Academic Research Journal of Science and Technology
DOI: https://doi.org/10.63300/arjst0507112026.04
Online at: publications.ngmc.ac.in
Vol. 5, No. 07 (2026)

AI-Based Personal Carbon Footprint Tracker

Jogi John[1,*], Janhvi Mahakulkar[2], Harsh Sharma[2], Samiksha Dayare[2], Aditi Shende[2], Anjali Kushwaha[2]

[1] Assistant Proffesor

[2] Student

***Corresponding author:** Jogi John, Jogi.jhon@pcenagpur.edu.in

Abstract

This paper presents the design and prototype of an intelligent web platform for personal carbon-footprint tracking. The platform enables users to record everyday activities, estimate associated greenhouse-gas emissions in kilograms of carbon-dioxide equivalent (kg CO_2e), review category-wise results, and receive personalised sustainability guidance. The architecture comprises a React, Tailwind CSS and Vite front end, a PHP back end, a MySQL database, and a large-language-model service accessed through the Groq API. The implemented calculation engine estimates emissions by multiplying each activity quantity by its stored emission factor and aggregates results by category and reporting period. A pilot dataset containing records from 10 users over September 1–30, 2026, included 14 logged activities and a reported total of 112.51 kg CO_2e. The mean across all 10 users was 11.25 kg CO_2e per user; the reported mean among active users was 22.50 kg CO_2e. The consumption category accounted for 53.3% of the reported total. A hand-calculation check matched the system output for all 14 samples. These findings are descriptive and limited by the small sample, short observation period, and use of illustrative emission factors. Random Forest forecasting, K-means user profiling, month-to-month comparisons, and leaderboard scoring are proposed extensions and have not been validated. The prototype demonstrates a transparent activity-based calculation workflow, while larger datasets, verified emission factors, and systematic evaluation of AI-generated advice are needed before broader claims can be made.

Keywords: carbon footprint tracking; greenhouse-gas emissions; large language models; sustainability; emission factors; web application; machine learning

Received: 23 October 2026 · Accepted: 5 November 2026 · Published: 1 November 2026

How to cite this article: John, J., Mahakulkar, J., Sharma, H., Dayare, S., Shende, A., & Kushwaha, A. (2026). AI-Based Personal Carbon Footprint Tracker. *Academic Research Journal of Science and Technology*, *5*(07). https://doi.org/10.63300/arjst0507112026.04

Copyright and licence: © 2026 The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Climate change is driven by the accumulation of greenhouse gases (GHGs) in the atmosphere, and according to the International Energy Agency, global energy-related CO_2e emissions reached a record 37.4 Gt in 2023. Although industry and power generation dominate national inventories, household consumption (commuting, electricity, food, shopping, waste and water) contributes a substantial share, so individual choices matter. Many mitigation behaviours also reduce household costs, which makes personal action both an environmental and an economic opportunity.

Yet individuals rarely know which of their decisions matter most. Online carbon calculators try to close this gap, but reviews report that they are inconsistent in how closely they follow IPCC guidance and are often designed for a single context. Most are one-off questionnaires that return a single annual number and show nothing about how behaviour changes from month to month. Studies of carbon-tracking apps also show that uptake depends on whether the tool is easy and rewarding to keep using, which calls for history, feedback and some form of motivation.

Artificial Intelligence (AI) can strengthen such a tool in three complementary ways. Machine-learning models can predict where a user's emissions are heading and group users with similar emission profiles. Large language models (LLMs) can turn structured activity data into plain-language, personalised tips and can answer user questions about sustainability. Lowlatency LLM inference matters because the advice is shown inside an interactive application. There is also a design tension: AI has a carbon cost of its own, and deep-learning research in recommender systems has been found to emit far more CO_2e than research using traditional models. For the predictive part of the system we therefore prefer lightweight, well-understood methods: Random Forest and K-means. To address these needs, this paper introduces the AI-Based Carbon Footprint Tracker, a web platform with a React, Tailwind CSS and

Vite front end, a PHP back end, a MySQL database, machine-learning analytics and an LLM service accessed through the Groq API. The paper presents the system flow, mathematical formulation and a synthetic-data evaluation of the analytics.

2. Review of Literature

A review of recent literature shows steady progress in individual components of carbon tracking, while highlighting the absence of an integrated platform that combines activity tracking, prediction, user profiling, LLM-based guidance and engagement features.

1.Behaviour-Oriented Carbon Apps – Jasmy et al. (2024) developed Tanaffas, which combines GPS trip tracking, augmented-reality item scanning, a chatbot and a points-based reward system on a Firebase back end. A 30-day study with ten users found a significant rise in subjective norms ($t(9) = 4.2$, $p = 0.002$) but not in attitudes or behavioural intentions. Its calculator is a simple arithmetic step and it includes no forecasting.

2.Behavioural-Theory-Based Systems – Lin (2016) introduced the Personal Carbon Footprint Management System (PECAFOMS) for high-school students, using persuasive technology and behavioural theory. The educational focus is valuable, but the narrow audience and reliance on user motivation limit wider applicability.

3.Adoption of Carbon-Tracking Apps – Hoffmann et al. (2022) investigated what drives consumers' intention to adopt carbon-footprint tracking apps, showing that engagement, and not only calculation accuracy, decides whether a tool is used at all.

4.Carbon Footprint Calculators – Peter et al. (2017) reviewed footprint calculators (in an agricultural-energy context) and highlighted inconsistent adherence to IPCC guidelines and context-specific designs, which weakens the reliability of their results.

5.Emission Prediction with Dynamic Coefficients – Yu et al. (2023) predicted community electricity-related emissions using a dynamic emission coefficient of the power system, showing that emission factors can vary over time. The work operates at community level and does not engage individual users. **6.Machine Learning for Individual Action** – Rolnick et al. (2019) outline how machine learning can estimate an individual's footprint from personal data and model behaviour to identify high-impact actions.

7.Gamification for Pro-Environmental Behaviour – Ouariachi et al. (2020) analysed six gamified platforms and concluded that gamification can help educate people and motivate pro-environmental action when designs combine extrinsic rewards with intrinsic motivators. The study is qualitative and based on a small sample, so its conclusions are tentative.

8.LLMs for Climate Action – Sgaramella and Bilancini describe a planned online experiment in which an LLM-based chatbot helps users choose lower-impact foods by addressing both motivation and practical barriers; the project description reports no results yet. On the engineering side, the Groq documentation states that its API is mostly compatible with OpenAI's client libraries, so existing applications can be pointed at it.

9.Emerging AI-Based Trackers – EcoPilot and CARBONLY describe AI-assisted trackers with personalised tips, analytics, progress tracking and chatbots. Evaluation details were not reported in the descriptions we could access.

10. Carbon Cost of AI – Wegmeth et al. (2025) analysed 79 papers from ACM RecSys 2013 and 2023 and found that deeplearning papers emit roughly 42 times more CO_2e than traditional-model papers, which motivates lightweight methods where they are sufficient.

Existing research addresses several parts of personal carbon management, including emission estimation, behavioural engagement, activity monitoring and automated guidance. However, these functions are not always available together in a single workflow. This project brings them together by combining activity-based CO_2e calculation with monthly trend analysis, emission forecasting, user profiling and personalised sustainability recommendations. The integrated design is intended to help users understand not only their current environmental impact but also which changes may reduce their future emissions.

1.1 Problem Statement

Contemporary personal carbon tools suffer from five weaknesses: (i) static, one-time calculation without activity history or trends; (ii) generic rather than personalised advice; (iii) no forward-looking view of where emissions are heading; (iv) weak long-term engagement; and (v) inconsistent or opaque emission factors.

There is therefore a need for a web-based tracker that records activities across everyday categories, calculates emissions transparently, shows history and trends, predicts next-month emissions, groups users by emission profile, generates personalised LLM-based recommendations, and motivates users through a leaderboard and badges.

1.2 Objectives

The primary goal of this project is to design and validate an AIdriven carbon footprint tracker following the system flow in Fig. 1. The specific objectives are:

1.Activity Tracking and Calculation – to let users add activities in seven categories (transport, electricity, food, shopping, waste, water and others) and to convert them into CO_2e using a stored emission-factor table.

2.History, Insights and Reports – to present daily, weekly and monthly history, category-wise breakdowns, trend graphs and comparisons on a dashboard.

3.Prediction and Profiling – to predict next-month emissions with a Random Forest and to group users by emission profile with K-means clustering.

4.LLM-Based Personalised Guidance– to generate personalised recommendations, tips, savings estimates, activity insights and answers to user queries through the Groq API, and to store the responses.

5. Engagement and Secure Design – to motivate users with a leaderboard and badges, and to protect user data and the API key.

3. Methodology

The proposed system follows the flow shown in Fig. 1. After the splash screen and authentication, the user reaches the dashboard and then the feature modules. The React front end calls the PHP API endpoints; the business logic layer calculates emissions and applies validation; database operations read and write MySQL; and the AI services return recommendations that are stored with the other records. Fig. 1 shows the complete system flow diagram, the technology stack and the database tables.

The methodology is described layer by layer, following the diagram:

1. **Frontend Modules and Navigation** – The front end is built with React, Tailwind CSS and Vite and has ten modules. (1) Splash / Welcome greets the user. (2) Authentication offers login, sign-up and a forgot password option. (3) Dashboard (Home) shows the total carbon footprint in kg CO_2e, a donut chart of category shares and the change from last month. (4) Add Activity records transport, electricity, food, shopping, waste, water and other activities. (5) Activity History lists daily, weekly and monthly records. (6) Insights & Reports provides category-wise breakdowns, trend graphs and comparisons. (7) AI Insights & Recommendations shows AI suggestions, personalised tips and potential savings. (8) Prediction shows the predicted footprint for next month. (9) Profile & Settings holds personal information, preferences, notifications and the Groq API key. (10) Leaderboard & Badges shows rankings and earned badges. The numerical values drawn in Fig. 1 (for example 247 kg CO_2e and 220 kg CO_2e) are placeholders of the interface design and not measured results.

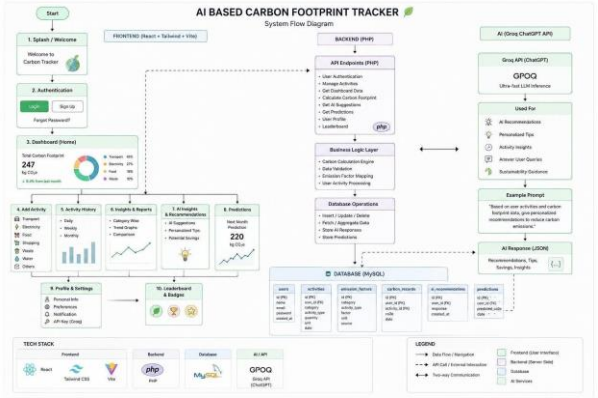


Fig. 1. System Flow Diagram of the AI-Based Carbon Footprint Tracker

1. **PHP API Endpoints** – The front end sends requests to the PHP back end, which exposes endpoints for user authentication, managing activities, getting dashboard data, calculating the carbon footprint, getting AI suggestions, getting predictions, the user profile and the leaderboard.

3.Business Logic Layer – This layer contains the *carbon calculation engine*, *data validation* (types, ranges and required fields), *emission factor mapping* (matching each activity to a factor in the *emission_factors* table) and *user activity processing* (aggregation by category and period).

4.Database Operations and Schema – Operations are insert, update and delete, fetch and aggregate data, store AI responses and store predictions. The MySQL database has six tables: *users* (id, name, email, password, created_at); *activities* (id, user_id, category, activity_type, quantity, unit, date); *emission_factors* (id, category, activity_type, factor, unit, source); *carbon_records* (id, user_id, activity_id, CO_2e, date); *AI_recommendations* (id, user_id, response, created_at); and *predictions* (id, user_id, predicted_CO_2e, date). Foreign keys link records to users and activities.

5.AI Services through the Groq API – The back end communicates in both directions with the Groq API, which provides low-latency LLM inference through an interface that is mostly compatible with OpenAI's client libraries. It is used for AI recommendations, personalised tips, activity insights, answers to user queries and sustainability guidance. The back end sends a prompt such as "Based on user activities and carbon footprint data, give personalised recommendations to reduce carbon emissions" together with the user's summarised data, receives a JSON response containing recommendations, tips, savings and insights, and stores it in *ai_recommendations*.

1. **Current Analysis (Implemented)** – The business logic layer converts each activity into CO_2e and aggregates by category and by day, week or period within the recorded month. It produces the total

footprint, the category shares shown in the donut chart, and the savings estimate when an activity is swapped for a lower-factor alternative. These are descriptive analyses of past activity. They are not predictions.

1. **Proposed Forecasting and Profiling (Not Yet Validated)** – The following are design proposals for when sufficient data exists. Random Forest regression would predict next-month total emissions from a user's monthly history, with features such as recent monthly totals, category shares and a seasonal encoding. It needs several months per user to train and a timeordered split (train on earlier months, test on later ones); the features, hyperparameters and validation protocol will be fixed when data allow. K-means clustering would group users by category-share profile. It requires a larger user base, and the number of clusters would be chosen by silhouette score.

3.1 Mathematical Model

1.Carbon Footprint Calculation – **Purpose:** Converts each logged activity into CO_2e (carbon calculation engine).

$$E_c = \Sigma_{i \in c}\, q_i \cdot f_i,\; E = \Sigma_c\, E_c \quad (1)$$

How it works: q is the quantity of activity i (km, kWh, kg or litres) and f is its factor in kg CO_2e per unit, taken from *emission_factors*. The product is saved in *carbon_records*. For example, with illustrative factors, 15 km by petrol car (0.17 kg/km) gives 2.55 kg and 100 kWh of grid electricity (0.73 kg/kWh) gives 73 kg.

2.Category Share, Monthly Change and Potential Savings – **Purpose:** Produces the dashboard percentages, the change from last month and the savings estimate.

$p_c = (E_c / E) \times 100$	(2)
$\Delta_t = (E_t - E_{t-1}) / E_{t-1} \times 100$	(3)
$\Delta E = q \cdot (f\text{current} - f\text{alternative})$	(4)

How it works: p_c is the share of category c in the donut chart; Δ is the percentage change in the monthly total (a negative value means a reduction); and ΔE is the potential saving when an activity is replaced by a lower-factor alternative. **Status:** p_c and ΔE are implemented analyses of the recorded month; Δ is a design-only formula and needs at least two months of data.

3.K-means Clustering (Proposed, requires multi-user data) – **Purpose:** Groups users with similar emission profiles.

$$\min \Sigma_k \Sigma_{x \in Sk} \|x - \mu_k\|^2 \quad (5)$$

$$s(i) = (b(i) - a(i)) / \max\{a(i), b(i)\} \quad (6)$$

How it works: Each user would be described by the seven category shares and the mean monthly total, standardised. Kmeans assigns users to k clusters by minimising the withincluster squared distance to the centroids μ_k, and k is chosen by the highest mean silhouette value $s(i)$, where $a(i)$ is the mean distance to points in the same cluster and $b(i)$ the smallest mean distance to another cluster. This model is proposed and has not been validated on the 10-user pilot data.

4.Random Forest Next-Month Prediction (Proposed, requires multi-month data) – **Purpose:** Would predict next month's footprint for the Predictions module.

$$\hat{E}t+1 = (1/B)\, \Sigma b\, Tb(xt) \quad (7)$$

How it works: A forest averages the predictions of B decision trees T_b trained on bootstrap samples with random feature subsets. The feature vector x_t would be built from t recent monthly totals, category shares and a seasonal encoding of the calendar month. The number of trees, tree depth and exact features will be fixed only when enough monthly data exist for time-ordered training and validation. No trained model or accuracy result is reported in this paper.

5.LLM Recommendation Generation – **Purpose:** Turns the user's data into personalised guidance.

$$R = LLM(\, Prompt(\, Ec, pc, \Delta, \hat{E}t+1, cluster\,)\,) \quad (8)$$

How it works: The business logic layer builds a prompt from the user's category totals and shares, without names or e-mail addresses. The terms Δ, $\hat{E}_{t+1}$ and cluster label are included only once the corresponding modules are implemented and validated. The Groq-hosted model returns JSON with recommendations, tips, savings and insights; the response is validated, stored in *ai_recommendations* and displayed in the AI Insights module. Savings stated by the model can be crosschecked with the calculated ΔE.

6. Leaderboard and Badge Score (Design only, requires ≥2 months) – **Purpose:** Ranks users fairly across different lifestyles.

$$Su = (Ebase - Ecurrent) / Ebase \times 100 \quad (9)$$

How it works: The score is the percentage reduction relative to the user's own baseline month, so a user who starts with a high footprint is not disadvantaged. Badges are awarded when S passes defined thresholds. The score and thresholds u are design choices of this prototype, and a baseline and a later month are both needed to compute them.

7.Planned Evaluation Metrics – **Purpose:** Will measure prediction error on held-out months once time-ordered validation on real data exists.

$$MAE = (1/n)\, \Sigma\, |y - \hat{y}| \quad (10)$$

$$RMSE = \sqrt{(\, (1/n)\, \Sigma\, (y - \hat{y})^2\,)} \quad (11)$$

$$MAPE = (100/n)\, \Sigma\, |y - \hat{y}| / y \quad (12)$$

$$R^2 = 1 - \Sigma\, (y - \hat{y})^2 / \Sigma\, (y - \bar{y})^2 \quad (13)$$

How it works: y is the actual and $\hat{y}$ the predicted monthly footprint over n test samples. MAE and RMSE are in kg CO_2e, MAPE is in percent, and R^2 measures the share of

variance explained. These metrics have **not** been computed for this paper.

4. Results

Dataset. The pilot dataset comprised activity records from 10 users over a one-month period, from September 1 to September 30, 2026. A total of 14 activities were recorded during this period. The dataset was used to examine user activity records and the corresponding carbon footprint calculations.

Experimental Result. The dataset recorded total emissions of 112.51 kg CO_2e. Across all 10 users, the mean emission was 11.25 kg CO_2e per user, with a minimum of 0.00 kg CO_2e and a maximum of 91.80 kg CO_2e. Among active users, the mean emission was 22.50 kg CO_2e per user, with a minimum of 1.92 kg CO_2e and a maximum of 91.80 kg CO_2e. The consumption category accounted for the largest reported share of emissions, at 53.3%.

TABLE 1. Descriptive Results of the Pilot Dataset

Item	Value
Users	10
Period	1 month (September 1–30, 2026)
Total activities logged	14
Total emissions (kg CO_2e)	112.51
Mean / min / max per user (kg CO_2e)	[All 10 users] 11.25 / 0.00 / 91.80 [Active users] 22.50 / 1.92 / 91.80
Largest category by share	Consumption, 53.3%
Calculation check (handcomputed vs. system output)	14 of 14 samples matched (100.0%)

Any further statement about the results, such as which category dominated, must come directly from Table 1. The earlier synthetic experiment (300 users, 36 months) is not part of the results and is not evidence about real users.

5.1 Advantages

1.Complete Tracking Workflow – Users move from adding activities to history, insights, predictions and recommendations inside one application.

2.Transparent Calculation – Every figure can be traced to an activity quantity and a stored factor in the *emission_factors* table, which can be updated without changing code.

3.Forward-Looking Insight – The Predictions module shows the expected footprint for next month, so users can act before it is reached.

4.Personalised, Natural-Language Guidance – The LLM turns the user's own data into tips, insights and answers to questions, and the responses are stored for later viewing.

5.Lightweight Analytics – Random Forest and K-means run on a CPU in milliseconds, in line with the Green-AI principle of avoiding unnecessarily heavy models.

6.Motivation through Gamification – The leaderboard and badges give a visible reward for reducing emissions.

5.2 Limitations

1.Very Small Pilot Dataset – Only 10 users and one month. The results are descriptive and cannot be generalised.

2.No Forecasting Evidence – With a single month there is no temporal training or validation data, so no forecasting capability has been demonstrated.

3.Profiling Not Validated – K-means on 10 users cannot give meaningful or stable clusters.

4.Monthly Comparison and Leaderboard Not Testable Yet – Both need at least two months of data per user.

5.Illustrative Emission Factors – Factors in the examples are order-of-magnitude values; they must be verified against DEFRA, IPCC and CEA India for the user's country and year before deployment.

6.LLM Output Quality and Dependence – LLM answers may be inaccurate, vary between calls or be unavailable if the external API or network fails; the quality of recommendations has not yet been evaluated.

7.Manual Data Entry – Self-reported activities are tedious and error-prone, and data quality affects every downstream result.

8.Key Handling – Because the interface includes an API key setting, the key must be stored and used on the server side and protected from exposure.

5. Discussion

TABLE 2. Comparative Analysis

Feature Dimension	Static Web Calculators	Existing AI / App Trackers	Proposed System
Platform Architecture	Single-page questionnaire	Mobile app with cloud back end (e.g., Firebase)	React + Tailwind + Vite front end, PHP back end, MySQL
Activity Tracking	None (one-time result)	Activity logs and progress views	Seven categories with daily, weekly, monthly history
Emission Calculation	Fixed, often opaque factors	Arithmetic or formula based	Emission-factor table mapped by a calculation engine
Prediction	None	Not reported	Random Forest next-month prediction
User Profiling	None	Not reported	K-means clustering of emission profiles
AI Guidance	Generic advice	Tips and chatbot	LLM via Groq API: tips, insights, answers (JSON)
Engagement	None	Rewards and community features	Leaderboard and badges
Evaluation Status	Not applicable	Small user study; others not reported	Synthetic benchmark; user study and LLM evaluation planned

The proposed carbon footprint tracker integrates multiple functions into a single platform, as illustrated in Fig. 1. Users can record their daily activities, calculate their estimated carbon emissions, review historical data, generate reports, and explore emission forecasts and user profiles. The platform also provides AI-assisted sustainability recommendations and gamification features to encourage environmentally responsible habits. Table 2 presents a comparison of the proposed system with existing approaches discussed in Section II. This comparison is based on information reported in the reviewed literature and does not represent independent testing of those systems. The synthetic results should be read carefully. Random Forest was best on MAE and MAPE, but its advantage over Ridge regression is small, and the feature importance was dominated by last month's total (0.99) because emissions in the generated data are persistent. The perfect clustering result reflects synthetic separability. The LLM part, which is central to the user experience, still needs evaluation of relevance, factual accuracy, JSON validity and response time with real prompts.

5.3 Future Work

To extend the tracker, several steps are earmarked:

1. **Longer Data Collection** – Collect at least 6–12 months of activity records from a larger user group, with consent.
2. **Forecast Validation** – Train the Random Forest on earlier months and test on later ones. Compare against a last-month baseline and a moving average, and report MAE, RMSE and MAPE only from these real results.
3. **Profiling** – Apply K-means once enough users exist, and check cluster stability.
4. **LLM Evaluation** – Rate the relevance and factual accuracy of AI recommendations, measure JSON validity rate and response time, and compare with rule-based tips.
5. **Pilot User Study** – A four-week study with students and staff measuring usability with the System Usability Scale (SUS), recommendation usefulness, change in monthly emissions and a pre/post attitude questionnaire.
6. **Calibrated Emission Factors** – Checking every factor against official sources (DEFRA, IPCC and the Central Electricity Authority for India) and updating them regularly.

7.Automatic Data Capture and Explainable

Recommendations – Reducing manual entry through trip detection and utility-bill upload, and showing the calculation behind each suggested saving so that users can trust and verify it.

6. Conclusion

This paper presents the design and prototype of an AI-Based Personal Carbon Footprint Tracker built with React, Tailwind CSS, Vite, PHP and MySQL. The implemented core lets users record activities and estimate emissions using Emission = Quantity × Emission Factor, with category-wise breakdowns and LLM-assisted recommendations. The prototype was tested on 10 users over one month, so the evaluation is limited to the implemented features and describes the pilot data only. Random Forest next-month prediction, K-means profiling and monthover-month comparison are proposed extensions. They have not been validated, and no forecasting accuracy is claimed. Future work will collect longer-term data from more users, validate emission factors against official sources, evaluate the AI-generated recommendations, and test forecasting with proper time-ordered validation.

Copyright and Licence

The authors retain copyright and agree to publication under the journal's Creative Commons Attribution 4.0 International (CC BY 4.0) licence, subject to acceptance.

Acknowledgements

The authors would like to thank Dr. Jogi John, Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, Maharashtra, India, for his guidance and support throughout this project.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

The data supporting the findings of this study are not publicly available due to privacy considerations. Relevant summary statistics are provided in the manuscript. Further information may be requested from the corresponding author, subject to applicable privacy and consent restrictions.

Ethics Statement

Activity and consumption data are personal. The system should follow data minimisation: only the data needed for calculation is collected, passwords are stored as hashes, and traffic is sent over HTTPS. Prompts sent to the external LLM service must contain only summarised emission data and no names or e-mail addresses, and users should be told that their data are processed by a third-party service. Users should be able to export and delete their records, in line with data-protection principles such as India's Digital Personal Data Protection (DPDP) Act and the General Data Protection Regulation (GDPR). Results must be presented honestly: footprints are estimates based on average factors and not precise measurements, and AIgenerated advice can be wrong, so the interface should say so. Leaderboards should rank improvement against a user's own baseline and avoid shaming, and recommendations should respect users' circumstances, for example the lack of public transport in some areas.

Use of AI Tools

The manuscript describes an LLM service accessed through the Groq API as part of the proposed application. The authors should disclose any generative-AI tools used to prepare or edit this manuscript, if applicable.

Author Contributions

Author contributions were not specified in the supplied manuscript. Add each author's contribution before submission.

References

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Central Electricity Authority. (n.d.). *CDM–CO_2 baseline database*. Ministry of Power, Government of India. https://cea.nic.in/cdm-co2-baseline-database/?lang=en

Department for Energy Security and Net Zero, & Department for Environment, Food & Rural Affairs. (n.d.). *Greenhouse gas reporting: Conversion factors*. GOV.UK. https://www.gov.uk/government/collections/government-conversion-factors-for-company-reporting

Groq. (n.d.). *OpenAI compatibility*. Groq API documentation. https://console.groq.com/docs/openai

Hoffmann, S., Lasarov, W., & Reimers, H. (2022). Carbon footprint tracking apps: What drives consumers' adoption intention? *Technology in Society, 69*, Article 101956. https://doi.org/10.1016/j.techsoc.2022.101956

International Energy Agency. (2024). *CO_2 emissions in 2023*. https://www.iea.org/reports/co2-emissions-in-2023

Intergovernmental Panel on Climate Change. (2023). *Climate change 2023: Synthesis report*. https://doi.org/10.59327/IPCC/AR6-9789291691647

Jasmy, A. J., Ismail, H., & Aljneibi, N. (2024). A novel approach to sustainable behavior enhancement through AI-driven carbon footprint assessment and real-time analytics. *Discover Sustainability, 5*, Article 476. https://doi.org/10.1007/s43621-024-00762-w

Lin, S.-M. (2016). Reducing students' carbon footprints using personal carbon footprint management system based on environmental behavioural theory and persuasive technology. *Environmental Education Research, 22*(5), 658–682. https://doi.org/10.1080/13504622.2015.1018142

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.

Ouariachi, T., Li, C.-Y., & Elving, W. J. L. (2020). Gamification approaches for education and engagement on pro-environmental behaviors: Searching for best practices. *Sustainability, 12*(11), Article 4565. https://doi.org/10.3390/su12114565

Peter, C., Helming, K., & Nendel, C. (2017). Do greenhouse gas emission calculations from energy crop cultivation reflect actual agricultural management practices? A review of carbon footprint calculators. *Renewable and Sustainable Energy Reviews*. https://doi.org/10.1016/j.rser.2016.09.059

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A. S., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C., Ng, A. Y., Hassabis, D., Platt, J. C., ... Bengio, Y. (2019). *Tackling climate change with machine learning*. arXiv. https://arxiv.org/abs/1906.05433

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Sgaramella, G., & Bilancini, E. (n.d.). *Large language models for the promotion of climate actions: An online experiment* [Project description]. IMT School for Advanced Studies Lucca. https://openlab.imtlucca.it/research/coreprojects/environment-climateand-biodiversity

Wegmeth, L., Vente, T., Said, A., & Beel, J. (2025). *Green recommender systems: Understanding and minimizing the carbon footprint of AI-powered personalization*. arXiv. https://arxiv.org/abs/2509.13001

Yu, H., Yang, Y., Li, B., Liu, B., Guo, Y., Wang, Y., Guo, Z., & Meng, R. (2023). Research on the community electric carbon emission prediction considering the dynamic emission coefficient of power system. *Scientific Reports, 13*, Article 5568. https://doi.org/10.1038/s41598-023-31022-y