Ethical Risks and Improvement Potentials of AI Tools in Carbon Footprint Calculation: The Case of XAI

Din-Yuang Huang
Holistic Education Center, Graduate Institute of Educational Leadership and Development, Fu-Jen Catholic University, 24205 New Taipei, Taiwan
060967@mail.fju.edu.tw

Keywords: AI Ethics, Carbon Footprint Calculation, Explainability, Transparency, Explainable AI (XAI), Sustainable Development.

Abstract: This study aims to explore the ethical risks from AI tools in carbon footprint calculation applications, especially focusing on Explainability and Transparency. Although AI provides convenient computing, the problem of "black box" may lead to a variety of ethical concerns caused by computing. This study first systematically reviews multiple documents and analyses to understand the initiatives for AI ethics and the risks arising from unexplainability and opacity. This study proposes Explainable Artificial Intelligence (XAI) as a solution to mitigate these ethical risks. The paper further explores how XAI can improve the transparency of carbon footprint calculation and its specific implementation by users through simulation of real-world scenarios. The Results show that XAI technology can transform AI's abstract predictions into mathematically rigorous and actionable evidence, enabling users to implement specific carbon reduction behaviours in their lives. Ultimately, this research contributes to the growing discourse on responsible AI by demonstrating how explainable models can foster trust, accountability, and sustainability in data-driven environmental decision-making.

1 INTRODUCTION

When we discuss the SDGs (Sustainable Development Goals) and ESG (Environmental, Social, and Governance), calculating carbon footprints is a crucial topic. For example, the ISO 1404X family of standards establishes a rigorous and comprehensive methodology, providing appropriate methods for different areas and yielding meaningful results. At the same time, many online carbon footprint calculators that provide personal carbon footprint calculations are used as calculation tools, and Generative AI (Hereinafter referred to as AI) can also be used to calculate the carbon footprint of personal life or organizations. AI technology has the potential to develop rapidly and be highly applicable in the field of sustainable development, especially in terms of carbon footprint calculation and optimization. As a convenient auxiliary tool, AI can provide a lot of important information and help users avoid complex calculations to obtain concise results.

While AI's efficiency comes with risks, we accept its output, but are AI's calculations biased? Are the results transparent within the AI's black box? Do the results adhere to the principles of sustainability and fairness? These issues, combined with human biases toward sustainability or green energy, can lead to biased judgments about the output. For example, Holmgren et al. conducted an experiment and found that people tend to incorporate compensatory green beliefs into less environmentally friendly objects, thereby lowering the calculated GHG CO2e values [1]. This reduced footprint may be due to averaging bias. Furthermore, regarding carbon footprint calculators, Murlow et al. noted in their research that these calculators vary greatly in details and boundaries and are often poorly designed to suit users. These AI operations can be subject to human error, leading to erroneous data output [2].

Based on questions above, this study aims to examine the ethical risks inherent in AI tools used in carbon footprint calculations. From an ethical perspective, this study examines how the concept of explainability impacts the reliability, credibility, and auditability of AI carbon footprint models. Furthermore, it addresses how data and algorithmic biases in AI carbon footprint calculations can lead to fairness issues, particularly in the attribution and

allocation of carbon emissions responsibility. Furthermore, based on these questions, this study explores how Explainable AI (XAI) technology can help avoid these ethical risks.

Before research begin, let's first define the boundaries and methods in this study. When discussing AI ethics, concepts like explainability, fairness, transparency, and accountability often come up. Since this study focuses on using AI tools for carbon footprint calculation, we will focus on the concepts of explainability and transparency. These two concepts are related to the calculation process, database, and coefficient selection when using AI to calculate carbon footprint. Since the purpose of this study is to clarify these two concepts and to demonstrate the importance of explainability and transparency through the help of the XAI model, this study does not attempt to develop a new and appropriate personal fitable model and only applies the XAI tools or models in practice. Secondly, to achieve this research result, this study will employ a literature review and synthesis method, as well as a conceptual analysis method, as the foundation for its thinking and research. This study will clarify the importance of these two concepts through a literature review, provide explanations and definitions for each, and then demonstrate their practical application and illustration using the XAI model.

2 LITERATURE REVIEW

The issues of explainability and transparency, along with the challenges posed by algorithmic "black box," have been central concerns since the earliest applications of AI tools. In this section, the literature is reviewed from three perspectives: the positions articulated by the United Nations and UNESCO, insights drawn from major trend reports, and contributions from academic scholarship. We will use the research method of Document Analysis to compare documents horizontally to clarify how different documents interpret explainability and transparency in AI, and whether these documents raise concerns about these concepts. This information will serve as the foundation for our discussion of these two concepts in Section 3.

Readers should pay attention here: the sources considered here are framed primarily within the educational domain. This orientation is deliberate, as carbon footprint calculation constitutes an integral component of Education for Sustainable Development (ESD). If processes of carbon accounting within the ESD context are marked by unexplainability and lack of transparency, the formulation of effective and contextually appropriate decarbonization strategies becomes considerably more difficult. In other words, without reliable interpretability in AI-supported assessments, sustainability education risks producing outcomes that are normatively appealing yet practically inadequate.

2.1 United Nations and UNESCO

The United Nations has, for several years, recognized the risks associated with AI "black box" and has emphasized the need for regulatory oversight of AI tools, particularly in relation to data governance. For instance, the *Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence* (2023) highlights explainability and transparency as essential requirements to ensure accountability. The objective of such regulation is to make explicit the logic and rationale underlying AI systems' outputs and decisions, thereby ensuring that this information is accessible and that users' trust can be strengthened [3].

By 2024, UNESCO extended its concern to the pedagogical context, focusing on potential issues that may arise when teachers and students engage with AI. In its proposed framework for developing teacher competencies, UNESCO underscored that AI outputs often carry elements of randomness. As a result, educators must develop an awareness of the black box nature of AI training processes and critically evaluate the results produced by such systems [4]. Moreover, the framework elaborates that explainability encompasses the languages used in training AI, cultural representativeness, methods of data collection and utilization, as well as the adaptability of AI systems to learners of different ages and abilities [4].

With regard to students, UNESCO recommended that learners be made aware of their right to request information concerning explainability from system designers and providers [5]. Such awareness, it argued, forms a foundational competency of a human-centered mindset [5]. Similarly, in the Governing AI for Humanity: Final Report, the United Nations emphasized that AI technologies are frequently characterized by opacity and deficiencies in interpretability [6]. The report further urged member states to advance research into the transparency of AI, thereby mitigating the risk of outcomes or decisions that may exceed human capacity for responsibility or redress [6].

Most recently, in 2025, UNESCO introduced additional guidance for the academic use of AI. The report *Guidance for Generative AI in Education and Research* explicitly acknowledged the longstanding challenge: "It has long been recognized that artificial neural networks (ANNs) are usually 'black box'; that is, that their inner workings are not open to inspection. As a result, ANNs are not 'transparent' or 'explainable', and it is not possible to ascertain how their outputs were determined" [7]. Consequently, UNESCO recommended that AI providers disclose trustworthy data and computational models, including the provenance of data, the modeling techniques employed, and the types of algorithms used, thereby ensuring both explainability and transparency for end-users [7].

2.2 Recommendations and Cautions in Trend Reports

Beyond the contributions of the United Nations and UNESCO, numerous foundations and research institutions with a focus on education regularly publish reports on AI-related issues. Many of these reports likewise highlight concerns regarding the "black box" nature of AI systems and offer recommendations for mitigating associated risks.

In a report on the ethical risks of AI prepared for SURF in the Netherlands, Sack and Little observed that AI tools pose threats to fairness and justice, one major cause being the underlying algorithms themselves. Such algorithms may perpetuate biases or enable exploitative practices [8]. Similarly, Molina and Medina, in a report commissioned by the World Bank on the use of AI in higher education across South America, argued that the lack of transparency in AI tools obstructs the effective promotion of certain professional practices and constrains the optimal use of AI. They further noted that algorithmic biases can produce inequitable outcomes [9]. They identified seven potential sources of bias, among which "Data Collection and Preparation" and "Problem Definition" are particularly relevant to the domain of carbon emission calculation [9].

Hoernig and colleagues, in a report for the Lisbon Institute of Public Policy, similarly emphasized that one of the primary harms of AI use is the lack of explainability and transparency. From an educational standpoint, they argued, users must first confront the problem of AI black box in order to monitor algorithmic bias, design curricula, or integrate AI responsibly into instruction [10]. The report also noted that the European Union's AI Act already requires disclosure of AI outputs and calls for measures to mitigate risks stemming from inadequate explainability and transparency, including training to identify bias [10].

In the United States, the Department of Education's Office of Educational Technology underscored similar concerns in its 2023 report. Cardona et al. emphasized the importance of enhancing transparency and providing explicit disclosure of relevant information [11]. They recommended the development of integrated AI policies grounded in ethical and just principles, with "Promote Transparency" identified as a core pillar [11]. Moreover, the report advocated for the adoption of inspectable, explainable, and overridable AI in educational contexts. Such systems would enable teachers to understand how AI evaluates student work, why it recommends certain resources, and the rationale behind specific pedagogical suggestions [11].

2.3 Other Categories of Literature

Many scholars have also mentioned the importance of ethical characteristics such as AI explainability and transparency. For example, Pikhart and Al-Obaydi (2025) interviewed teachers and mentioned that the respondents expressed concerns about the validity and reliability of AI data [12]. These concerns include the belief that AI data may contain inaccuracies or erroneous information and the need for further verification [10]. Even those respondents who acknowledged that AI data has some reliability believe that verifying the information is important because AI can produce data that is difficult to assess for accuracy. These respondents generally believe that data generated by AI is generally unreliable. Vasquez III et al. note that AI algorithms can be subject to bias and potentially lead to unfairness, emphasizing that potential biases should be carefully considered when developing AI and its algorithms. Furthermore, attention should be paid to the lack of transparency in AI's decision-making process [13]. Vredenburgh (2024)'s research on the application of AI in public policy emphasizes that AI should be transparent to avoid harm to human rights in politics or justice [14].

To sum up, explainability and transparency of AI are important because these two ethical qualities affect every aspect of our daily lives. From the perspective of carbon emission calculations, these two characteristics are not only related to whether the calculation process can produce reasonable conclusions, but also to whether the results obtained due to this unclear process will cause harm to the user.

The reviewed literature highlights the central importance of AI explainability and transparency within the domains of education and sustainable development. The United Nations and UNESCO emphasize that, in order to uphold ethical use and accountability, the decision-making logic and computational processes of AI systems must remain understandable and accessible. Trend reports further caution that algorithmic bias and opacity may compromise fairness, and that for both educators and students, understanding the "black box" characteristics of AI is essential to safeguarding the quality of teaching and learning. Academic studies corroborate this perspective, noting that lack of transparency in AI data, reliability, and public policy applications can produce potential harms for users and society. From the perspective of carbon footprint calculation, these ethical attributes affect not only the reasonableness of computational processes but also the potential risks of adverse outcomes for users. Overall, the literature underscores that maintaining explainability and transparency in AI-supported carbon footprint assessment is foundational to ensuring ethical rigor, auditability, and fairness.

3 ON THE CONCEPTS OF EXPLAINABILITY AND TRANSPARENCY

Based on the preceding literature analysis, we have established that explainability and transparency are of critical importance for AI. The question that follows is: how should these two concepts themselves be explained? In other words, how can the concepts of explainability and transparency be rendered explainable and transparent?

Vredenburgh (2024), drawing on Kafka's novel The Trial, illustrates the dangers of untransparency and unexplainability: the protagonist is arrested, yet neither he nor anyone else knows the reason for his arrest. Although everyone insists he must be guilty, no one can specify his crime [14]. This allegory reflects the problem of AI "black box." Vredenburgh argues that explainability refers to the capacity of inference models to be understood by users. Using decision trees as an example, he notes that the branching nodes of decisions should be traceable; in this sense, AI outputs can be considered transparent. Vredenburgh thus situates these terms within the domain of political philosophy. Balasubramaniam et al. (2023) approach the concepts from an ethical standpoint, viewing transparency as a non-functional requirement that is nonetheless crucial due to its dependence on trust within system development. Both explainability and transparency are context-dependent and require careful attention to ethical principles in real-world applications [15].

The findings of Vredenburgh and Balasubramaniam align with Palacio et al. (2021), who note that although these concepts are widely regarded as important, there is no universally accepted standard, and their interpretation and application remain diverse [16]. This plurality affords significant flexibility for the design of XAI. Palacio et al. ultimately define explanation as "the process of describing one or more facts, such that it facilitates the understanding of aspects related to said facts (by a human consumer)." Compared to the previous two studies, this definition similarly emphasizes the process of being understood, though the focus differs across contexts. These conceptual insights resonate with Fedytskyi's (2025) discussion of transparency, which highlights its relation to data bias and decision-making processes. Both terms relate to the reference process users use to make decisions [17].

The above description is similar to the OECD's definition of explainability and transparency in Section 1.3 of The OECD Recommendation on AI [18]. In the web version, the OECD provides us with further explanation. Transparency emphasizes the disclosure of information during interactions, allowing users to understand AI's operations and obtain meaningful information to make more informed choices. It also promotes public and stakeholder understanding. Regarding explainability, the OECD states that "explainability means enabling people affected by the outcome of an AI system to understand how it arrived at." [19]

According to this, and considering that this research is related to carbon footprint calculation, whether the user is motivated by sustainable development education or practical needs such as carbon inventory work, when we use the terms explainability and transparency, we will primarily focus on ensuring that the AI coefficient selection and operation process can be understood. The former emphasizes the calculation process, while the latter means the source of the selected coefficients or database: both are important bases in carbon emission calculations. In our research, we will specifically consider transparency because it concerns the fairness of the calculation base and results. The 2023 EDPS (European Data Protection Supervisor) report mentioned, "The concepts of transparency, interpretability and explainability in the context of AI have no formal definition, and are sometimes used

interchangeably." [20] And according to EDPS, "A transparent AI system enables accountability by allowing stakeholders to validate and audit its decision-making processes, detect biases or unfairness, and ensure that the system is operating in alignment with ethical standards and legal requirements." [20] This is because it concerns whether the coefficients and regulations selected by AI when calculating carbon footprints are available for users to review. As for another common word, interpretability, we will not discuss it in this study because it is related to the interpretation of results rather than the actual calculation of AI when calculating carbon footprint

4 ETHICAL RISK ANALYSIS OF AI TOOLS IN CARBON FOOTPRINT CALCULATION

Building on the results of the preceding literature review, we examined how scholars have explained the concepts of explainability and transparency, both of which are primarily tied to the computational processes involved in the use of AI [21]. For example, in the case of Taiwan, the implementation of organizational carbon inventories is typically guided by five foundational steps in training and education: defining boundaries, identifying sources, performing calculations, reporting, and verification [21]. Beyond the act of calculation itself, both boundary setting and source identification are directly related to methodological choices and the selection of coefficients. If AI systems fail to provide adequate reference materials, users are exposed to the risk of being unable to discern the reliability of the data produced.

4.1 Ethical Characteristics of AI

Within ethical frameworks, AI is expected to embody, though not be limited to, the following characteristics: explainability, fairness, transparency, accountability, privacy, and security. UNESCO's thical recommendations highlight that AI-generated outcomes must not cause harm, nor should AI-driven decisions lead to irreversible consequences [3]. Although the report primarily references life-and-death decision-making scenarios, the principle holds equal importance in carbon footprint calculations, as such computations directly inform strategic planning, policy decisions, and financial commitments to emission reduction [3]. When AI ethics link transparency and explainability to accountability and responsibility, they entail whether sufficient explanations are provided to facilitate user understanding, and whether algorithms and internal reasoning processes are disclosed to the public or regulatory authorities as part of essential information transparency.

One of the primary ethical risks in using AI for carbon footprint calculations is the lack of explainability. Carbon footprints can be computed through categorical classification and subsequently presented as numerical data in public documents such as sustainability reports. However, for users lacking specialized training in carbon inventory methodologies, the calculation process and classification methods are often not their primary concern. Instead, they focus on the outcomes and numerical results, which guide their decisions in autonomous emission reduction efforts. In such contexts, transparency and explainability in the computational process demand a greater degree of trust. Therefore, when using AI tools to conduct carbon inventories, in the absence of coefficients, supporting information, and calculation processes, the data obtained poses a certain degree of risk from an ethical perspective.

4.2 The Case of Carbon Footprint Calculators

Now we can take the example of a carbon footprint calculator to illustrate the impact of this lack of interpretability risk. The carbon footprint calculator is a simple tool that helps people understand the carbon emissions generated by their activities. It has appeared on the Internet many years ago. By setting and inputting conditions, users can easily calculate the CO2e GHG they produce within a certain period of time. The carbon footprint calculator is widely accepted by users because it avoids complicated calculation processes and can obtain the corresponding carbon emissions simply by inputting the activity figures.

Although such calculators are user-friendly, reducing human activity into a formula for quantitative computation is not a straightforward task. As early as a decade ago, West et al. (2015) pointed out the inherent complexity of these calculations and their connection to behavioral change [22]. They suggested that future designs should account for changes in product prices and emission factors over time, as such changes may render the underlying methodology outdated. Designers must also navigate the trade-off between precision and usability.

Excessive complexity in data entry, pursued in the name of accuracy, risks deterring user participation.

From a user-centered perspective, Williamson (2022) reported the question which we call the "black box problem" [23]. Based on her experience, the annual household carbon footprint estimated for her family ranged from 8 to 42 tons of CO2e, depending on the level of detail in different calculators. She observed that different tools draw on varying datasets and assumptions, yet these assumptions rarely capture the specific lifestyle of an individual user relative to national averages. This underscores the necessity for calculators to provide transparent and sufficient data, references, and disclosure of computational processes.

DuPuis and Mulvaney (2024) mentioned the black box problem of this carbon footprint calculation. Although the carbon footprint calculator is a simple and convenience tool, it has been criticized on three problems: issues of accuracy and completeness in its design, its effectiveness in achieving the goal of changing individual behavior, and related scientific research issues in climate governance [24]. The first of these is related to the opacity issue mentioned in this article; the inaccuracies of carbon footprint calculators include inconsistent results, a lack of transparency into the data sources used in their algorithms, and errors even in the data used for reference. Taking meat calculations as an example, they pointed out that factors ignored by carbon footprint calculators lead to differences in carbon footprint calculations. Such differences may cause the calculation results to be distorted, and ultimately it is difficult to provide practical help for global emission reduction efforts.

The impact is not just about the difference in the calculation results, but also about applying the results to possible wrong decisions in life. Improper design of carbon reduction methods, or increasing emission reduction efforts in areas where emission reduction results are not significant, will affect the adoption and implementation of emission reduction strategies. If the carbon footprint calculator website used has a payment mechanism where users can donate to help reduce emissions, how should users ensure that their quota is in line with actual results? Therefore, the black box aspects of carbon footprint calculations in AI applications include difficulties in auditing and verification, particularly in understanding why AI models produce specific carbon footprint values. Furthermore, when no reference coefficient for emissions reduction is specified, the AI can be tricky to understand which data sources it uses to generate its results. Due to the difficulty of auditing and verification, it is difficult for users to assess whether they are biased by data, sampling or measurement, thus giving rise to fairness issues.

These ethical challenges are related to opacity and the absence of explainability, which also blur accountability and responsibility. While UNESCO's Ethical Impact Assessment framework provides a template for evaluation [3], its effective application depends on both AI developers and users proactively engaging in self-assessment. Yet when AI systems generate erroneous results or cause ethical harms, the unresolved question remains: who bears responsibility - the data providers, AI developers, model deployers, or the end-users?

5 XAI AS AN ETHICAL SOLUTION TO THE BLACK-BOX PROBLEM IN CARBON FOOTPRINT CALCULATION

As discussed earlier, the issue of the "black box" in AI decision-making poses significant ethical challenges. Because carbon footprint calculation involves the use of reference coefficients and the specification of contextual conditions, employing AI as a computational tool may generate discrepancies in outcomes - particularly when users are unable to clearly define all scopes, categories, and conditions. In such cases, whether through simplified carbon footprint calculators or AI-based computational frameworks, the results may vary substantially. To address these black-box concerns in carbon footprint analysis, the introduction of Explainable AI (XAI) offers a potential pathway for mitigating these ethical dilemmas.

5.1 The Concept of XAI

Taking IBM as an example, the company defines XAI as "a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms." According to this perspective, predictions must be accurate, the underlying techniques should be traceable, and the decision-making processes must adequately meet human needs. Proper use of XAI can enhance users' confidence in the outputs while simultaneously assisting designers in building and planning AI models with a sense of responsibility. In this respect, XAI is regarded as a key enabler of responsible AI, positioning AI models within a framework of trust and transparency [25].

EDPS further highlights that XAI embodies transparency, interpretability, and explainability – features that help users evaluate decision-making processes and understand the principles behind model predictions. EDPS also identifies two modes through which XAI facilitates interpretability in AI tools [20]:

- White Box: Employing simple and understandable models (e.g., decision trees or linear regression) that provide clear explanations of decision logic, thereby allowing decision-makers to grasp the rationale underlying specific outcomes.
- Post-hoc Explanations for Black Box Models: Offering retrospective explanations of decision processes. These include local explanations such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations), as well as global explanations that provide broader interpretive frameworks.

Management Solutions also defines XAI as "a set of processes and methods that enable users to understand and trust the results and products created by machine learning algorithms. This discipline is crucial for an organization to build trust when using AI models, helping to characterize model accuracy, fairness, transparency and understanding of results in AI-based decision making" [26]. Their report further distinguishes between "interpretability" and "explainability." The former emphasizes human-centered understanding, defined as "the ability to explain or present in terms that are understandable to a human being," and refers to the degree to which a person can grasp the cause of a decision. The latter, in contrast, focuses on algorithms and internal logics, defined as "the dimension of a model output that describes how the output of the model was produced," demonstrating the extent to which the internal mechanics of a machine learning system can be expressed in human terms [26].

In sum, XAI may be concisely described as an approach that distinguishes itself from conventional AI by ensuring that the information it generates is interpretable and that its processes of outcome generation and decision-making can be rendered intelligible. XAI thus embodies the explainability and transparency that contemporary discourse expects from ethically responsible AI systems.

5.2 Is Such an Operation Ethically Sound?

If we apply the concept of XAI to carbon footprint calculation, and recall the discussion in Section 4.2 where the carbon calculator was used as an example, the importance of explainability and transparency in carbon accounting becomes evident. XAI, when integrated into carbon footprint calculation, should allow users to trace the sources of coefficient references, or at the very least, help them understand why AI recommends a particular mitigation strategy.

Drawing upon the three procedural steps in carbon accounting – boundary definition, source identification, and calculation – XAI's transparency primarily facilitates bias detection and diagnostic assessment. When XAI can reveal variations in model performance across different data subsets (such as geographical regions, income groups, or industrial sectors), users can scrutinize whether unfair coefficient references exist or whether location-specific sensitivities have been properly accounted for. As highlighted in the United Nations' Governing AI for Humanity: Final Report (2024), AI governance is far from evenly distributed worldwide. For instance, among 54 African countries, 48 were reported in 2024 to lack adequate AI governance mechanisms [6].

The distinctive features of XAI can provide substantial support in this context. When data inputs are adaptively trained and model algorithms optimized, users can access more accurate information and results regarding carbon footprint calculations. Enhanced precision not only improves trust but also increases the acceptability of the outcomes. Particularly significant is the capacity of XAI to furnish a "chain of evidence" in model decision-making. This includes documentation of data sources, coefficient selection, boundary settings, and constraints – key elements in carbon accounting. Such mechanisms enable AI to provide more reliable and consistent explanations, thereby helping to avoid potential greenwashing practices in carbon reporting.

In this way, by enhancing both explainability and transparency, the data produced by AI systems gains qualities of accountability and traceability. These qualities are indispensable for meeting the ethical requirements imposed on the use of AI in carbon footprint calculations.

5.3 Example: Does a Clearer Calculation Process Enhance Explainability?

Building on the concepts previously outlined, if explainability and transparency are integrated into carbon footprint calculation, users will be able to verify the ethical responsibilities tied to

accountability. In this section, two illustrative cases are used to advance the discussion:

1) Does employing the most recent version of ChatGPT-5 for carbon footprint calculations enhance the possibility of generating comprehensible results?
2) Does the use of XAI techniques – such as LIME and SHAP – offer more meaningful outcomes in carbon accounting?

Both cases are based on a carbon reduction project carried out and supervised by the author during an academic conference held in June 2025. One of the author's responsibilities in this conference was to establish a baseline for the event's carbon emissions, so that similar large-scale events in the future could benefit from a standardized reference for carbon footprint assessments. For this purpose, the research team employed Gemini to construct the computational model and adhered to ISO 14064-1 to categorize potential carbon emissions into six distinct categories. The resulting calculations were subsequently cross-verified using ChatGPT-4.

5.3.1 A Clearer Calculation Process

Taking ChatGPT as an example, in August 2025 OpenAI released GPT-5, which provides enhanced computational capacities. On August 10, 2025, the author conducted a comparative inquiry into the two versions, with particular attention given to whether GPT-5 exhibited operational features consistent with XAI principles. The test case involved paper lunchbox recycling, which was one of the subprojects in the aforementioned conference. During the event, the working team collected and recycled a total of 633 paper lunchboxes. The results highlight the differences between GPT-4 and GPT-5 in carbon emission calculations.

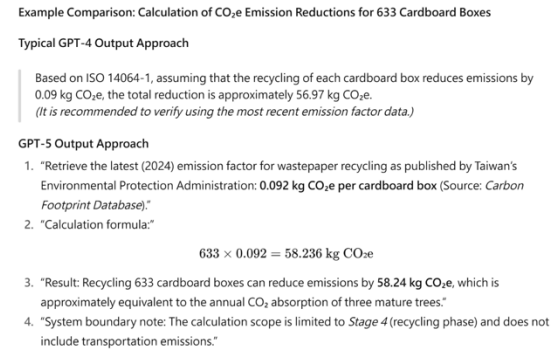

Figure 1: The example of differences between GPT-4 and GPT-5 in carbon emission calculation.

The above chart illustrates the computational differences between GPT-4 and GPT-5. While both algorithms compute "633 recycled cartons," the computational process and content presented differ significantly. Using GPT-4, we can directly obtain the basis and results of the calculation, but without the computational process and its limitations. Whereas GPT-5 went further by offering not only the basis and results but also the explicit calculation formulas and system boundary settings. Particular attention should be directed toward the notion of system boundary. This is a critical step in carbon footprint assessment, corresponding to the "boundary" concept discussed earlier. As shown in GPT-5's output (point 4 in the figure), the results are constrained within Stage 4, thereby offering substantial support and evidence regarding whether transportation should be incorporated into the calculations. This feature enables users to more clearly distinguish the categories of emissions, thereby enhancing both the clarity and accountability of the carbon accounting process.

5.3.2 Application of LIME and SHAP Models

The calculation process in Figure 1 above provides an example of transparency. Introducing LIME and SHAP can also enhance our understanding of data. This relates to the concept of interpretability, that is, whether data can be understood in a more reasonable way. This forms the second case study presented in this research.

This paper does not delve into the principles of LIME and SHAP. Related explanations can be found in the earlier discussion of XAI [26]. We can also build upon the statements made by Dwivedi and his team [27]: both methods are characterized by their ability to provide interpretations that measure the impact of having a certain value for a given feature in comparison to the prediction. Thus, when data is produced, it is no longer just numbers or results, but rather a set of symbols that can be understood and given meaning.

In this section, regarding the calculation of conference-related carbon emissions, we use Stage 3: Carbon emissions from commuting as an example. Based on Gemini's calculation results, the researcher can compare the reasonably expected amounts of carbon emissions. For instance, a scenario may be set up comparing driving a 2000cc gasoline SUV from Taichung, Taiwan, with traveling by high-speed rail. Traditional data would only provide raw figures, such as the estimated CO2 emissions produced by driving

the vehicle. However, by incorporating LIME and SHAP, the analysis can be extended to include driving habits, thus providing more meaningful explanatory insights.

We can compare scenarios constructed by SHAP: for example, comparing aggressive driving with taking the high-speed rail. If we assume a driver tends to drive aggressively, coupled with a vehicle of higher engine displacement, the total carbon emissions for a single trip could reach 28 kg CO2e. In contrast, traveling by high-speed rail would result in only 3 kg CO2e. The choice of "mode of transportation" offsets most of the emissions. When these two approaches are compared, it becomes possible to quantify and evaluate the environmental benefits of choosing high-speed rail, providing strong persuasive power.

5.4 Why Is XAI Important for Carbon Footprint Calculation? Ethical Reflections

In the two examples discussed in Sections 5.3.1 and 5.3.2, we observed that XAI provides us with clearer and more explicit explanations, making the interpretation of data more meaningful. For instance, in the case of carbon emissions from driving, if the data is merely explained at a numerical level, it does not inherently carry accountability. However, when the SHAP model is applied, the data becomes directly linked to vehicle engine displacement and driving behavior. This kind of modeling not only resonates with users but also prepares the groundwork for future carbon-related transactions by ensuring that the outcomes are understandable and fair for trading.

An article published on the Sustainability Director website reminds us that XAI has significant implications for carbon pricing. The transparency of carbon pricing implies data disclosure, and under such circumstances, it builds our understanding of and trust in mechanisms for sustainable economic transitions. XAI can provide the transparency required for carbon pricing: through its application, it can certify the transparency of the origin of carbon credits. Since XAI can offer the necessary tools and techniques to address transparency challenges in the field of carbon pricing – through evaluating carbon sequestration projects, collecting and processing complex datasets, and ensuring accountability in carbon price setting – it establishes a rigorous methodological foundation for high-quality, reliable data. This introduction of XAI provides quality-assured data through rigorous methodologies and user-centric design, ensuring that different users receive the most appropriate data. This article also addresses the issue of transparency atrophy, meaning that even with XAI, black box migrations can still occur [28].

In this sense, the use of XAI in carbon footprint calculation, based on its properties of explainability and transparency, creates the possibility of accountability and fairness in the outcomes. Thus, the introduction of XAI into carbon footprint assessment can meet ethical requirements in its application.

6 CONCLUSIONS

Since carbon footprint calculation is an essential component of sustainable development, obtaining reliable data becomes critically important. Reliable data requires clarity regarding database sources, the use of accurate and verifiable coefficients, and a transparent and reviewable calculation process. We cannot demand that every user undergo professional training or obtain certification before engaging in carbon reduction efforts, because such practices are a vital part of sustainability education and should be accessible as actions that anyone can take in daily life. Therefore, when using AI as an assistive tool, AI with explainability and transparency becomes particularly important: users need to understand how their data is generated, so that they can in turn understand how to take concrete actions to reduce emissions.

In discussing the introduction of XAI as a support for carbon footprint calculation, this study argues that explainability and transparency are crucial when using AI tools, and that XAI has the potential to meet these requirements (though its effectiveness still depends on users' professional competence and the depth of the instructions provided). While this paper used LIME and SHAP as examples of XAI applied to carbon footprint calculation, these cases are only preliminary illustrations. Future research could explore how to deepen the application of XAI or even develop new XAI methods better suited for environmental data and carbon footprint models. Furthermore, due to the limitations of this study, the concept of interpretability was not addressed in detail – only briefly referenced in examples. Since this concept relates to fairness, justice, and accountability, it is recommended for future dedicated research.

The integration of XAI technologies in the future will not only provide data with explainability and transparency in carbon footprint calculations, but also – through comparisons of different models – help users understand how they can implement emission reduction strategies in real life and even enable them to clearly measure the actual numerical outcomes of

their carbon reduction efforts. For this study was based on document analysis, while it provided some conceptual clarity, the limited AI experimentation left some room for improvement. Future research is recommended to conduct qualitative interviews to understand user experiences and design more comprehensive experiments. Through actual AI implementation and analysis of the results, we can understand how the introduction of XAI improves the concepts of explainability and transparency.

REFERENCES

[1] M. Holmgren, H. Andersson, and P. Sörqvist, "Averaging bias in environmental impact estimates: Evidence from the negative footprint illusion," Journal of Environmental Psychology, vol. 55, pp. 48–52, 2018.

[2] J. Mulrow, K. Machaj, J. Deanes, and S. Derrible, "The state of carbon footprint calculators: An evaluation of calculator design and user interaction features," Sustainable Production and Consumption, vol. 18, pp. 33–40, 2019.

[3] UNESCO, "Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence," UNESCO, 2023. [Online]. Available: https://www.unesco.org/en/articles/ethical-impact-assessment-tool-recommendation-ethics-artificial-intelligence.

[4] F. C. Miao and C. Mutlu, AI Competency Framework for Teachers. Paris, France: UNESCO, 2024. [Online]. Available: https://doi.org/10.54675/ZJTE2084.

[5] F. C. Miao, S. Kelly, and L. Natalie, AI Competency Framework for Students. Paris, France: UNESCO, 2024. [Online]. Available: https://doi.org/10.54675/JKJB9835.

[6] United Nations, Governing AI for Humanity: Final Report. New York, NY, USA: United Nations, 2024. [Online]. Available: https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf.

[7] F. C. Miao and H. Wayne, Guidance for Generative AI in Education and Research. Paris, France: UNESCO, 2023. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000386693

[8] T. Sack and B. Little, The Risks of Personalising Higher Education with Artificial Intelligence: Ethical Risk Report. Utrecht, Netherlands: SURF, 2025. [Online]. Available: https://www.surf.nl/files/2025-04/the-risks-of-personalising-higher-education-with-artificial-intelligence-nati-sack-ben-little.pdf.

[9] E. Molina and E. Medina, AI Revolution in Higher Education: What You Need to Know. Washington, DC, USA: The World Bank, 2025. [Online]. Available: https://documents1.worldbank.org/curated/en/099757104152527995/pdf/IDU-b1e5ef00-75ff-4ba4-a4b6-84899c3ea968.pdf.

[10] S. Hoernig, A. Ilharco, T. Pereira, and R. Pereira, Generative AI and Higher Education: Challenges and Opportunities. Lisbon, Portugal: Institute of Public Policy, 2024. [Online]. Available: https://www.ipp-jcs.org/en/2024/09/24/report-11-generative-ai-and-higher-education-challenges-and-opportunitiesreport-11.

[11] M. Cardona, R. J. Rodríguez, and K. Ishmael, Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations. Washington, DC, USA: U.S. Department of Education, Office of Educational Technology, 2023. [Online]. Available: https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf.

[12] M. Pikhart and L. H. Al-Obaydi, "Reporting the potential risk of using AI in higher education: Subjective perspectives of educators," Computers in Human Behavior Reports, vol. 18, Art. 100693, 2025, doi: 10.1016/j.chbr.2025.100693.

[13] E. Vasquez III, J. D. Basham, B. Jimenez, and M. T. Marino, Artificial Intelligence: The Impact of AI on Education for All Learners. Center for Innovation, Design, and Digital Learning, 2024. [Online]. Available: https://ciddl.org/wp-content/uploads/2025/03/Artificial-Intelligence-The-Impact-of-AI-on-Education-for-All-Learners.pdf.

[14] K. Vredenburgh, "Transparency and explainability for public policy," LSE Public Policy Review, vol. 3, no. 3, p. 4, 2024, doi: 10.31389/lseppr.111.

[15] N. Balasubramaniam, M. Kauppinen, A. Rannisto, K. Hiekkanen, and S. Kujala, "Transparency and explainability of AI systems: From ethical guidelines to requirements," Information and Software Technology, vol. 159, Art. 107797, 2023.

[16] S. Palacio, A. Lucieri, M. Munir, S. Ahmed, J. Hees, and A. Dengel, "XAI Handbook: Towards a unified framework for explainable AI," in Proc. 2021 IEEE/CVF Int. Conf. Computer Vision Workshops (ICCVW), Montreal, QC, Canada, 2021, pp. 3759–3768, doi: 10.1109/ICCVW54120.2021.00420.

[17] R. Fedytskyi, "Explainable AI (XAI): Transparency in AI decisions," Medium, 2025. [Online]. Available: https://medium.com/@roman_fedyskyi/explainable-ai-xai-transparency-in-ai-decisions-dccfb0a06592.

[18] OECD, "Transparency and explainability (Principle 1.3)," in Policies, Data and Analysis for Trustworthy Artificial Intelligence, OECD, 2025. [Online]. Available: https://oecd.ai/en/dashboards/ai-principles/P7.

[19] OECD, The OECD Recommendation on AI. Paris, France: OECD, 2019. [Online]. Available: https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.

[20] European Data Protection Supervisor, "TechDispatch on Explainable Artificial Intelligence," 2023. [Online]. Available: https://www.edps.europa.eu/data-protection/our-work/publications/techdispatch/2023-11-16-techdispatch-22023-explainable-artificial-intelligence_en.

[21] Industrial Development Administration, Ministry of Economic Affairs, Greenhouse Gas Inventory Methods and Analysis. Taipei, Taiwan: MOEA, 2025. [Online]. Available: https://ghg.tgpf.org.tw/CVData/CVData_more?id=a69e3f7686054dfabffd8e7c3b8c95fd.

[22] S. E. West, A. Owen, K. Axelsson, and C. D. West, "Evaluating the use of a carbon footprint calculator: Communicating impacts of consumption at household level and exploring mitigation options," Journal of Industrial Ecology, vol. 20, no. 3, pp. 396–409, 2015, doi: 10.1111/jiec.12372.

[23] R. Williamson, "What goes into the carbon calculator black box?" COSMOS, 2022. [Online]. Available: https://cosmosmagazine.com/australia/carbon-footprint-calculators.

[24] E. M. DuPuis and D. Mulvaney, "Opening the black box: Carbon-footprint calculators, meat consumption, and the 'wicked problem' of metric governance," Sustainability: Science, Practice and Policy, vol. 20, no. 1, 2024, doi: 10.1080/15487733.2024.2390232.

[25] IBM, "What is explainable AI?" IBM, 2023. [Online]. Available: https://www.ibm.com/think/topics/explainable-ai.

[26] Management Solutions, Explainable Artificial Intelligence (XAI): Challenges of Model Interpretability. Madrid, Spain: Management Solutions, 2023. [Online]. Available: https://www.managementsolutions.com/sites/default/files/minisite/static/22959b0f-b3da-47c8-9d5c-80ec3216552b/iax/pdf/explainable-artificial-intelligence-en.pdf.

[27] R. Dwivedi, D. Dave, H. Naik, S. Singhal, O. Rana, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, and R. Ranjan, "Explainable AI (XAI): Core ideas, techniques and solutions," ACM Computing Surveys, vol. 55, no. 9, pp. 1–33, 2023.

[28] Sustainability Director, "Explainable AI for carbon price transparency," Sustainability Directory, 2025. [Online]. Available: https://prism.sustainability-directory.com/scenario/explainable-ai-for-carbon-price-transparency.