 million loan to finance the construction and operation of a combined cycle gas-fired power plant.	$105 million USD	50%	Project finance	Utilities	2.884	151410 t.CO_2
7	A federal bank purchases $2 billion worth of sovereign bonds from Country Z as part of its foreign exchange reserve strategy	$2 billion USD	5%	Sovereign debt	Federal Reserve banks, credit intermediation, and related activities	0.069	6900 t.CO_2