

LLM-Driven Estimation of Personal Carbon Footprint from Dialogues

Shuqin Li¹ Huifang Du² Haofen Wang^{2*}

¹Hangzhou Dianzi University Information Engineering College, Hangzhou, China

²Tongji University, Shanghai, China

shuqinlee9683@gmail.com, duhuifang@tongji.edu.cn, carter.whfcarter@gmail.com

Abstract

Personal Carbon Footprint (PCF) Estimation is crucial for raising individual environmental awareness by linking daily activities to their environmental impact. However, existing tools are limited by fragmented scenarios and labor-intensive manual data entry. We present PCCT, an LLM-powered system that combines conversational understanding with emission knowledge grounding for PCF Estimation. We address two key challenges: (1) resolving incomplete activity information across turns through knowledge-guided and context-aware tracking, and (2) accurately mapping emission factors using multi-step LLM inference and vector-based similarity search. The system dynamically combines knowledge-guided activity extraction, and context-aware memory management, generating accurate carbon footprint estimates. We validate the effectiveness with the *CarbonDialog-1K* benchmark, comprising 1,028 annotated user activity narratives. Experimental results demonstrate that our method outperforms baseline systems in accuracy, while subjective evaluations show superior appropriateness, usability, efficiency, and naturalness.

1 Introduction

Personal Carbon Footprint (PCF) estimation plays a pivotal role in fostering individual environmental awareness by translating daily activities into environmental impacts (Mancini et al., 2016; Lan- nelongue et al., 2021). Current tools, though effective in narrow domains like transportation tracking, face systemic limitations: (1) they rely on fragmented scenario definitions that hinder cross-domain analysis and (2) demand laborious manual data entry, which prevents long-term user engagement (Scrucca et al., 2021; Dreijerink and Paradies, 2020; Chen et al., 2016).

Recently, large language models (LLMs) have demonstrated impressive semantic understanding

and reasoning capabilities (Radford et al., 2019; Mann et al., 2020; Bi et al., 2024). These advancements offer a promising path toward passive carbon footprint estimation through conversational interfaces. We inquire *how to accurately calculate the PCF from the casual conversations by leveraging the LLM's capabilities?*

In order to calculate the PCF, we need two parts: the user's activity parameters and the emission factor accordingly (Mariette et al., 2022). The first challenge arises from the progressive nature of conversational data. Users tend to describe their activities across multiple conversation turns with varying levels of detail and precision. For example, a user might initially mention "driving to work" and later specify "in an electric car for 15 kilometers." This requires the system to reason over the conversation history and Emission Factor database and infer the missing information or ask the user for clarification when necessary (Tu et al., 2024). Even with complete parameters, matching activities to emission factors (EFs) (Solazzo et al., 2021) is a challenging task (Balaji et al., 2023; Wood et al., 2017; Oehlert et al., 2022). Traditional practitioners either rely on manually curated mappings from millions of entries of EF database, which are time-consuming and prone to errors or use coarse-grained rule-based systems that are not scalable and accurate (Scrucca et al., 2021). This demands precise integration of domain knowledge and context-aware reasoning (Deng et al., 2023).

In this work, we propose a Progressive Contextual Carbon Tracking (PCCT) framework. Figure 2 presents our framework architecture. The PCCT framework consists of three components that work together to process multi-turn conversations about daily activities: (1) Knowledge-Guided Progressive Activity Extraction, which identifies user activities while leveraging emission factor knowledge to guide the extraction process; (2) Context-Aware Memory Management, which main-

* Corresponding author

tains structured representations of activities and parameters across conversation turns; and (3) Carbon Footprint Calculation. To validate performance, we construct the *CarbonDialog-1K* benchmark containing 1,028 annotated user activity narratives with ground-truth emission values. Experimental comparisons against a rule-based expert system demonstrate our method’s superior accuracy across all metrics. Complementary user studies reveal 30% higher engagement rates compared to conventional carbon tracking tools.

Our contributions are as follows:

- We propose a Progressive Contextual Carbon Tracking (PCCT) framework that calculates carbon footprint progressively in dialogues guided by knowledge. To our knowledge, this work represents the first use of LLMs for PCF estimation from conversations.
- We introduce CarbonDialog-1K to facilitate the research on carbon footprint estimation from natural language.
- Extensive experimental results demonstrate our system’s superior performance in accuracy and usage effectiveness.

2 Related work

2.1 Carbon Footprint Calculation

Carbon footprint calculation has been a cornerstone of environmental research (Wiedmann and Minx, 2008), focusing on quantifying the greenhouse gas emissions associated with various activities, products, and organizations. Early methodologies, such as life cycle assessment (LCA), provided frameworks for calculating emissions across supply chains and industrial processes (Agyei Boakye et al., 2023; Chen et al., 2021). These approaches were later adapted to assess the environmental impact of individual actions, giving rise to the concept of Personal Carbon Footprint (PCF). The calculation of PCF involves translating daily activities, such as transportation, energy consumption, and dietary habits, into measurable emissions. Traditional methods rely on structured data inputs, often requiring users to manually provide detailed information such as vehicle type, fuel consumption, and travel distance (Scrucca et al., 2021). Some work is proposed to facilitate this process, enabling emissions tracking in specific domains such as commuting and household energy use (Dreijerink and

Paradies, 2020; Balaji et al., 2023). However, these approaches are limited by their fragmented nature, as they often fail to account for cross-domain activities or incomplete user inputs. Recent advancements have attempted to address these limitations by integrating automated data collection methods, such as GPS tracking for transportation emissions (Chen et al., 2016; Gately et al., 2017) and smart meter integration for energy usage (Miao et al., 2024). Despite these improvements, existing tools remain labor-intensive and lack the flexibility to handle diverse and complex user activities, leading to low user engagement and long-term abandonment.

2.2 NLP Techniques for Carbon Footprint Estimation

The application of Natural Language Processing (NLP) techniques to estimate carbon footprints has gained significant attention in recent years, driven by the need for scalable and automated solutions to address the challenges of environmental impact assessment (Murphy, 1998). Traditional methods rely on manual annotation or supervised classification approaches (Sousa and Wallace, 2006), which require large labeled datasets. Recent works have explored to automate the estimation process. For instance, term frequency-inverse document frequency (TF-IDF) features are employed to classify companies into American Industry Classification System (NAICS) codes based on web-scraped text data (Wood et al., 2017; Oehlert et al., 2022). While effective, these methods are limited by their reliance on large labeled datasets and their inability to generalize to unseen or zero-shot scenarios. CaML automates the Economic Input-Output based Life Cycle Assessment (EIO-LCA) (Hendrickson et al., 1998) process by using semantic text similarity matching with the pre-trained model SBERT (Reimers and Gurevych, 2019) to screen and rank product and industry sector matches (Balaji et al., 2023). However, despite the progress made, current NLP-based approaches still struggle to perform accurate carbon emission calculations due to their inadequate context-aware reasoning capabilities, which fail to fully capture the complex relationships in environmental data.

3 Dataset Construction

To enable research on carbon footprint estimation through natural conversations, we construct a com-

prehensive dataset containing 1,028 multi-turn dialogues (*CarbonDialog-1K*) through a two-stage process: (1) building a daily activity emission factor database and (2) generating natural multi-turn dialogues. Rather than collecting real-world conversations and manually annotating them, we adopt a reverse engineering approach (Wang et al., 2015) to ensure comprehensive coverage and calculation accuracy.

3.1 Activity-EF Database Construction

The first stage focuses on building a comprehensive database of emission factors (EFs) for daily activities. We first collect a diverse set of daily activities that potentially contribute to personal carbon footprints. To ensure comprehensive coverage, we leverage large language models to generate a wide range of activities in 6 categories: transportation, energy consumption, food and beverages, consumer goods, services, and entertainment. For each activity, we query the Clima¹ API to obtain standardized emission factors.

Given the significant regional variations in emission factors due to differences in energy mix, infrastructure, and economic development, we focus on activities with well-documented emission factors in selected regions (New Zealand, United Kingdom, and Germany, China, and United States).

3.2 Dialogue Generation

We then adopt the Activity-EF Database to generate dialogues through a reverse engineering approach. We first construct the ground truth activities and their carbon footprints, then generate dialogues that would lead to these calculations. The process consists of three main phases: (1) activity set construction, (2) narrative and parameter generation, and (3) progressive dialogue synthesis. Algorithm 1 formalizes this process.

For each dialogue, we first sample 3-5 activities from the database that are compatible both logically (e.g., avoiding conflicting transportation modes) and regionally (sharing the same region for consistent emission factors). Each activity is then assigned temporal context and enriched with necessary parameters required by its emission factor calculation:

¹<https://www.climatiq.io>, Clima¹ maintains a database of emission factors from authoritative sources such as government environmental agencies and research institutions.

Algorithm 1 Dialogue Generation

```

1: Input: Activity-EF database  $\mathcal{E}$ , Sample size  $N$ 
2: Output: Dialogue dataset  $\mathcal{D}$ 
3: for  $i = 1$  to  $N$  do
4:   Select region  $r$  randomly
5:    $n \leftarrow \text{RandActivityNum}()$ 
6:    $A_r \leftarrow \text{SampleActivities}(n, \mathcal{E}, r)$ 
7:   for each activity  $a \in A_r$  do
8:     Assign temporal context to  $a$ 
9:      $D_a \leftarrow \text{GenerateDescription}(a)$ 
10:     $P_a \leftarrow \text{GenerateParameter}(a)$ 
11:     $C_a \leftarrow \text{CalculateCarbonFootprint}(a)$ 
12:   end for
13:    $C_r \leftarrow \text{SumCarbonFootprint}(\{C_a\})$ 
14:    $A_r \leftarrow \text{GroupActivities}(A_r)$ 
15:    $T \leftarrow \text{ToDialogue}(A_r, \{D_a, P_a, C_a\}, C_r)$ 
16:    $\mathcal{D} \leftarrow \mathcal{D} \cup \{T, C_r\}$ 
17: end for
18: return  $\mathcal{D}$ 

```

$$CO_2e = \sum_{i=1}^n P_i \times EF_i \quad (1)$$

where P_i represents the activity parameter (e.g., distance traveled, fuel consumed) and EF_i is the corresponding emission factor from our database.

The selected activities are then transformed into natural narratives using LLMs, with parameters strategically distributed across multiple dialogue turns. Then, we break an activity’s description into multiple turns. This design mirrors typical human conversation patterns, where users often begin with basic activity descriptions before gradually adding specific details. Table 1 shows an example dialogue constructed through this process.

3.3 Quality Control

We implement a comprehensive quality control pipeline integrating automated filtering, manual review, and LLM-assisted validation. Initially, automatic filters exclude anomalous data by removing: 1) activities with emission factors beyond three standard deviations from category means, indicative of industrial rather than personal behavior; 2) parameter outliers exceeding typical personal usage scenarios (e.g., flights over 20,000 km, meals exceeding \$500); and 3) activities incompatible with personal carbon footprints (e.g., industrial production, commercial shipping).

Subsequently, we validate remaining activities via a hybrid approach, combining manual verifi-

User: Hi, I'd like to track my carbon footprint for today.
System: Hello! I'll help you track your carbon footprint. Please tell me about your activities for today.
User: I swung by a cafe on my way to work and picked up a coffee and pastry for \$5. Then I took a taxi to the city centre. That's all.
System: Thank you. I've recorded the cafe purchase. For the taxi ride, I'll need to know the distance traveled. Could you tell me how long was your taxi ride?
User: The taxi ride was about 5km.
System: Thank you. Here's your carbon footprint breakdown:
1. Cafe purchase: $\$5 \times 1.1224 \text{ kg/NZD} = 5.612 \text{ kg COe}$
2. Taxi ride: $5\text{km} \times 0.018894 \text{ kg/km} = 0.09447 \text{ kg COe}$
Total Carbon Footprint: 5.71 kg COe
Would you like to add any other activities?

Table 1: Example dialogue from our dataset showing users tend to describe their activities progressively across multiple turns and the system ask for missing information incrementally and calculate the carbon footprint when all information is provided.

Metric	Value
Total Dialogues	1,028
Unique Activities	196
Avg. Turns / Dialogue	6.96 (± 3.47)
Avg. Activities / Dialogue	2.64 (± 0.82)
Avg. COe / Dialogue	16.65 (± 13.82) kg

Table 2: Main statistics of the dataset

cation (20% of the dataset) with LLM-based validation (80%), examining activity compatibility, parameter reasonableness, and calculation accuracy. The LLM-based validation achieves 89% consistency with human evaluators on a test set of 50 dialogues. Overall, our process filters out approximately 10% of initial dialogues, significantly enhancing dataset quality.

3.4 Dataset Statistics

Table 2 presents the key statistics of our dataset. The final dataset comprises 1,028 multi-turn conversations distributed across seven regions (New Zealand, United Kingdom, Germany, France, United States, China, and Global). Figure 1 shows the distribution of activities across six major categories. Food & Beverages represents the largest category with 784 activities, followed by Consumer Goods (598) and Transportation (528). For average emissions, Energy and Services categories contribute the highest average emissions.

The comprehensive annotation and diverse activity coverage make this dataset valuable for advancing research in conversational carbon footprint estimation. The dataset will be made publicly available to support further research in this important

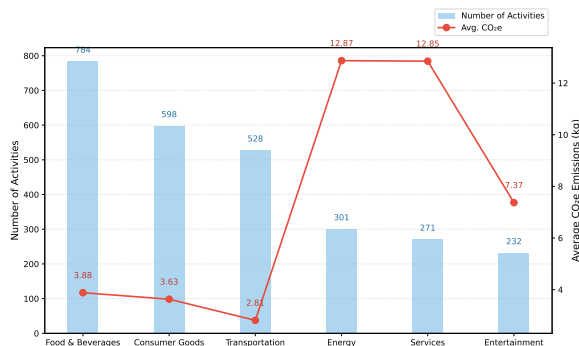


Figure 1: Category-wise distribution of activities and their carbon footprint. The bars (blue) show the number of activities in each category, while the line (red) shows the average CO₂e emissions per activity.

area².

4 Method

Our approach addresses the challenges of carbon footprint calculation through natural conversations by introducing a Progressive Contextual Carbon Tracking (PCCT) framework. Figure 2 presents our framework architecture.

The PCCT framework consists of three components that work together to process multi-turn conversations about daily activities: (1) Knowledge-Guided Progressive Activity Extraction, which identifies user activities while leveraging emission factor knowledge to guide the extraction process; (2) Context-Aware Memory Management, which maintains structured representations of activities and parameters across conversation turns; and (3)

²Dataset available at: <https://github.com/shuqinlee/Chat2CarbonFootprint.git>

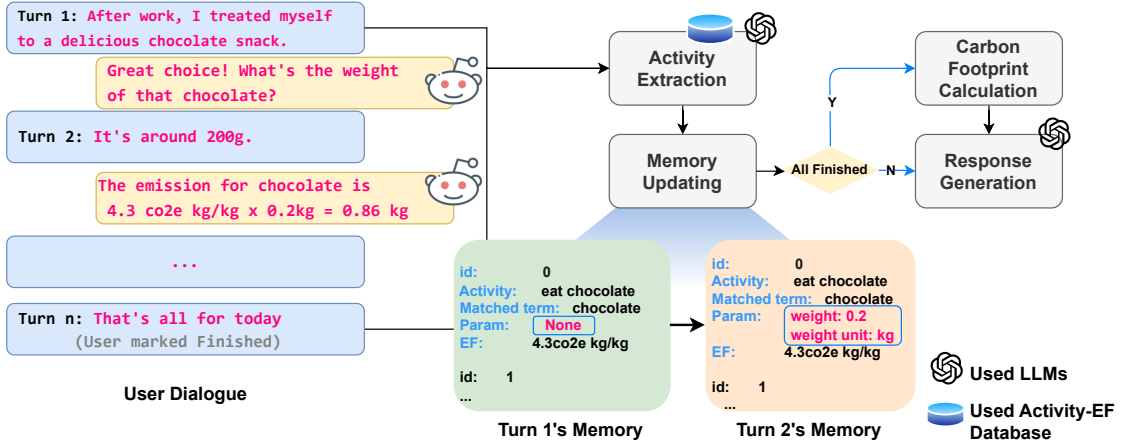


Figure 2: Overview of our Progressive Contextual Carbon Tracking (PCCT) framework. The system processes user inputs through three main components: (1) Activity Extraction, which identifies activities and parameters guided by Activity-EF Database, (2) Memory Updating, which maintains a context memory of previously extracted activities and their parameters, enabling incremental updates and refinements as new information becomes available, (3) Carbon Footprint Calculation.

Carbon Footprint Calculation

4.1 Knowledge-Guided Progressive Activity Extraction

A key challenge in conversational carbon footprint calculation is that users tend to describe their activities progressively across multiple turns, often with varying levels of detail and precision. For example, a user might initially mention "driving to work" and later specify "in an electric car for 15 kilometers." This natural communication pattern creates the challenge of correctly identifying and updating activities with their parameters across multiple turns. Our knowledge-Guided approach addresses these challenges through a bidirectional interaction between LLMs and activity-emission factor database.

We develop a progressive extraction process where emission factor knowledge guides parameter identification (Figure 3). The system first uses LLMs to recognize activity descriptions from user utterances. These descriptions are then matched against our emission factor database, which identifies the most relevant emission factors along with their required parameters. For example, if a user mentions "driving to work," the system identifies potential emission factors that might require parameters such as distance and passengers.

4.1.1 Efficient Emission Factor Retrieval

A critical support component for our knowledge-guided approach is an efficient retrieval system that

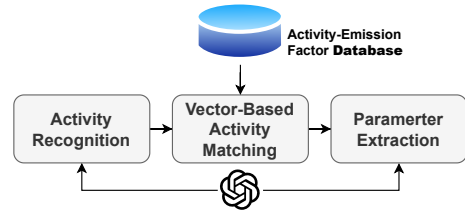


Figure 3: Knowledge-Guided Progressive activity extraction process. The system maintains awareness of previously extracted activities and their parameters, enabling it to identify both new activities and updates to existing ones.

can quickly match described activities with appropriate emission factors. Matching activities with emission factors from large databases is typically time-consuming and requires professional expertise (Balaji et al., 2023). To address this challenge, we implement a vector-based retrieval system using FAISS library (Douze et al., 2024).

The retrieval system operates in two stages. Firstly, we construct offline index by pre-computing dense vector representations for all emission factors in our database, incorporating both activity descriptions and their associated parameters. Next, during conversation, extracted activities are encoded into the same vector space and matched against the pre-built indices. This enables sub-second retrieval of relevant emission factors, even from databases containing thousands of entries which guides the subsequent parameter extrac-

tion stage.

4.2 Context-Aware Memory Management

To maintain a structured representation of all extracted information across turns, our memory module maintains two interconnected components: 1) Activity Registry maintains all identified activities with their emission factor mappings, tracking the state of each activity (complete or partially specified); 2) Parameter Registry tracks all extracted parameters for each activity and maintains all missing information. This including their values, units, extraction confidence, and source turn information.

In each turn, the activity extraction module identifies incremental activity information and updates the memory rather than creating duplicates. It provides a complete view of missing parameters, enabling targeted follow-up questions; and it maintains extraction confidence scores that help prioritize which missing information is most critical to obtain.

4.3 Carbon Footprint Calculation

The final component of our framework transforms the structured activity and parameter information into accurate carbon footprint calculations. Once the system has collected sufficient information about the user’s activities, it applies the appropriate emission factors to calculate the carbon footprint. For each activity, the system applies the appropriate emission factor to the validated parameters, calculating the carbon footprint using the formula: $CO_2e = \sum_{i=1}^n P_i \times EF_i$ where P_i represents the activity parameter (e.g., distance traveled) and EF_i is the corresponding emission factor.

After the calculation for each activity, the system generates detailed explanations of the calculation process, breaking down the contribution of each activity to the total carbon footprint and gives advice.

5 Objective Experiments

We conduct objective experiments to evaluate the effectiveness of our PCCT framework against baseline approaches. Our evaluation focuses on three key aspects: activity recognition accuracy, parameter extraction completeness, and carbon footprint calculation accuracy.

5.1 Experimental Setup

Dataset. We evaluate all systems using our CarbonDialog-1K dataset. For testing, we use a

randomly selected subset of 560 dialogues, ensuring balanced coverage across activity categories and regions. The remaining dialogues are used for system development and parameter tuning.

Metrics. We assess system performance using several key metrics. For activity recognition, we measure the F1-score for correctly identifying activities from user descriptions. Parameter extraction performance is evaluated using both the F1-score for parameter identification and the Missing Critical Parameter Rate (MCPR) for missing critical parameters. Finally, to assess emission calculation accuracy, we compute the Mean Absolute Error (MAE) in kg CO₂e and Mean Absolute Percentage Error (MAPE).

Baseline System. Our goal is to estimate personal carbon footprint (PCF) from open-ended, multi-domain text-based dialogues. To evaluate this, we compare PCCT with a rule-based calculator (**RBC**) that uses keyword matching and fixed templates to compute emissions from structured inputs. This reflects conventional carbon tools based on forms or explicit prompts. GPS-based trackers focus mainly on transport, require continuous sensing and device integration, and cover a limited range of activities. As they are not comparable to general-purpose, text-based dialogue systems, we exclude them from our evaluation.

Implementation Details. Our PCCT implementation uses a combination of LLMs and vector similarity search. For the LLM component, we employ the DeepSeek-V3 model (Liu et al., 2024) for activity extraction and parameter identification. The vector similarity search uses FAISS (Douze et al., 2024) with 768-dimensional embeddings from sentence transformer (Reimers and Gurevych, 2019) for emission factor retrieval.

5.2 Results and Analysis

5.2.1 Overall Performance

Table 3 presents the overall performance comparison between our PCCT system and the rule-based baseline.

The results demonstrate that PCCT significantly outperforms the rule-based approach across all metrics. Our system achieves a substantial improvement in activity recognition and parameter extraction. This can be attributed to the knowledge-augmented activity extraction that effectively matches user descriptions with standard-

System	Activity F1 (%)	Parameter F1 (%)	MCPR (%)	MAE (kg CO ₂ e)	MAPE (%)
Rule-Based Calculator	30.8	18.5	45.0	9.7	74.3
PCCT (Ours)	74.2	75.5	15.5	4.8	39.6

Table 3: Overall performance comparison of PCCT vs. Rule-based Calculator. Our system outperforms the rule-based calculator across all metrics. MCPR reveals the missing rate of critical parameter. MAPE reveals relative error of carbon footprint.

Category	MAPE (%)	
	RBC	PCCT
Transportation	55.8	28.5
Food & Beverages	68.5	42.3
Consumer Goods	95.8	45.8
Entertainment	85.5	38.5
Services	115.6	58.6
Energy	95.4	65.4

Table 4: Category-level performance comparison sorted by MAPE (lower values indicate better performance)

ized emission factors, (2) the context-aware memory management that maintains coherent information across turns, and (3) the progressive parameter-gathering strategy that ensures the completeness of critical information.

The improved activity recognition and parameter extraction directly translate to calculation accuracy, with PCCT achieving a 50% lower Mean Absolute Error (MAE: 4.8 vs 9.7 kg CO₂e) and a reduction of 34.7 percentage points in Mean Absolute Percentage Error (MAPE: 39.6% vs 74.3%). Importantly, while the rule-based system can only attempt calculations for the small subset of activities where it successfully extracts all parameters (18.5% of cases), PCCT maintains high accuracy while handling a much broader range of activities and conversation patterns.

5.2.2 Category-level Performance

A detailed analysis of performance across different activity categories reveals significant variations in calculation accuracy, as shown in Table 4.

Our analysis reveals several important patterns that highlight the effectiveness of our knowledge-augmented approach:

Transportation activities show the highest accuracy for both systems, with PCCT achieving the lowest MAPE (28.5%). The well-defined parameter requirements (primarily distance) benefit from our progressive extraction approach, allowing the system to focus on gathering specific, critical information.

Food & Beverages and Consumer Goods show moderate error rates with PCCT (MAPE: 42-46%). These categories represent the most frequent activities in our dataset, they demonstrate the value of our retrieval system.

Energy and Services present the greatest challenges for both systems, with the highest error rates even with PCCT (MAPE: 58-65%). These categories involve inducing parameters that the users may not themselves know, and the system may not have access to. For example, the amount of electricity consumed by a service may not be known, therefore this brings in more uncertainty.

The performance gap between PCCT and RBC is consistent across all categories, with PCCT reducing MAPE by 48-57%. Note that the rule-based system’s MAPE values are only applicable to the small subset of activities where it successfully extracted all critical parameters (Activity F1: 30.8%, Parameter F1: 18.5%). Therefore, the performance gap between PCCT and RBC is even larger for these categories.

These results demonstrate that while certain activity types remain challenging for carbon footprint calculation, PCCT’s integrated approach of knowledge-augmented extraction, context-aware memory, and progressive parameter gathering significantly improves accuracy across all categories.

6 Subjective Experiments

To further evaluate the effectiveness of our system in comparison to the baseline system carbon footprint calculation methods, we conducted a controlled subjective experiment.

6.1 Experiment Design

We recruited 20 participants, including environmental enthusiasts, general users, and sustainability researchers, to calculate their carbon footprint for 10 predefined daily activities of varying complexity. Participants used our system, a conversational interface that guided them through activity details, resolved incomplete information, and provided carbon footprint estimates with confidence bounds.

They also used traditional tools like spreadsheets or online calculators requiring manual data entry. Participants rated their experience on a 7-point Likert scale for usability (Zwakman et al., 2020), appropriateness (Torrey et al., 2013; Peng et al., 2019), efficiency (Siro et al., 2022), and naturalness (CAO et al., 2023), and provided qualitative feedback through semi-structured interviews. Task completion time, result accuracy, and user engagement metrics (e.g., errors, frustration) were recorded.

6.2 Results

Usability. Our system received higher usability ratings (mean = 4.6, SD = 0.5) than the baseline system (mean = 3.2, SD = 0.8). Users praised the intuitive interface of our system, while the baseline system users found data entry frustrating.

Appropriateness. Both methods scored similarly (Our system: mean = 4.7, SD = 0.4; The baseline system: mean = 4.5, SD = 0.6), though our system’s context-aware explanations were appreciated.

Efficiency. Our system was more efficient (mean = 4.5, SD = 0.5) than the baseline system (mean = 2.8, SD = 0.9), with task completion times of 12 minutes vs. 28 minutes.

Naturalness. Our system scored higher on naturalness (mean = 4.8, SD = 0.3) compared to the baseline system (mean = 2.5, SD = 0.7), with its conversational interface being a key factor.

6.3 Statistical Analysis

A two-way ANOVA confirmed significant differences between methods ($F(1, 236) = 45.3, p < 0.001$) and dimensions ($F(3, 236) = 12.7, p < 0.001$), with an interaction effect ($F(3, 236) = 8.2, p < 0.001$). Post-hoc tests showed our system outperformed the baseline system in usability, efficiency, and naturalness ($p < 0.001$), but not in appropriateness ($p > 0.05$).

6.4 Qualitative Feedback

Participants praised our system for its intuitive, conversational interface and transparency, with one noting, "It was so easy to use—I didn’t need prior knowledge of carbon emissions," and another appreciating its guidance and confidence-bound estimates. In contrast, the baseline system users found data entry and emission factor lookup tedious and error-prone, with comments like, "I spent more time looking up emission factors than calculating,"

and frustration over the lack of guidance. Many wished for explanatory features similar to our system. Feedback highlighted our system’s strengths in usability, transparency, and engagement while revealing inefficiencies in rule-based system.

7 Conclusion

We introduce PCCT, a novel framework for calculating personal carbon footprints through natural conversations. By integrating knowledge-guided activity extraction, context-aware memory management, and progressive parameter gathering, PCCT bridges the gap between casual dialogue and precise carbon footprint estimation. Experimental results demonstrate PCCT’s significant advantages over traditional approaches, achieving substantial improvements in both automated understanding and calculation accuracy. Our subjective experiments further validate PCCT’s practical value, with users particularly appreciating its intuitive interface and conversational guidance. Our work establishes a promising direction for making carbon footprint estimation more accessible through natural interactions.

Limitations

While PCCT demonstrates promising results in personal carbon footprint estimation, several limitations should be acknowledged:

First, the accuracy of our system heavily depends on the quality and coverage of the emission factor database. For emerging or region-specific activities, the system may struggle to find appropriate emission factors, potentially leading to less accurate calculations. This limitation is particularly evident in service-related activities where standardized emission factors are often lacking. Second, our evaluation, while comprehensive, is limited to a curated benchmark dataset and controlled user studies. Real-world deployment may present additional challenges, such as handling extremely rare activities and adapting to regional variations in emission factors, and maintaining consistent performance across different user demographics. Third, the PCCT pipeline relies on LLM inference, which itself incurs a non-negligible carbon cost. Future work will fine-tune an expert model to a 7 B parameter variant, further reducing inference emissions.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (U23B2057, 62176185), and the Youth Fund Special Project at Hangzhou Dianzi University Information Engineering College (KYP0323010).

References

- Amma Asantewaa Agyei Boakye, Terrie Boguski, Sarah Cashman, Christoph Koffler, Ashley Kreuder, Manish Kumar, Naveen Kumar Vipparla, and Lisa Peterson. 2023. At the intersection of life cycle assessment and indirect greenhouse gas emissions accounting. *The International Journal of Life Cycle Assessment*, 28(4):321–335.
- Bharathan Balaji, Venkata Sai Gargeya Vunnava, Geoffrey Guest, and Jared Kramer. 2023. Caml: Carbon footprinting of household products with zero-shot semantic text similarity. In *Proceedings of the ACM Web Conference 2023*, pages 4004–4014.
- Xiao Bi, Deli Chen, Guanting Chen, Shanhuang Chen, Damai Dai, Chengqi Deng, Honghui Ding, Kai Dong, Qiushi Du, Zhe Fu, et al. 2024. Deepseek llm: Scaling open-source language models with longtermism. *arXiv preprint arXiv:2401.02954*.
- Jianqin CAO, Jingyu ZHANG, Liang ZHANG, and Xiaoyu WANG. 2023. The psychological structure and influence of interactive naturalness. *Acta Psychologica Sinica*, 55(1):55.
- Chen Chen, Zengfeng Zhao, Jianzhuang Xiao, and Robert Tiong. 2021. A conceptual framework for estimating building embodied carbon based on digital twin technology and life cycle assessment. *Sustainability*, 13(24):13875.
- Shimon Chen, Shlomo Bekhor, David M Broday, et al. 2016. Aggregated gps tracking of vehicles and its use as a proxy of traffic-related air pollution emissions. *Atmospheric Environment*, 142:351–359.
- Zhu Deng, Jinjie Liu, Biao Luo, Can Yuan, Qingrun Yang, Lei Xiao, Wenwen Zhou, and Zhu Liu. 2023. Autopcf: Efficient product carbon footprint accounting with large language models. *arXiv preprint arXiv:2308.04241*.
- Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvassy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. 2024. The faiss library. *arXiv preprint arXiv:2401.08281*.
- LJM Dreijerink and GL Paradies. 2020. How to reduce individual environmental impact? a literature review into the effects and behavioral change potential of carbon footprint calculators. *TNO Report*, 11148.
- Conor K Gately, Lucy R Hutyra, Scott Peterson, and Ian Sue Wing. 2017. Urban emissions hotspots: Quantifying vehicle congestion and air pollution using mobile phone gps data. *Environmental pollution*, 229:496–504.
- Chris Hendrickson, Arpad Horvath, Satish Joshi, Octavio Juarez, Lester Lave, H Scott Matthews, Francis C McMichael, and Elisa Cobas-Flores. 1998. Economic input-output-based life cycle assessment (eio-ica). *mental*.
- Loïc Lannelongue, Jason Grealey, and Michael Inouye. 2021. Green algorithms: quantifying the carbon footprint of computation. *Advanced science*, 8(12):2100707.
- Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. 2024. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.
- Maria Serena Mancini, Alessandro Galli, Valentina Nicolucci, David Lin, Simone Bastianoni, Mathis Wackernagel, and Nadia Marchettini. 2016. Ecological footprint: refining the carbon footprint calculation. *Ecological indicators*, 61:390–403.
- Ben Mann, N Ryder, M Subbiah, J Kaplan, P Dhariwal, A Neelakantan, P Shyam, G Sastry, A Askell, S Agarwal, et al. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 1:3.
- Jérôme Mariette, Odile Blanchard, Olivier Berné, Olivier Aumont, Julian Carrey, AnneLaure Ligozat, Emmanuel Lellouch, Philippe-Emmanuel Roche, Gaël Guennebaud, Joel Thanwerdas, et al. 2022. An open-source tool to assess the carbon footprint of research. *Environmental Research: Infrastructure and Sustainability*, 2(3):035008.
- Bo Miao, Jie Wen, Qiujiu Yuan, Kaisaier Aisikaer, Daming Zhao, and Zhao Zhao. 2024. Research on carbon footprint analysis method of a-class single-phase cost-controlled intelligent energy meter product. In *2024 8th International Conference on Electrical, Mechanical and Computer Engineering (ICEMCE)*, pages 558–563. IEEE.
- John B Murphy. 1998. Introducing the north american industry classification system. *Monthly Lab. Rev.*, 121:43.
- Christine Oehlert, Evan Schulz, and Anne Parker. 2022. Naics code prediction using supervised methods. *Statistics and Public Policy*, 9(1):58–66.
- Zhenhui Peng, Yunhwan Kwon, Jiaan Lu, Ziming Wu, and Xiaojuan Ma. 2019. Design and evaluation of service robot’s proactivity in decision-making support process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.
- Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.
- Flavio Scrucca, Grazia Barberio, Valentina Fantin, Pier Luigi Porta, and Marco Barbanera. 2021. Carbon footprint: Concept, methodology and calculation. *Carbon Footprint Case Studies: Municipal Solid Waste Management, Sustainable Road Transport and Carbon Sequestration*, pages 1–31.
- Clemencia Siro, Mohammad Aliannejadi, and Maarten de Rijke. 2022. Understanding user satisfaction with task-oriented dialogue systems. In *Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval*, pages 2018–2023.
- Efisio Solazzo, Monica Crippa, Diego Guizzardi, Marilena Muntean, Margarita Choulga, and Greet Janssens-Maenhout. 2021. Uncertainties in the emissions database for global atmospheric research (edgar) emission inventory of greenhouse gases. *Atmospheric Chemistry and Physics*, 21(7):5655–5683.
- Ines Sousa and David Wallace. 2006. Product classification to support approximate life-cycle assessment of design concepts. *Technological Forecasting and Social Change*, 73(3):228–249.
- Cristen Torrey, Susan R Fussell, and Sara Kiesler. 2013. How a robot should give advice. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 275–282. IEEE.
- Qingshi Tu, Jing Guo, Nan Li, Jianchuan Qi, and Ming Xu. 2024. Mitigating grand challenges in life cycle inventory modeling through the applications of large language models. *Environmental Science & Technology*, 58(44):19595–19603.
- Yushi Wang, Jonathan Berant, and Percy Liang. 2015. [Building a semantic parser overnight](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1332–1342, Beijing, China. Association for Computational Linguistics.
- Thomas Wiedmann and Jan Minx. 2008. A definition of ‘carbon footprint’. *Ecological economics research trends*, 1(2008):1–11.
- Sam Wood, Rohit Muthyala, Yi Jin, Yixing Qin, Nilaj Rukadikar, Amit Rai, and Hua Gao. 2017. Automated industry classification with deep learning. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 122–129. IEEE.
- Dilawar Shah Zwakman, Debajyoti Pal, Tuul Triyason, and Chonlameth Arpnikanondt. 2020. Voice usability scale: measuring the user experience with voice assistants. In *2020 IEEE International Symposium on Smart Electronic Systems (iSES)(Formerly iNiS)*, pages 308–311. IEEE.