An Awareness Framework for Sustainable Selection of LLMs in Business

Natascha Brughitta Anchia and Andreas Martin [0000-0002-7909-7663]

FHNW University of Applied Sciences and Arts Northwestern Switzerland, School of Business, Riggenbachstrasse 16, 4600, Olten, Switzerland
natascha.brughittaanchia@outlook.com, andreas.martin@fhnw.ch

Abstract. The environmental impacts of large language models (LLMs) often remain invisible in business adoption. This paper presents an awareness framework to support the sustainable selection of LLMs, developed using a design science research approach within the marketing department of a major European engineering and technology company. Addressing the lack of transparency and emissions data from LLM providers, the artefact calculates electricity use, carbon emissions, and material impacts of inference tasks and visualises them in an interactive dashboard. Evaluation workshops with stakeholders from marketing, sustainability, and AI strategy confirmed the framework's potential to foster awareness, support sustainable decision-making, and align AI use with corporate environmental goals and the UN Sustainable Development Goals (SDG). The framework is transferable to other business contexts.

Keywords: Sustainable Artificial Intelligence · Large Language Models (LLM) · Business AI Adoption · Design Science Research

1 Introduction

The rapid adoption of large language models (LLMs) in business offers efficiency gains but raises growing concerns about environmental sustainability [23,8,29]. While organisations increasingly align with climate goals and frameworks such as the UN Sustainable Development Goals (SDGs), the environmental footprint of LLMs remains largely invisible in decision-making processes [33,7,10,21,12]. High energy use, carbon emissions, and material consumption during inference are often overlooked, due to limited transparency from model providers and a lack of accessible sustainability tools [27].

Although sustainability reporting is gaining traction, businesses still lack mechanisms to evaluate the environmental and financial impacts of LLMs in real-world functions [27,33]. Decisions are often based on performance or novelty, rather than sustainability. The absence of standardised methods to compare general-purpose and task-specific models further hinders responsible selection [5]. This gap is particularly evident in marketing departments, where LLM use is rising but sustainability remains underexplored [8,13,19,34].

This paper addresses the following research question: *How can awareness for the sustainable selection of large language models be raised among business users and decision-makers?*

To answer this, we propose an awareness framework developed using a Design Science Research (DSR) approach. It was applied within the marketing department of a major European engineering and technology company. Drawing on interviews, documentation, and literature, relevant sustainability and cost dimensions were identified. Then these dimensions were used to develop an artefact: a task-based evaluation model with dashboard visualisation.

This artefact enables stakeholders to assess and compare the sustainability impacts of different LLMs for standardised marketing tasks. By visualising energy use, carbon emissions, material impacts, and costs, it supports sustainability-oriented decision-making. Evaluation workshops with marketing, sustainability, and AI stakeholders confirmed its potential to raise awareness and align AI use with environmental objectives.

The remainder of this paper is structured as follows: Section 2 reviews related work. Section 3 explains the applied research method. Section 4 presents the organisational context and problem awareness phase. Section 5 introduces the framework and artefact. Section 6 discusses the evaluation. Section 7 outlines implications, and Section 8 concludes with future directions.

2 Related Work

Research at the intersection of artificial intelligence (AI), sustainability, and business adoption is gaining momentum. While early enthusiasm around LLMs focused on productivity and innovation, recent studies have highlighted their significant environmental footprint. This section reviews literature on the environmental impacts of AI and LLMs, key sustainability metrics for impact assessment, and the distinction between general-purpose and task-specific models—relevant for informed business model selection.

2.1 Environmental Impact of AI and LLMs

Although AI's innovation potential is widely discussed, its environmental costs remain underrepresented in business literature [23]. Recent studies quantify emissions, electricity use, and resource depletion across model lifecycles [4,20,22]. For instance, training BLOOM (176B parameters) resulted in 24.7 tons of CO_2 equivalents (CO_2eq) from usage alone, increasing to 50.5 tons including hardware and operational computation [24]. Other estimates reach up to 280 tons CO_2eq, equivalent to the lifetime emissions of five cars, depending on model and setup [30,35].

Inference—often overlooked—can exceed training emissions, especially in business contexts where general-purpose models are continuously embedded in applications [27]. Generative and image-based tasks are also more energy-intensive than simpler classification tasks, underlining the need for transparent energy-use insights in AI deployment.

2.2 Sustainability Metrics and Life Cycle Assessment

While no universal methodology exists for measuring LLM environmental impacts, tools and metrics for training and inference phases have emerged [4,6]. Libraries and tools such as CodeCarbon or eco2AI estimate emissions based on server location, hardware, electricity use, and Power Usage Effectiveness (PUE) values [20].

Life Cycle Assessment (LCA), standardised in ISO 14044 [18], offers a framework for measuring environmental impacts from raw material extraction to disposal. Applied to AI, it captures emissions from infrastructure, manufacturing, transport, and usage [4,23]. While full "cradle-to-grave" LCAs remain complex, focusing on inference and deployment yields valuable insights [24].

2.3 Experimental Approach

Experimental approaches measure energy usage of AI models directly. CPU/GPU energy consumption is tracked using tools like *nvidia-smi*, DCGM, or Scaphandre, often combined with libraries like CodeCarbon or Carbontracker [6,16,2]. These methods require executable models and, for inference, API deployment. Measurements rely on public datasets and are converted into emissions using carbon intensity coefficients and PUEs [31,25].

2.4 Calculation Approach

Calculation-based approaches use data-driven estimations and assumptions, allowing for standardisation and reproducibility. These often build on LCA frameworks and decompose model impacts into compute, storage, and network components [4,14]. The scope varies but usually excludes end-user devices. The methodology used in this research is based on [14], which provides traceable formulas to assess inference emissions and material usage for model comparisons. Table 1 summarises common sustainability metrics in the literature. Despite increasing awareness, few tools currently enable real-time comparison of AI models using these metrics in business settings.

Table 1. Overview of sustainability metrics used in AI impact assessment

Metric	**Description**	**Authors**
Abiotic Depletion Potential (ADP) in kg Sb-eq	Depletion of non-renewable abiotic resources such as minerals and metals.	[4,11]
Global Warming Potential (GWP) in g CO_2-eq	Potential contribution to climate change from greenhouse gas emissions.	[14,23]
Primary Energy (PE) in Wh	Total primary energy demand, including direct and indirect usage.	[4,11]

2.5 General-purpose vs. Task-specific Models

Understanding the trade-offs between general-purpose and task-specific models is essential for sustainable model selection. General-purpose models, also referred to as generative or multi-purpose, can perform a variety of tasks without retraining and are highly adaptable [32]. Based on transformer architectures, they offer cross-domain flexibility but come with high energy and data demands [27].

In contrast, task-specific models are built for dedicated use cases (e.g., classification, translation), using labelled datasets to optimise performance and efficiency for a single task [17]. Their narrower scope improves explainability, energy efficiency, and performance in well-defined contexts [3]. They often require less compute and training data, lowering environmental impacts.

Nonetheless, current practice often favours general-purpose models even where task-specific ones would suffice [27], highlighting the need for tools that support nuanced, sustainability-informed decisions.

2.6 Conclusion

This section reviewed literature on the environmental impact of AI, sustainability metrics, and model classification. A clear gap exists in providing business users with transparent tools to compare models from a sustainability perspective. Existing studies are often generic and lack operational relevance for departments like marketing. This paper addresses that gap by focusing on real-world business tasks and proposing a practical artefact for informed and sustainable AI adoption. The next section outlines the research methodology.

3 Research Method

This study adopts the Design Science Research (DSR) methodology, widely used in information systems to solve practical problems by developing and evaluating artefacts [15]. The process follows the three-cycle view by Hevner, comprising relevance, rigor, and design cycles.

The relevance cycle connects the research with its practical context. In this study, the key challenge was a lack of awareness and consideration of AI models' environmental impacts in deployment decisions. Interviews and internal documents from a leading European engineering and technology firm revealed the need for a tool that visualises the environmental and cost implications of LLM use, particularly for marketing tasks.

The rigor cycle anchors the artefact in the scientific knowledge base. A literature review on AI sustainability, environmental impact metrics, and socio-technical decision-making informed the selection of evaluation dimensions. Based on these insights and organisational needs, the artefact's objectives were defined to include electricity consumption, cost, and environmental indicators such as greenhouse gas emissions and abiotic resource use.

In the design cycle, the artefact was developed and iteratively refined. A formula-based methodology for calculating LLM inference impacts was implemented in a Microsoft Excel model, and results were visualised in a Power BI dashboard. Key design requirements included accessibility for non-technical stakeholders, practical relevance, and the ability to simulate various model-task scenarios.

The study followed five phases: problem awareness, suggestion, development, evaluation, and conclusion. The resulting artefact, a task-based spreadsheet and dashboard, supports awareness and informed decision-making on LLM usage in business contexts.

4 Problem Awareness and Context Analysis

This phase aimed to identify the key challenges organisations face when evaluating the environmental sustainability and cost-effectiveness of LLM usage in business. The study focused on the marketing department of a major European engineering and technology company and addressed the sub-research question: *"What challenges exist in assessing the environmental sustainability and cost-effectiveness of AI models for tasks in marketing departments?"*

4.1 Application Scenario

The company is a major European multinational headquartered in Solothurn, Switzerland, and part of a larger corporate group with over 400,000 employees in more than 60 countries, generating €90.3 billion in 2024. It operates across four business sectors: mobility, industrial technology, consumer goods, and energy/building technology.

This study focuses on one business unit in the consumer goods division, specifically its marketing department, which includes over 100 employees across central, regional, and national levels. It is divided into six teams: Strategic Marketing, Online Touchpoints, Offline Touchpoints, Central Marketing Enabling, Business Unit Marketing, and User Experience.

The Central Marketing Enabling team, responsible for multilingual marketing content across all channels, served as the primary context for this study.

Following the release of ChatGPT-3, the company launched an enterprise-wide generative AI initiative with executive backing. This included secure, policy-compliant internal tools: a general-purpose assistant and a task-specific content tool for marketing, reflecting a focus on productivity and responsible AI use.

To explore challenges in evaluating LLM sustainability, five semi-structured interviews were conducted with stakeholders from digital transformation, sustainability, in-house AI consulting, AI management, and marketing. Interviews were transcribed and thematically analysed. Internal documents were also reviewed to assess whether sustainability is defined as a tool selection criterion.

Table 2 summarises the interview participants and their contributions.

Table 2. Overview of interview participants and their contributions

Role	Expertise	Contribution to findings
Head of Marketing	Responsibility for multimedia marketing content creation and strategy. Currently, ambassador for AI activities in marketing.	Involved in testing AI tools for tasks such as market analysis, optimising retail displays, translating videos, and other marketing processes. Acknowledges a "zero awareness" level regarding the energy consumption of AI within department and company.
Sustainability Department Representative	Expertise in corporate sustainability topics, with a particular focus on scope three emissions, circular economy, circular business models, material usage, recycled content, and supply chain.	Aware of the potential negative environmental effects of AI and notes that standards like the GHG Protocol are still being defined in this area for such new topics.
GenAI Portfolio Manager	Portfolio Manager for the GenAI initiative, involved in selecting GenAI tools across the company.	Sustainability is not explicitly integrated into the current evaluation criteria for the GenAI portfolio. The focus for evaluating and prioritising tools and use cases is primarily on business impact and technical feasibility.
Head of Digital Transformation	Leadership role with expertise in driving digital transformation within the company.	Notices lack of awareness regarding the energy consumption of AI among users and within their department. Interested in integrating sustainability into their digital strategy, but currently not a primary selection criterion, as the focus is on short-term financial benefits and cost reduction.
GenAI Consulting / Inhouse Consulting	Consulting and supporting the adoption and implementation of generative AI solutions across different divisions at the organization.	Aware of environmental impacts of AI and acknowledges that sustainability is currently not a primary evaluation criterion in their consulting work. Focus is on quickly helping departments leverage AI tools for efficiency gains and coping with new technologies.

4.2 Findings

The findings revealed a lack of awareness around the environmental impacts of LLM inference. Four key challenges were identified:

Lack of Awareness of AI's Environmental Impacts Most employees, including end users and decision-makers, are unaware of the environmental footprint of LLMs. While sustainability or AI experts showed some awareness of issues like energy and water use, the topic is largely absent from organisational discourse. One participant noted, "there is zero awareness. Nobody asks how much energy AI consumes."

This invisibility is due to the abstract nature of digital energy use. Furthermore, the company's current AI narrative prioritises speed and efficiency over sustainability. As one interviewee put it, "everyone sees the benefit in GenAI, not the cost side."

Absence of Standardised Measurement and Accounting Even where awareness exists, no standardised methodology is used to quantify or report environmental impacts. The GHG Protocol lacks clear guidance on digital AI services, especially third-party or cloud-based ones. Emissions from AI use are assumed to fall under Scope 3, which is already difficult to define and track. While physical products are assessed using LCA methods, no equivalent exists for AI services.

Short-Term Focus and Financial Constraints AI-related investment decisions are primarily cost-driven. Sustainability is rarely part of formal evaluation processes. The focus is on short-term gains, such as cost savings or productivity increases. Unless environmental benefits align with financial returns, sustainability is unlikely to become a decisive factor. As one stakeholder stated, "being more sustainable doesn't always translate into short-term business benefits."

Fragmented Decision-Making and Tool-Driven Adoption AI tools are adopted across departments with limited central coordination. There is no governance framework that systematically includes sustainability experts in AI evaluations. Tool selection is often opportunistic, driven by availability or vendor offerings. Long approval cycles and varying digital literacy levels further hinder coherent, strategic adoption. Some departments implement tools more quickly through external partners than internal IT. This fragmented landscape increases the risk of unsustainable AI deployment.

Summary In short, the company lacks awareness, metrics, and accountability frameworks for the environmental impacts of LLMs. Without clear indicators and alignment with business objectives, sustainability will remain a low priority in AI strategy and deployment.

5 Framework Design and Artefact Development

This section presents the design and instantiation of the awareness artefact, which supports decision-makers and business users in assessing the environmental and financial impacts of LLM usage. The artefact consists of a sustainability evaluation model and an interactive Power BI dashboard, developed using the DSR approach. Its goal is to enable informed, sustainability-aligned model selection.

5.1 Design Principles and Methodology

The framework follows three key principles: *transparency*, *business relevance*, and *comparability*. Transparency ensures documented formulas and assumptions. Business relevance grounds the artefact in real-world use cases. Comparability allows benchmarking across sustainability indicators like CO_2 emissions, electricity use, and cost.

Due to limited measurement infrastructure at the company site, a calculation-based approach was adopted. The method by EcoLogits [14], designed specifically for LLM inference, was applied. It builds on a simplified but robust LCA covering manufacturing, raw materials, transport, and usage, excluding end-of-life impacts due to data scarcity. Its transparency, reproducibility, and inclusion in tools like the AI Energy Score [26] made it suitable for text-based marketing tasks. The scope focuses exclusively on inference, omitting training, networks, and user devices.

5.2 Definition and Standardisation of Tasks

To ensure practical relevance, the artefact is based on real-world LLM use cases within the Central Marketing Enabling team. This team produces content such as product descriptions, translations, and digital marketing text. While AI could support a broad range of tasks, current usage is concentrated on text generation.

Four representative text-based tasks were identified and standardised: Amazon product descriptions, bullet points, website product descriptions, and social media posts. To define a typical inference, internal examples were analysed and tokenised using Hugging Face's tokenizer tool [36]. The resulting average output lengths form the basis for impact calculations. An overview of these tasks and their token counts is provided in Table 3.

Table 3. Standardised marketing tasks and output lengths

Task	**Average Output Tokens**
Amazon Product Description	210
Amazon Bullet Points	250
Website Product Description	215
Social Media Post	95

5.3 Model Selection and Classification

The artefact integrates general-purpose and task-specific models, based on data from [26], [9], and [28]. Classification follows the AI Energy Score framework [26], which groups models into small ($\leq$ 20B), medium (21–66B), and large ($>$ 66B) parameter categories.

5.4 Calculation Methodology

The artefact quantifies electricity use, CO_2 emissions, cost, and material impacts, applying the *GenAI Impact* methodology [14], which separates *usage impacts* from *embodied impacts*.

GPU energy usage during inference is based on the number of output tokens $\#T_{\text{out}}$ and active parameters P_{active}:

$$E_{\text{GPU}} = \#T_{\text{out}} \times (\alpha \times P_{\text{active}} + \beta \pm 1.96\sigma)$$

where $\alpha = 8.91 \times 10^{-5}$, $\beta = 1.43 \times 10^{-3}$, and $\sigma = 5.19 \times 10^{-4}$.

The number of GPUs required depends on the model's memory demand:

$$\text{GPU}(P_{\text{total}}, Q, M_{\text{GPU}}) = \frac{1.2 \times (P_{\text{total}} \times Q/\#\text{GPU}_{\text{installed}})}{M_{\text{GPU}}}$$

Server energy (excluding GPUs) is calculated via inference duration ΔT and a fixed 1 kW power draw:

$$W_{\text{server/GPU}} = 1\,\text{kW}$$

$$E_{\text{server/GPU}}(\Delta T) = \Delta T \times W_{\text{server/GPU}} \times \frac{\text{GPU}}{\#\text{GPU}_{\text{installed}}}$$

Total server energy combines both components:

$$E_{\text{server}} = E_{\text{server}\setminus\text{GPU}} + \text{GPU} \times E_{\text{GPU}}$$

To account for cooling and networking overhead, a PUE factor is applied:

$$E_{\text{request}} = \text{PUE} \times E_{\text{server}}, \quad \text{PUE} = 1.2$$

Energy use is converted to CO_2 equivalents using an emission factor F_{em}:

$$I_{\text{request},u} = E_{\text{request}} \times F_{\text{em}}$$

Embodied impacts are allocated proportionally to the inference duration over a five-year hardware lifecycle ΔL:

$$I_{\text{request},e} = \frac{\Delta T}{\Delta L} \times I_{\text{server},e}$$

$$I_{\text{server},e} = \frac{\text{GPU}}{\#\text{GPU}_{\text{installed}}} \times I_{\text{server}\setminus\text{GPU},e} + \text{GPU} \times I_{\text{GPU},e}$$

with $I_{\text{GPU},e} = 0.00509$ and $I_{\text{server}\backslash\text{GPU},e} = 0.25$.

The total environmental impact per request is:

$$I_{\text{request}} = I_{\text{request},u} + I_{\text{request},e}$$

Assumptions include NVIDIA A100 80GB GPUs, 4-bit quantization, constant PUE, global electricity mix, and fixed hardware lifetimes.

An overview of all applied formulas, including two user-friendly equivalence representations, is provided in Table 4.

Table 4. Overview of Formulas with Description

Formula	**Description**
(1) $E_{\text{GPU}} = \#T_{\text{out}} \times (\alpha P_{\text{active}} + \beta \pm 1.96\sigma)$ *with* $\alpha = 8.91 \times 10^{-5},\ \beta = 1.43 \times 10^{-3},\ \sigma = 5.19 \times 10^{-4}$	GPU energy consumption per output token.
(2) $\text{GPU}(P_{\text{total}}, Q, M_{\text{GPU}}) = \frac{M_{\text{model}}(P_{\text{total}}, Q)}{M_{\text{GPU}}}$, *with* $\text{M}_{\text{model}}(P_{\text{total}}, Q) = 1.2 \times \frac{P_{\text{total}} \times Q}{\#\text{GPU}_{\text{installed}}}$	Number of GPUs needed to load the model.
(3) $E_{\text{server}\backslash\text{GPU}}(\Delta T) = \Delta T \times W_{\text{server}\backslash\text{GPU}} \times \frac{\text{GPU}}{\#\text{GPU}_{\text{installed}}}$ *with* $\Delta T = \#T_{\text{out}} \times (\alpha P_{\text{active}} + \beta \pm 1.96\sigma)$	Server-shell energy consumption without GPUs.
(4) $E_{\text{server}} = E_{\text{server}\backslash\text{GPU}} + \text{GPU} \times E_{\text{GPU}}$	Energy consumed inside the rack during inference, combining (1)–(3).
(5) $E_{\text{request}} = \text{PUE} \times E_{\text{server}}$	Total electricity drawn to satisfy the prompt (Wh).
(6) $I^u_{\text{request}} = E_{\text{request}} \times F_{\text{em}}$	Indicator for effect on global warming (gCO_2eq).
(7) $I^e_{\text{server}} = \frac{\text{GPU}}{\#\text{GPU}_{\text{installed}}} \times I^e_{\text{server}\backslash\text{GPU}} + \text{GPU} \times I^e_{\text{GPU}}$, *with* $I^e_{\text{GPU}} = 0.00509,\ I^e_{\text{server}\backslash\text{GPU}} = 0.25$	Embodied impacts of the server (gCO_2eq).
(8) $I^e_{\text{request}} = \frac{\Delta T}{\Delta L} \times I^e_{\text{server}}$	Use of metals and minerals (kg Sb eq).
(9) $I_{\text{request}} = I^u_{\text{request}} + I^e_{\text{request}}$	Total environmental impacts of the request (gCO_2eq), usage + embodied.
(10) Equivalent smartphones charged = $\frac{E_{\text{request}}}{28{,}466\,\text{Wh} - (22\,\text{h} \times 0.411\,\text{W})}$	Approximate number of full smartphone charges using the same energy.
(11) Equivalent smartphone charging minutes = $\frac{E_{\text{request}} \times 60}{9.5}$	Approximate number of minutes a smartphone could be charged using the same energy.

5.5 Dashboard Development and Structure

To calculate the environmental impacts of LLMs during inference, the formulas described in the preceding chapter were applied to the defined marketing tasks and implemented in a Microsoft Excel spreadsheet. Independent variables such as total parameters, active parameters, and number of output tokens were drawn from the model and task definitions. The spreadsheet estimates energy consumption, CO_2 emissions, electricity costs, and material usage. These results were visualised in an interactive Power BI dashboard designed to support non-technical business users. The dashboard includes three main views.

The *Calculation tab* (Figure 1) allows users to select a task and model and displays absolute environmental and financial impacts, including kg Sb-eq, Wh, gCO_2, CHF, and everyday equivalents such as smartphone charges.

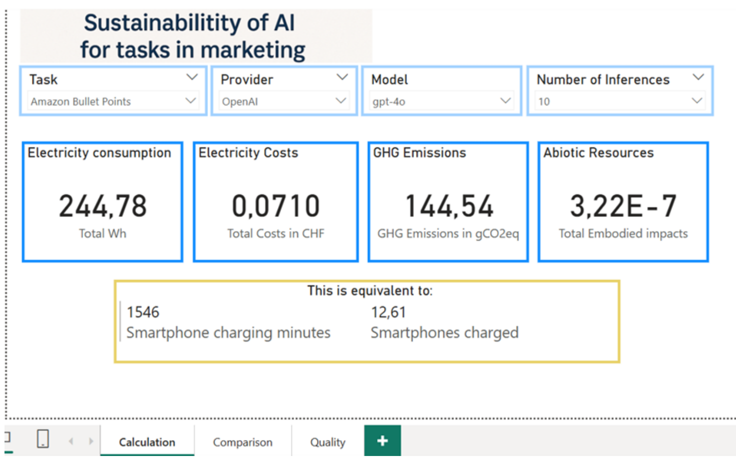

Fig. 1. Power BI dashboard: Calculation tab

The *Comparison tab* (Figure 2) enables direct comparison of models for a selected task, highlighting relative differences in energy use and emissions.

Finally, the *Quality tab* (Figure 3) integrates qualitative performance scores from the Artificial Analysis Intelligence Index [1], supporting more holistic evaluation by including aspects beyond sustainability.

5.6 User Input and Automation

The artefact requires only one active user input: the expected number of output tokens for the selected task. All other calculations are automatically performed in the background based on predefined logic and model data, making the system highly usable and accessible.

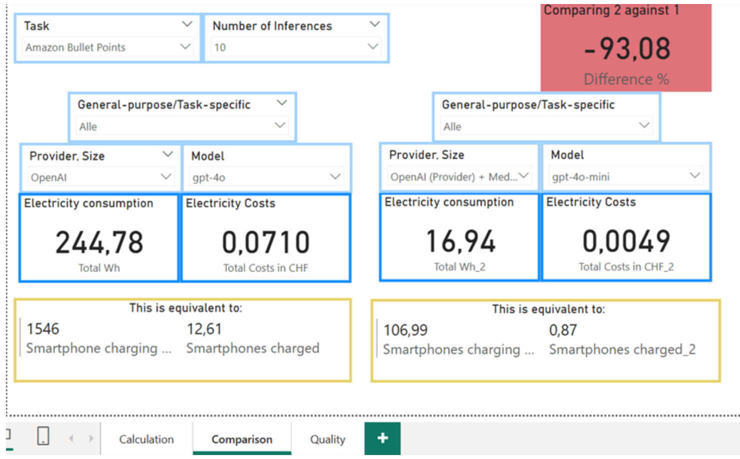

Fig. 2. Power BI dashboard: Comparison tab

5.7 Conclusion

This artefact translates complex sustainability indicators into a visual and accessible tool for business users. It enables awareness and supports more informed model selection decisions, aligning AI adoption with corporate environmental goals.

6 Evaluation

This section assesses the artefact's effectiveness in raising awareness of the environmental and cost implications of LLM usage. Evaluation workshops with content creators, strategic decision-makers, and sustainability professionals were held to determine whether the dashboard supports informed model selection.

6.1 Evaluation Method

Three key user groups were selected:

- *Users:* marketing content creators using LLMs.
- *Management:* decision-makers in AI tool selection and digital strategy.
- *Sustainability:* staff involved in environmental reporting and corporate sustainability.

Each participant joined an individual workshop (except the content creators, who were interviewed together). After a brief walkthrough of the artefact and its objectives, participants completed a task simulation comparing the impacts of creating a social media post using ChatGPT-4 and ChatGPT-4o-mini. This prompted open discussions based on pre-defined questions.

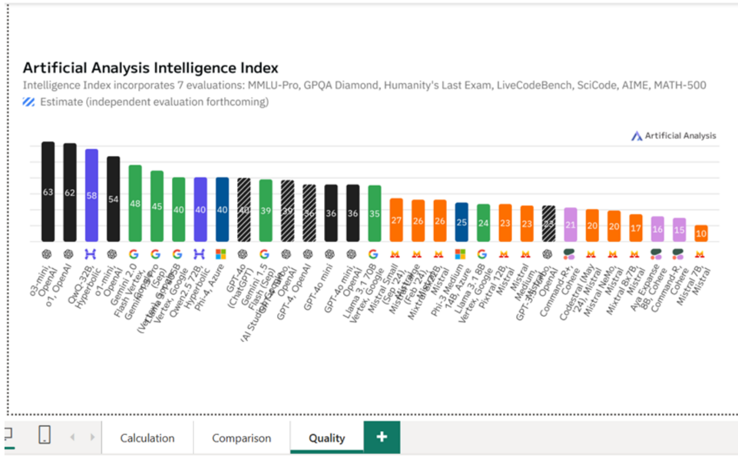

Fig. 3. Power BI dashboard: Quality tab

6.2 Evaluation Results

Awareness and Understanding The artefact significantly improved participants' awareness. All reported limited prior knowledge of LLM-related energy use and environmental impact. Visualising effects using relatable units (e.g., smartphone charges) consistently led to "aha moments." The 99.44% energy difference between GPT-4 and GPT-4o-mini surprised all participants and sparked reflection.

While electricity use and CO_2 emissions were widely understood, abiotic resource consumption was perceived as abstract. Suggestions included adding more real-world equivalences (e.g., spoons, trees, coffee cups) to improve comprehension.

Decision-Making Support Participants agreed the artefact could inform model choices, but noted that widespread adoption would require a shift in culture, leadership endorsement, and greater awareness. Although sustainability matters, it competes with other priorities like performance, server location, cost, and security. As a result, participants saw sustainability decisions as a strategic rather than individual concern.

The dashboard's insights could support internal reporting and align with company-wide sustainability targets. However, limited provider transparency and the absence of reporting standards remain barriers. AI-related emissions currently fall under Scope 3 and are not subject to legal disclosure, but tracking them now may provide a future advantage and position the company as a responsible innovator.

Feedback and Improvement Suggestions Across groups, participants provided constructive feedback:

- Clearer explanation of sustainability metrics, especially for abiotic resources.

- Additional equivalences for CO_2 (e.g., trees, car mileage).
- Default model settings of sustainable models for frequent tasks in the internal chatbot.
- Integration of gamification or visual alerts to promote suitable model selection.

These suggestions were noted for further development and potential iteration.

6.3 Conclusion

The evaluation confirms that the artefact improves awareness of LLM sustainability impacts and supports more informed decisions. By translating abstract metrics into accessible visuals, the tool addresses a blind spot in AI adoption. The feedback validates the framework's relevance while also pointing to opportunities for enhancement.

7 Discussion and Limitations

The evaluation confirmed the artefact's potential to support sustainability-oriented model selection in business. Stakeholders recognised the added value of visualising environmental and cost metrics, especially as such data had previously been absent from decision-making workflows. However, the evaluation also surfaced notable challenges.

Firstly, while the dashboard improved awareness, this does not automatically lead to behavioural change. Decision-makers showed interest in the tool, but end users preferred pre-configured or default options that simplify sustainable choices. This points to a gap between awareness and action and highlights the need for organisational mechanisms such as intelligent defaults or embedded sustainability prompts to guide low-effort decisions.

Secondly, a knowledge gap persists across user groups. Most lacked the expertise to assess trade-offs between model performance and environmental impact, limiting their ability to independently interpret the dashboard. Broader adoption of sustainable AI practices may therefore require targeted training and improved internal communication.

The evaluation itself also had limitations. The framework was tested within a single department and assessed in one round of stakeholder workshops, limiting generalisability. Broader validation across functions and industries is needed to assess adaptability and long-term behavioural impact.

Lastly, the artefact currently focuses only on inference impacts for text-based LLMs. Expanding to other modalities, such as image or video generation, and including training-related emissions would increase relevance and provide a more holistic view of AI's environmental footprint.

8 Conclusion and Future Work

The main contribution of this paper is the development of an awareness framework for raising awareness of the environmental impacts of LLMs. This framework enables organisations to make more informed, sustainability-oriented model selection decisions by assessing the environmental and financial implications of LLM usage in business. Developed using a DSR approach, the artefact allows users to compare general-purpose and task-specific models across key sustainability indicators. Based on real marketing tasks and stakeholder input, it translates complex metrics—such as energy consumption, CO_2 emissions, and material use—into an accessible decision-support tool.

The evaluation showed that the artefact improves awareness and can support more sustainable AI adoption. Stakeholders appreciated the visualisation of impacts, especially where LLM inference emissions had previously been overlooked. Task-based calculations and relatable equivalents made the results easy to understand and apply in decision-making.

However, some limitations remain. The framework was tested in a single department and only once, limiting broader applicability. Also, awareness alone does not guarantee behavioural change; structural support like intelligent defaults or workflow integration is likely needed.

Future work should expand the framework to cover multi-modal use cases, such as image or video generation, and include training-related emissions. Additionally, future research should explore embedding sustainability considerations into internal AI governance and reporting processes and introducing "sustainable" AI models as default options for users. Longitudinal studies would help evaluate its long-term effect on decisions and emissions reporting.

References

1. Artificial Analysis: Artificial Analysis Intelligence Benchmarking Methodology (Feb 2025), `https://artificialanalysis.ai/methodology/intelligence-benchmarking`
2. Bannour, N., Ghannay, S., Névéol, A., Ligozat, A.L.: Evaluating the carbon footprint of NLP methods: a survey and analysis of existing tools. In: Moosavi, N.S., Gurevych, I., Fan, A., Wolf, T., Hou, Y., Marasović, A., Ravi, S. (eds.) Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing. pp. 11–21. Association for Computational Linguistics, Virtual (Nov 2021). `https://doi.org/10.18653/v1/2021.sustainlp-1.2`, `https://aclanthology.org/2021.sustainlp-1.2`
3. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F.: Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion **58**, 82–115 (Jun 2020). `https://doi.org/10.1016/j.inffus.2019.12.012`, `https://www.sciencedirect.com/science/article/pii/S1566253519308103`
4. Berthelot, A., Caron, E., Jay, M., Lefèvre, L.: Estimating the environmental impact of Generative-AI services using an LCA-based methodology. Procedia CIRP **122**, 707–712 (Jan 2024). `https://doi.org/10.1016/j.procir.2024.01.098`, `https://www.sciencedirect.com/science/article/pii/S2212827124001173`

5. Bousquette, I.: These AI Models Are Pretty Mid. That's Why Companies Love Them. Wall Street Journal (Jul 2024), https://www.wsj.com/articles/these-ai-models-are-pretty-mid-thats-why-companies-love-them-710a0f72
6. Budennyy, S.A., Lazarev, V.D., Zakharenko, N.N., Korovin, A.N., Plosskaya, O.A., Dimitrov, D.V., Akhripkin, V.S., Pavlov, I.V., Oseledets, I.V., Barsola, I.S., Egorov, I.V., Kosterina, A.A., Zhukov, L.E.: eco2AI: Carbon Emissions Tracking of Machine Learning Models as the First Step Towards Sustainable AI. Doklady Mathematics **106**(1), S118–S128 (Dec 2022). https://doi.org/10.1134/S1064562422060230, https://doi.org/10.1134/S1064562422060230
7. Chofreh, A.G., Goni, F.A.: Review of Frameworks for Sustainability Implementation. Sustainable Development **25**(3), 180–188 (2017). https://doi.org/10.1002/sd.1658, https://onlinelibrary.wiley.com/doi/abs/10.1002/sd.1658, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sd.1658
8. Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., Zemmel, R.: The economic potential of generative AI. Tech. rep., McKinsey&Company (Jun 2023)
9. Chung, J.W., Liu, J., Ma, J.J., Wu, R., Kweon, O.J., Xia, Y., Wu, Z., Chowdhury, M.: The ML.ENERGY Benchmark: Toward Automated Inference Energy Measurement and Optimization (May 2025). https://doi.org/10.48550/arXiv.2505.06371, http://arxiv.org/abs/2505.06371, arXiv:2505.06371 [cs]
10. Dauvergne, P.: Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs. Review of International Political Economy **29**(3), 696–718 (May 2022). https://doi.org/10.1080/09692290.2020.1814381, https://doi.org/10.1080/09692290.2020.1814381, publisher: Routledge _eprint: https://doi.org/10.1080/09692290.2020.1814381
11. Desroches, C., Chauvin, M., Ladan, L., Vateau, C., Gosset, S., Cordier, P.: Exploring the sustainable scaling of AI dilemma: A projective study of corporations' AI environmental impacts (Jan 2025). https://doi.org/10.48550/arXiv.2501.14334, http://arxiv.org/abs/2501.14334, arXiv:2501.14334 [cs]
12. Fozooni, A., Nazari, S., Jamalpur, A.: Prioritizing big data applications in E-commerce considering sustainable development indicators. Journal of Future Sustainability **4**(3), 169–178 (2024). https://doi.org/10.5267/j.jfs.2024.9.002, http://www.growingscience.com/jfs/Vol4/jfs_2024_15.pdf
13. Fraiwan, M., Khasawneh, N.: A Review of ChatGPT Applications in Education, Marketing, Software Engineering, and Healthcare: Benefits, Drawbacks, and Research Directions (Apr 2023). https://doi.org/10.48550/arXiv.2305.00237, http://arxiv.org/abs/2305.00237, arXiv:2305.00237
14. GenAI Impact: LLM Inference - EcoLogits (2024), https://ecologits.ai/latest/methodology/llm_inference/
15. Hevner, A.R.: A Three Cycle View of Design Science Research. Scandinavian Journal of Information Systems **19** (2007), https://core.ac.uk/download/pdf/301370809.pdf
16. Husom, E.J., Goknil, A., Shar, L.K., Sen, S.: The Price of Prompting: Profiling Energy Use in Large Language Models Inference (Jul 2024). https://doi.org/10.48550/arXiv.2407.16893, http://arxiv.org/abs/2407.16893, arXiv:2407.16893 [cs] version: 1
17. Ichien, N., Liu, Q., Fu, S., Holyoak, K.J., Yuille, A.L., Lu, H.: Two Computational Approaches to Visual Analogy: Task-Specific Models Versus Domain-General Mapping. Cognitive Science **47**(9), e13347 (2023). https://doi.org/10.1111/cogs.13347, https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.13347, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cogs.13347
18. International Organization for Standardization: ISO 14044:2006(en), Environmental management — Life cycle assessment — Requirements and guidelines (Jul 2006), https://www.iso.org/obp/ui/#iso:std:iso:14044:ed-1:v1:en
19. Kshetri, N., Dwivedi, Y.K., Davenport, T.H., Panteli, N.: Generative artificial intelligence in marketing: Applications, opportunities, challenges, and research agenda. International Journal of Information Management **75**, 102716

(Apr 2024). https://doi.org/10.1016/j.ijinfomgt.2023.102716, https://www.sciencedirect.com/science/article/pii/S026840122300097X
20. Lacoste, A., Luccioni, A., Schmidt, V., Dandres, T.: Quantifying the Carbon Emissions of Machine Learning (Nov 2019), http://arxiv.org/abs/1910.09700, arXiv:1910.09700 [cs]
21. Lee, S.U., Perera, H., Liu, Y., Xia, B., Lu, Q., Zhu, L., Cairns, J., Nottage, M.: Integrating ESG and AI: A Comprehensive Responsible AI Assessment Framework. AI and Ethics **5**(5), 5121–5148 (Oct 2025). https://doi.org/10.1007/s43681-025-00741-5, https://link.springer.com/10.1007/s43681-025-00741-5
22. Li, P., Yang, J., Islam, M.A., Ren, S.: Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models (Oct 2023). https://doi.org/10.48550/arXiv.2304.03271, http://arxiv.org/abs/2304.03271, arXiv:2304.03271 [cs]
23. Ligozat, A.L., Lefevre, J., Bugeau, A., Combaz, J.: Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessment of AI Solutions. Sustainability **14**(9), 5172 (Jan 2022). https://doi.org/10.3390/su14095172, https://www.mdpi.com/2071-1050/14/9/5172, number: 9 Publisher: Multidisciplinary Digital Publishing Institute
24. Luccioni, A.S., Viguier, S., Ligozat, A.L.: Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model (Nov 2022), https://arxiv.org/abs/2211.02001v1
25. Luccioni, S., Gamazaychikov, B., Hooker, S., Pierrard, R., Strubell, E., Jernite, Y., Wu, C.J.: Light bulbs have energy ratings — so why can't AI chatbots? Nature **632**(8026), 736–738 (Aug 2024). https://doi.org/10.1038/d41586-024-02680-3, https://www.nature.com/articles/d41586-024-02680-3, bandiera_abtest: a Cg_type: Comment Publisher: Nature Publishing Group Subject_term: Climate change, Computer science, Energy, Machine learning, Sustainability
26. Luccioni, S., Gamazaychikov, B., Strubell, E., Hooker, S., Jernite, Y., Wu, C.J., Mitchell, M.: AI Energy Score Leaderboard (2025), https://huggingface.co/spaces/AIEnergyScore/Leaderboard
27. Luccioni, S., Jernite, Y., Strubell, E.: Power Hungry Processing: Watts Driving the Cost of AI Deployment? In: The 2024 ACM Conference on Fairness, Accountability, and Transparency. pp. 85–99. ACM, Rio de Janeiro Brazil (Jun 2024). https://doi.org/10.1145/3630106.3658542, https://dl.acm.org/doi/10.1145/3630106.3658542
28. Rincé, S., Banse, A., Defour, V.: EcoLogits Calculator (2025), https://huggingface.co/spaces/genai-impact/ecologits-calculator
29. Rohde, F., Wagner, J., Meyer, A., Reinhard, P., Voss, M., Petschow, U., Mollen, A.: Broadening the perspective for sustainable artificial intelligence: sustainability criteria and indicators for Artificial Intelligence systems. Current Opinion in Environmental Sustainability **66**, 101411 (Feb 2024). https://doi.org/10.1016/j.cosust.2023.101411, https://www.sciencedirect.com/science/article/pii/S1877343523001586
30. Rohde, F., Wagner, J., Reinhard, P., Petschow, U., Meyer, A., Voß, M., Mollen, A.: Nachhaltigkeitskriterien für künstliche Intelligenz. Berlin (Dec 2021), https://www.alexandria.unisg.ch/server/api/core/bitstreams/68a50206-cda3-4aa8-b5ca-e9f3a1c5e140/content
31. Samsi, S., Zhao, D., McDonald, J., Li, B., Michaleas, A., Jones, M., Bergeron, W., Kepner, J., Tiwari, D., Gadepally, V.: From Words to Watts: Benchmarking the Energy Costs of Large Language Model Inference (Oct 2023). https://doi.org/10.48550/arXiv.2310.03003, http://arxiv.org/abs/2310.03003, arXiv:2310.03003 [cs]
32. Triguero, I., Molina, D., Poyatos, J., Del Ser, J., Herrera, F.: General Purpose Artificial Intelligence Systems (GPAIS): Properties, definition, taxonomy, societal implications and responsible governance. Information Fusion **103**, 102135 (Mar 2024). https://doi.org/10.1016/j.inffus.2023.102135, https://www.sciencedirect.com/science/article/pii/S1566253523004517

33. Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S.D., Tegmark, M., Fuso Nerini, F.: The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications **11**(1), 233 (Jan 2020). https://doi.org/10.1038/s41467-019-14108-y, https://www.nature.com/articles/s41467-019-14108-y, publisher: Nature Publishing Group
34. Wang, N., Li, S., Wang, C., Zhao, L.: Current Status and Emerging Trends of Generative Artificial Intelligence Technology: A Bibliometric Analysis. Journal of Internet Technology **25**(3), 477–485 (May 2024), https://jit.ndhu.edu.tw/article/view/3065, number: 3
35. Wright, D., Igel, C., Samuel, G., Selvan, R.: Efficiency is Not Enough: A Critical Perspective of Environmentally Sustainable AI (Sep 2023). https://doi.org/10.48550/arXiv.2309.02065, http://arxiv.org/abs/2309.02065, arXiv:2309.02065 [cs, stat]
36. Xenova: The Tokenizer Playground - a Hugging Face Space by Xenova, https://huggingface.co/spaces/Xenova/the-tokenizer-playground