Measurement 268 (2026) 120654

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Environmental sustainability of Artificial Intelligence: A GUM-based, user-centered measurement framework

Leopoldo Angrisani [a], Mauro D'Arco [a], Egidio De Benedetto [a],*, Luigi Duraccio [a], Immacolata Esposito [a], Annarita Tedesco [b]

[a] University of Naples Federico II - Department of Information Technology and Electrical Engineering, Naples, Italy
[b] University of Naples Federico II - Department of Public Health, Naples, Italy

ARTICLE INFO

Keywords:
Artificial intelligence
Carbon footprint
Life-cycle assessment
Measurements
Metrology
Sustainability
Measurement sustainability

ABSTRACT

Given the increasing adoption of artificial intelligence (AI) across a broad spectrum of applications, along with the urgent need for sustainable development, understanding the environmental sustainability of AI pipelines has become increasingly relevant. In this regard, however, the current state of the art lacks a reliable methodology for measuring environmental sustainability from a user-centered perspective (*i.e.*, by considering all the operations typically performed by end users), which is essential for achieving awareness of the actual sustainability in the development and adoption of AI models. Starting from these considerations, this paper employs a rigorous methodology based 1) on the ISO standards for sustainability assessment and 2) on the Guide to the Expression of Uncertainty in Measurement (GUM) to measure and aggregate the *Carbon Footprint* required by each stage of an AI pipeline. To the best of the authors' knowledge, this work constitutes the first to integrate GUM-based approaches into a user-driven AI pipeline. To illustrate the methodology, a case study on the use of AI models in biosignal processing is presented. Without losing generality, the results provide useful insights for implementing more sustainable AI practices, enabling a reliable, environment-oriented assessment of AI pipelines and guiding decisions toward reduced environmental impact.

1. Introduction

The concept of *Sustainability*, defined as *"meeting the needs of the present generation without compromising the ability of future generations to meet their own needs"* [1], is a cornerstone of contemporary scientific and social discourse. Among the dimensions of Sustainability, Environmental Sustainability has become particularly critical due to the urgent need to mitigate the impacts of excessive resource consumption and pollution, which contribute significantly to climate change and global warming [2].

In this scenario, the role of Information and Communication Technologies (ICTs) has garnered growing attention. In fact, the ICT sector encompasses a wide range of technologies that can provide tools to enhance resource management and system efficiency; however, at the same time, the very implementation of ICTs may adversely affect the environment. This dual effect is captured by the concepts of *Green by ICT* and *Green of ICT*, respectively [3].

An example of this duality is represented by Artificial Intelligence (AI), one of the core ICTs [4–6]. As highlighted in [7], AI enables large-scale data analysis, instrumental in supporting ecosystem preservation and improving energy and waste management [8]. Nevertheless, such data processing demands can require significant computational power, thus leading to increased energy consumption and emissions [9]. For this reason, understanding the environmental impact of AI technology has become a crucial concern.

In the literature, several studies offer estimates of the environmental impact of AI from different perspectives. For instance, some studies provide large-scale assessments that consider the deployment and operation of large models, such as those used in Natural Language Processing (NLP), focusing on the energy demands of data centers [10–12]. Other studies address the environmental impact of AI from a more *user-centered* perspective [13–17], *i.e.*, analyzing the effects of AI usage by end users [18]. These studies either employ existing tools or develop custom mathematical models specifically designed to estimate environmental impact. However, although such tools are paving the way toward greater user awareness of the environmental impact of AI adoption, they still present some limitations:

- The environmental impact estimation mainly focuses on the training and validation of the AI models, often overlooking the effort

* Corresponding author.
E-mail address: egidio.debenedetto@unina.it (E. De Benedetto).

https://doi.org/10.1016/j.measurement.2026.120654
Received 30 December 2025; Received in revised form 20 January 2026; Accepted 29 January 2026
Available online 30 January 2026

required by users for data collection, data preparation, and model finalization.

- The emphasis is primarily on energy consumption, often neglecting the environmental impact associated with the life-cycle of hardware and software components.
- Most importantly, there is a lack of standardized approaches for coupling environmental impact estimations with their associated uncertainty, which undermines the reliability of sustainability assessments and may lead to misleading conclusions.

Starting from these considerations, this paper aims to provide a reliable approach for measuring the environmental impact associated with the use of AI pipelines from a *user-centered* perspective, where the expression *user-centered* explicitly refers to considering all the operations that are typically performed by end-users throughout the practical use of the AI pipeline. It is important to point out that, to ensure trustworthy outcomes, the adopted methodology leverages ISO standards for sustainability assessment and the *Guide to the Expression of Uncertainty in Measurement* (GUM) principles, Notably, the proposed methodology addresses not only the energy-related impact but also the life-cycle effects of hardware and software components throughout all the stages of the AI pipelines (from data acquisition and preparation to model validation and finalization, without being limited to the case of a user employing pre-trained models), also quantifying the uncertainty related to the resulting outcomes. To the best of the authors' knowledge, this work constitutes the first time that such approaches are integrated into a user-driven AI pipeline.

Without losing generality, the methodology was applied to a case study involving biosignal data classification through AI, a topic of growing relevance [19]. This case study addressed how to measure the environmental impact related to end users' employment of AI pipelines; also, it serves as a practical example to assess how one could intervene to improve sustainability without degradation in performance.

The paper is organized as follows. Section 2 provides an overview of sustainability assessment, outlining the foundational concepts employed in Section 3 to introduce the adopted methodology. Section 4 presents a case study demonstrating the methodology's application. Section 5 discusses the experimental results obtained. Finally, conclusions are drawn and the future work is outlined.

2. Background

The *Sustainable Development Goals* (SDG) of the *United Nations 2030 Agenda SDG 13* of the *2030 Agenda* [20,21] highlight the importance of *"taking urgent action to combat climate change and its impacts"*. Such emphasis on combating climate change aligns with the ISO *14000* series of standards [22,23], which were designed to guide organizations in environmental management. Building on these standards, the International Telecommunication Union (ITU) has also proposed a set of recommendations for the assessment of the environmental impact of ICT products and systems [24]. At the core of this assessment is the quantification of greenhouse gas (GHG) emissions over a specified timeframe, as these emissions are the primary drivers of contemporary global warming. This process is represented by the concept of *Carbon Footprint*, namely the most widely adopted indicator in environmental assessments,[1] expressed in kilograms of equivalent carbon dioxide ($CO_{2,eq}$). This metric standardizes emissions by comparing the radiative forcing of various GHGs to that of carbon dioxide. The process involves converting the emissions of gases such as methane and ozone into equivalent CO_2 amounts based on their global warming potential (GWP), enabling consistent evaluation. Hence, in this framework, the Carbon Footprint serves as a quantitative metric of the environmental impact of ICTs.

However, the estimation of an ICT product's or service's Carbon Footprint cannot rely on the direct measurement of GHG emissions, as these emissions predominantly originate from upstream processes. For instance, consider the case of a laptop: although its use phase involves electricity consumption that leads to GHG emissions, it is not feasible to directly quantify the amount of $CO_{2,eq}$ attributable to the device itself, since these emissions are primarily associated with the generation and distribution of electricity. Consequently, the estimation of the Carbon Footprint requires the application of indirect modeling approaches. This is generally performed by means of the *Life-Cycle Assessment (LCA)* methodology [22], which considers the GHG emissions related to its entire life cycle, including manufacturing, distribution, usage, and disposal. According to this framework, the process is structured into three key steps:

1. *Checklist of the items*: this initial step involves developing a comprehensive inventory of the items comprising the specific product or service under evaluation. This inventory acts as the primary database for the estimation. Clearly defined boundaries, tailored to the specific case at hand, delineate what can be included or excluded. A standard checklist encompassing six categories, namely (i) hardware, (ii) software, (iii) consumables and supportive products, (iv) site infrastructure, (v) goods transportation, and (vi) personnel movements, is often adopted.
2. *Impact Assessment*: each item of the checklist undergoes an analysis to distinguish between *embodied* and *operational* impacts. Embodied impacts arise from the production, transportation, packaging, and disposal processes and can often be mitigated or avoided through the reuse of items, aligning with circular economy principles [25]. Operational impacts, on the other hand, pertain to the item's energy consumption during use. Despite the operational impacts are unavoidable, they can be quantified over the specified period. Considering both types of impacts enables a comprehensive environmental evaluation of each item.
3. *Impact Aggregation*: the final step aggregates the embodied and operational impacts of all the items considered to estimate the overall Carbon Footprint of the product or system under evaluation.

As a foundational approach, this ISO-ITU framework has established the basis for assessing the environmental impact of ICTs, particularly serving as the cornerstone for current tools used to estimate the Carbon Footprint of AI models. However, since these tools exhibit the limitations outlined in the Introduction, which hinder the achievement of reliable outcomes, a methodology that appropriately integrates the described framework with the principles of the GUM can prove valuable in providing not only an estimate but a fully-fledged measurement result of the Carbon Footprint.

3. Materials and methods

This section describes the key stages of AI pipelines, along with the methodology adopted for measuring the related Carbon Footprint.

3.1. Stages of AI pipelines

From a user-centered perspective, an AI pipeline represents the entire workflow through which AI models are developed and finalized by end users. This encompasses the whole process that can typically be divided into four key stages:

1. *Data Collection*: this stage encompasses the acquisition of data, which may be sourced from publicly available datasets or collected directly through on-site experimental campaigns.

[1] Other indicators, such as water or material footprint, may also be employed to evaluate environmental sustainability. Nevertheless, the methodological considerations and analytical framework developed in this study are, to a large extent, transferable to these alternative indicators.

2. *Data Preparation*: after collection, data are typically prepared for model training by handling outliers or missing values, normalizing or standardizing, or performing feature engineering to enhance the model's performance.
3. *Model Validation*: in this stage, the AI model is validated using appropriate validation strategies. Typically, the dataset is divided into *training data* and *test data*. The training data are used to enable the model to learn underlying patterns. The model's performance is subsequently assessed on *test data*. Additional hyperparameter optimization can typically be performed to enhance the model's effectiveness.
4. *Model Finalization*: it involves training the AI model on the entire dataset to extract the maximum amount of information and refine its predictive capabilities. This stage ensures the model is fully optimized and ready for application in real-world scenarios.

A comprehensive measurement of the environmental impact of AI pipelines requires evaluating the Carbon Footprint of each stage individually, considering both the embodied and operational contributions.

3.2. Sustainability measurement

The proposed methodology for measuring the sustainability of AI pipelines adopts the ISO-ITU framework described in Section 2 and integrates it with the GUM guidelines to achieve a reliable measure of the Carbon Footprint.

In particular, following the approach outlined in [26,27], *Supplement 1* of the GUM [28] is leveraged. Specifically, for each stage of the AI pipeline, the proposed methodology is based on the following steps:

1. *Checklist of the items*: in line with the ISO-ITU framework described in Section 2, the first step involves developing a checklist of items characterizing the stage. Depending on the specific case study at hand, it is explicitly stated which categories are included in the impact assessment and which are excluded.
2. *PDFs assignment*: in this step, the embodied and operational impacts associated with all items belonging to the stage. In particular, a probability density function (PDF) is assigned to each impact. These PDFs represent the incomplete knowledge of a *true value* of the Carbon Footprint of each item, which depends on the modeling strategy and the source used for the estimation, *e.g.* literature, technical reports, or specific software for LCA. Notably, the choice of the PDF type may be carried out in accordance with the guidance provided by the GUM, or, when it is not feasible, based on the available knowledge through a subjective Bayesian approach, wherein expert judgment is formally incorporated into the uncertainty representation.
3. *PDFs propagation*: the individual PDFs are then aggregated through Monte Carlo (MCM) simulations, involving a finite number of extractions for each input variable, under the assumption that the overall Carbon Footprint of the stage under evaluation is the sum of the Carbon Footprints of each item belonging to the stage.
4. *Coverage interval*: the final step allows the determination of a coverage interval representing the measure of the overall Carbon Footprint of the stage. This step aids in identifying the most impactful stages of the entire AI pipeline, which is crucial for recognizing where interventions can enhance the pipeline's sustainability without significantly compromising performance.

In Fig. 1, a sketch of the proposed methodology is represented.

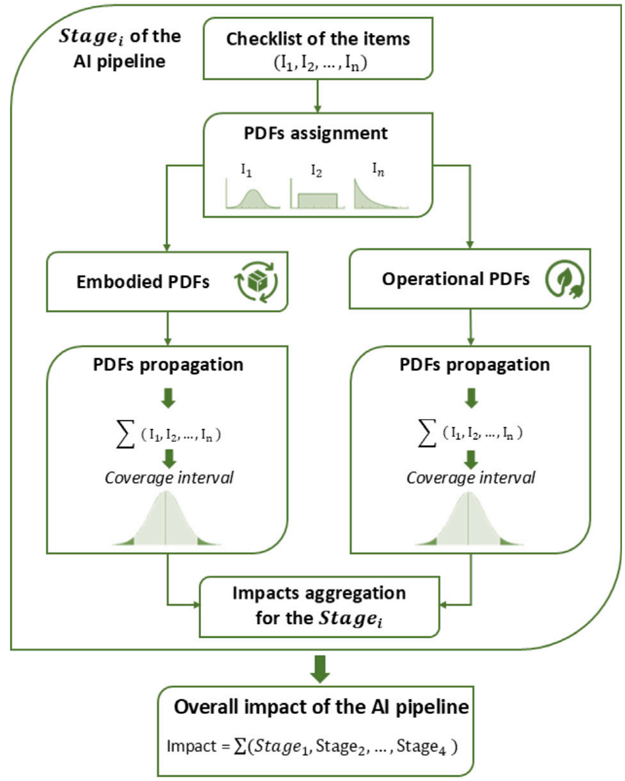

Fig. 1. Graphical representation of the methodology used to measure the environmental sustainability of an AI pipeline.

4. Case study

As a case study, the classification of biosignals, in particular Steady-State Visually Evoked Potentials (SSVEPs), was considered [29]. These biosignals represent the brain's response to visual stimulation. Since they can be easily acquired through electroencephalography (EEG), they are of significant value in the healthcare domain. In fact, proper acquisition and processing of SSVEPs assist patients by enabling alternative communication channels, facilitating attention monitoring, and providing control over assistive devices, such as wheelchairs [30].

4.1. AI model

Without losing generality, the Filter-Bank Canonical Correlation Analysis (FBCCA) algorithm [31] was chosen as the AI model to be employed.[2] It consists of three main steps: (i) applying a filter bank to the original EEG signal; (ii) performing CCA operation between the sub-band components of the EEG signal and sinusoidal reference signals; and (iii) classifying the resulting outputs.

The filter bank decomposes the original EEG signal X into multiple sub-band components X_s, where $s = 1, 2, \ldots, S$ is the sub-band index, by applying a set of filters with distinct pass-bands. For each sub-band component X_s, standard CCA is performed to calculate the correlation coefficients between the sub-band signal X_s and a sinusoidal reference signal Y_n, with the index $n = 1, 2, \ldots, N$ corresponding to the frequencies of the visual stimulation. Notably, each sinusoidal reference signal Y_n is constructed by considering a specified number

[2] The choice of the FBCCA model was motivated by its recognition as a state-of-the-art classification approach for SSVEPs. Nevertheless, the proposed methodology is not model-specific and can be analogously applied to alternative models, including neural networks.

of harmonics H. Hence, for each frequency n and sub-band s, the calculated correlation coefficient is denoted as ρ_s^n, and the vector of these correlation coefficients is represented as:

$$\rho^n = \left[\rho_1^n, \rho_2^n, \dots, \rho_s^n, \dots, \rho_S^n\right] \tag{1}$$

Subsequently, a weighted sum of the squared correlation coefficients, $\tilde{\rho}^n$, is computed across all sub-band components for each stimulation frequency n:

$$\tilde{\rho}^n = \sum_{s=1}^{S} w(s) \cdot (\rho_s^n)^2 \tag{2}$$

To account for the decrease in the Signal-to-Noise Ratio (SNR) of SSVEP harmonics at higher frequencies, the weighting function $w(s)$ is applied to the sub-band components. The weights are defined as:

$$w(s) = s^{-a} + b \tag{3}$$

where a and b are parameters that undergo optimization to improve classification performance.

After obtaining N features $\tilde{\rho}^n$ (one for each stimulation frequency), the signal classification process involves identifying the stimulation frequency f_z ($z = 1, \dots, N$) corresponding to the feature $\tilde{\rho}^z$ with the highest value. This frequency f_z corresponds to the frequency of the visual stimulation that induced the SSVEP in the EEG signal X considered.

Overall, the performance of the FBCCA algorithm depends on four key parameters: (i) the number of sub-bands S, (ii) the number of harmonics H considered for the sinusoidal reference signals in the CCA, and (iii) the parameters a and b that weight the correlation coefficients for each stimulation frequency.

4.2. Data collection

Since data collection can be conducted in two different ways, two scenarios were considered:

1. *Scenario #1*: download of a public benchmark dataset. Concerning this scenario, the SSVEP dataset used in [32] was chosen. It is a widely recognized benchmark for SSVEPs processing, featuring 64-channel EEG recordings from 35 subjects, who performed a cue-guided target selection task with 40 visual stimuli. The stimuli were rendered at frequencies between 8.0 Hz and 15.8 Hz (0.2-Hz span), with each trial lasting up to 5 s. Each participant completed six acquisition blocks of 40 trials each. The size of the dataset was about 100 GB.
2. *Scenario #2*: acquisition of new data through on-site experiments and upload online. This scenario involved an experimental campaign with 30 healthy participants. Participation was voluntary and uncompensated, with written informed consent obtained from all participants. The study was also approved by the Ethical Committee of Psychological Research at the University of Naples Federico II. SSVEPs were elicited using flickering visual stimuli generated by an application running on an Augmented Reality Head-Mounted Display (AR HMDs). Each participant completed five acquisition blocks, each consisting of eight trials corresponding to eight flickering stimuli ranging from 8 to 15 Hz with 1-Hz span. EEG signals were recorded with an EEG headset featuring 8 EEG channels and a sampling rate of 250 sps. Also in this case, the resulting dataset had a size of about 100 GB.

Both scenarios required essential components, such as laptops, internet modems, and cloud services. However, the second scenario involves additional equipment, including the EEG signal acquisition instrument and the AR HMD.

4.3. Data preparation

With reference to data preprocessing strategies, neither normalization nor standardization was required for the FBCCA algorithm to perform efficiently. As the dataset preparation already included a notch filter to eliminate disturbances caused by powerline interference, no additional filtering was applied. However, standard data preprocessing steps were carried out to format and prepare the input data for optimal compatibility with the AI model.

4.4. Model validation

Given the significant inter-individual variability in SSVEP signals [33], Leave-One-Subject-Out Cross-Validation (LOSO CV) was used for validation. LOSO CV is a variation of k-fold cross-validation, where each fold corresponds to a specific subject in the dataset. In each LOSO iteration, a grid search was conducted on the training data to find the optimal combination of parameters (S, H, a, b) that yielded the highest accuracy. This combination was then tested on the test data, producing a test accuracy value. After all iterations, the algorithm's performance was measured by obtaining the mean and standard uncertainty of the test accuracy across all subjects, this latter through a type-A evaluation.

In this work, the following grid was employed for both scenarios described in Section 4.2, with a total of 2160 possible combinations: S = 1:8 with a step of 1; H = 1:6 with a step of 1; a = 0:2 with a step of 0.25; and b = 0:1 with a step of 0.25.

4.5. Model finalization

Following the validation process aimed at evaluating the performance, the FBCCA algorithm underwent a final training phase using the entire dataset and all parameter combinations from the grid search. This approach ensured that the finalized model maximized its generalization capability, making it highly effective for future real-world applications [34].

5. Experimental results

This Section details the experimental results obtained at each stage of the AI pipeline, following the 4-step methodology presented in Section 3.2.

5.1. Checklist of the items

The checklist of items was developed based on the two different scenarios mentioned in Section 4.2: (i) downloading a public dataset and (ii) acquiring new data. For the specific case study considered, the primary categories were hardware and software, as aspects such as the travel of individuals and the storage of goods and services were outside the scope of this work.

- *Scenario #1*: the data download phase involved the use of a cloud server, an internet modem, and a laptop (namely a Lenovo Thinkpad T15g), the latter employed throughout the entire pipeline. The dataset was preprocessed, validated, and finalized considering the FBCCA model implemented in MATLAB environment, serving as the software platform for all stages of the pipeline.
- *Scenario #2*: the data acquisition phase required an EEG headset, namely the *g.tec UNICORN Hybrid Black* [35], and a stimulation platform, namely the *Microsoft Hololens 2* AR HMD [36], for rendering the visual stimuli. The data upload phase involved a cloud server, an internet modem, and a laptop. As in Scenario #1, this laptop was used for all the stages of the pipeline with MATLAB software.

This checklist is also summarized in Table 1.

Table 1
Checklist of the items for both scenarios of the AI pipeline.

AI pipeline stage	Scenario #1	Scenario #2
Data collection	Laptop Modem Server	Laptop Modem Server AR HMD EEG headset Software
Data preparation	Laptop Software	Laptop Software
Model validation	Laptop Software	Laptop Software
Model finalization	Laptop Software	Laptop Software

5.2. PDFs assignment

The impact assessment of each item of the provided checklist was carried out as follows:

- Regarding the *Embodied* impacts, which arise from the production, transportation, packaging, and disposal processes, each item underwent a modeling procedure.
  Due to incomplete information regarding all these processes associated with the hardware items of the checklist, assumptions were made using different reasonably attributable values. The open-source software *OpenLCA* [37] was used for the modeling. *OpenLCA* provides access to various databases for conducting the LCA of the items. Specifically, the *Environmental Footprints* (EF) database [38] was utilized, as it offers extensive data on the production and disposal of a wide range of products and services. The adoption of the EF database, developed under the guidance of the European Commission, ensures that the embodied impact estimates are traceable to documented and harmonized life-cycle inventory models, consistent with established and widely adopted environmental assessment standards. As outputs, uniform PDFs were obtained for the impacts of the modem, EEG headset, and AR HMD, while a normal PDF was derived for the laptop.[3] Notably, the embodied impact of the cloud server infrastructure was excluded from the analysis, as its realization serves multiple purposes and is not strictly dependent on the specific case study under consideration. Only the operational impact, which depends on the case study, was considered. Finally, with regards to software items, the embodied contribution of MATLAB was not addressed, as MathWorks reports negligible emissions [39].
- Since all the considered items are electronic equipment, the *operational* impacts were estimated based on the energy consumption required by each item to perform its function within the corresponding stage of the AI pipeline.
  The first key element in this estimation was the time required for each stage in both scenarios. A reasonable estimation of the time needed for data download or upload was obtained based on the dataset size and the connection speeds. In particular, the download and upload speeds were obtained from the online report *Speedtest Global Index*, provided by Ookla [40]. In this case study, the connection speeds from Italy were used. The analysis was based on monthly connection speed data over the last 12 months. The mean value of the connection speeds was calculated, and the Type-A evaluation of the standard uncertainty was performed. The results, considering a normal PDF after verification with the χ^2 test, outlined download and upload speeds in the ranges respectively equal to 71.72 ± 2.19 Mbps and 19.89 ± 0.12 Mbps. Based on these values, the time required to download and upload the 100-GB considered dataset was estimated at 99 % coverage level as $2.86 \div 3.36$ h and $11.00 \div 11.35$ h, respectively. Moreover, the time required for dataset acquisition in the case study was estimated to 112 working hours. Instead, the stages of data preparation, model validation, and model finalization required 8, 336, and 24 working hours, respectively.
  The second key element for the energy consumption estimation was the power data of the items. Such data were mainly sourced from the datasheets of the hardware items. In particular, given the absence of further information provided by the manufacturer, a delta PDF was applied for items with a single power value, while a uniform PDF was used for items with a range of power values.[4] Energy-related PDFs were then derived with MCM simulations, starting from the PDFs of power consumptions and execution times, considering a number of extractions equal to 10^6. The only exception was represented by the server, for which the energy-related PDF was obtained starting from a PDF corresponding to the energy consumption of the server per GB, considered as a uniform PDF in the range equal to $99 \div 112$ Wh/GB according to the data download scenarios described in [41]. Therefore, by knowing the total database size, the final energy-related PDF for the server was determined.
  For the sake of clarity, it is specified that the energy-related PDF associated with the modem accounts exclusively for its activation and for the power required to transmit and receive signals, as declared by the manufacturer. Consequently, the impact assessment of the modem is treated as independent from that of the server: although the two devices are functionally connected, the modem's operational impact is modeled solely on its nominal power consumption, whereas the server's impact is estimated based on its computational workload.
  In the next step, all the energy PDFs were converted into carbon-related PDFs using a conversion factor based on monthly sampled historical data over the last 12 months from Italy [42]. Notably, this conversion factor is inherently dependent on geographical and temporal assumptions, as it reflects the electricity generation mix of the considered area. Variations in the geographical context directly affect the resulting carbon-related impacts, since different regions are characterized by different emission intensities per unit of energy. For this reason, the geographical area was explicitly fixed, and Italy was selected as the reference context, consistent with the location where the AI pipeline was executed in the case study. By retrieving from the available data the mean and the standard uncertainty through a Type-A evaluation, it was possible to associate each kWh with a range of emissions equal to 0.29 ± 0.02 kg CO_2,eq, considering a normal PDF after verification with the χ^2 test. Consequently, the carbon-related PDFs were obtained by applying MCM simulations with a number of extractions equal to 10^6. Clearly, the operational emissions related to software were excluded, as they were already accounted for in the hardware impact.

[3] As indicated in Section 2, such PDF choices were motivated by the Authors' prior knowledge of the measurand values, in the absence of official information from the manufacturers. This assumption is consistent with the Bayesian interpretation adopted in GUM Supplement 1, according to which prior information may be legitimately incorporated into the uncertainty evaluation process [28].

[4] It is worth mentioning that power could also be measured using a dedicated measuring instrument. However, this would introduce an additional environmental impact that would need to be quantified, and that lies outside the AI pipeline itself.

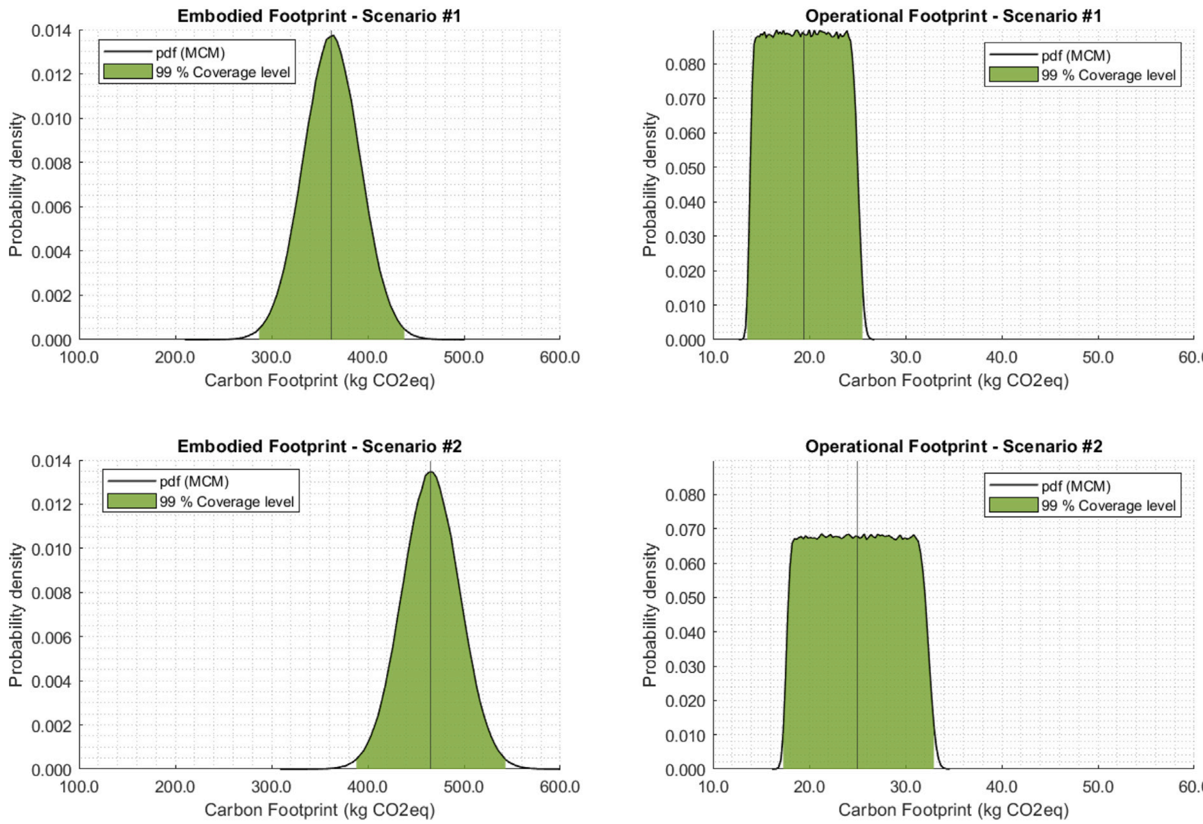

Fig. 2. Output PDFs of the embodied and operational impacts of the AI pipeline for both Scenario #1 and Scenario #2.

Table 2
Embodied and operational impacts of the AI pipeline for the scenario #1 (99 % coverage level).

Item	Embodied impact [kg CO_2,eq]	Operational impact [kg CO_2,eq]
Laptop	246.66 ÷ 393.34	10.77 ÷ 21.98
Modem	32.13 ÷ 52.91	0.04 ÷ 0.05
Server	–	2.83 ÷ 3.37
Software	–	–
Total	287.58 ÷ 437.57	13.70 ÷ 25.42

Table 3
Embodied and operational impacts of the AI pipeline for the scenario #2 (99 % coverage level).

Item	Embodied impact [kg CO_2,eq]	Operational impact [kg CO_2,eq]
Laptop	246.66 ÷ 393.34	14.25 ÷ 29.09
Modem	32.13 ÷ 52.91	0.16 ÷ 0.17
Server	–	2.82 ÷ 3.37
AR HMD	94.03 ÷ 98.98	0.13 ÷ 0.14
EEG headset	5.02 ÷ 7.99	≈0.01
Software	–	–
Total	389.13 ÷ 541.8	17.45 ÷ 32.82

5.3. PDFs propagation and coverage intervals

Once the carbon-related PDFs for each item were obtained, they were aggregated through MCM simulations using 10^6 extractions, separately for embodied and operational contributions, based on the assumption that the overall environmental impact is the sum of the individual impacts. This additive measurement model implicitly assumes statistical independence among the input quantities. While some degree of correlation may exist in practice, *e.g.*, between execution time, energy consumption, and hardware or software configurations, this assumption is considered reasonable in the present context, as the modeled quantities are treated at the level of individual pipeline stages, even when the same computational device is employed across multiple stages. The individual and total impacts are reported in Tables 2 and 3, for scenario #1 and #2, respectively, with results expressed as 99 % coverage intervals of the output PDFs. Additionally, Fig. 2 illustrates the obtained PDFs. Clearly, different (yet justifiable) choices in uncertainty modeling may lead to different coverage intervals, since the adopted PDFs determine how variability and lack of knowledge are represented and, consequently, how uncertainty is propagated to the final results.

As can be seen, the contributions in the second scenario are higher than the first one. This is due to the use of additional equipment for data acquisition, such as the AR HMD and the EEG headset. It is also evident that embodied contributions are significantly higher than operational ones. This can be explained by the fact that embodied contributions relate to the stages of production, packaging, transportation, and disposal, whereas operational contributions only pertain to the single use of the AI pipeline. Moreover, considering that embodied contributions can be lowered by utilizing existing equipment, which is in line with the principles of circular economy, a more in-depth analysis will focus on operational contributions.

To this aim, for a more comprehensive assessment of the environmental impact of AI pipelines, additional insights related to the impact of each stage separately are provided. This allows for identifying the stage with the highest impact and understanding which actions can be taken to mitigate the environmental impact without degrading performance. The results for both scenarios are summarized in Table 4, highlighting that *Model Validation* has produced the highest impact, followed by *Data Collection*.

Table 4
Operational impacts of the AI pipeline stage in both scenarios (99 % coverage level).

AI pipeline stage	Scenario #1 [kg CO_2,eq]	Scenario #2 [kg CO_2,eq]
Data collection	2.97 ÷ 3.59	6.74 ÷ 10.98
Data preprocessing	0.23 ÷ 0.47	0.23 ÷ 0.47
Model validation	9.74 ÷ 19.90	9.74 ÷ 19.90
Model finalization	0.69 ÷ 1.42	0.69 ÷ 1.42

Clearly, strategies to reduce environmental impact can include shortening hyperparameters optimization [43]. However, this approach may result in decreased model performance. Therefore, when incorporating sustainability as a decision-making factor, balancing both performance requirements and environmental impact becomes essential. It becomes crucial to evaluate trade-offs to ensure that sustainability efforts do not compromise the quality or effectiveness of AI models.

In this study, the grid search process was extended to include additional configurations beyond the initial setup of 2160 combinations, which required a computational time of 336 h. Two alternative configurations were evaluated, reducing the computational time to 168 h and 48 h, respectively. Fig. 3 presents a detailed comparison of the environmental impact and related classification accuracy for three grid search configurations, highlighting the relationship between environmental impact and model performance. The results reveal clear trade-offs, which are discussed below.

First, the environmental impact is strongly influenced by the duration of the grid search. As expected, the operational impact increases with longer durations. The 336-h configuration exhibits the highest environmental cost, with emissions ranging from about 10 to 20 kg $CO_{2,eq}$, reflecting the significant computational effort involved. In contrast, the 168-h configuration reduces emissions by approximately half, with values ranging from about 5 to 10 kg $CO_{2,eq}$. The 48-h configuration, as the shortest duration tested, achieves the lowest environmental footprint, with emissions between about 1 and 3 kg $CO_{2,eq}$. In terms of classification accuracy, longer grid search durations generally lead to better performance. The 336-h configuration achieves the highest accuracy, with values ranging from about 80 % to 98 %. The 168-h configuration provides similar accuracy levels, while significantly reducing environmental costs, making it an attractive compromise. In fact, the 48-h configuration, while environmentally favorable, shows a substantial decrease in accuracy, with a range of about 67 % to 91 %. This highlights the potential performance limitations of shorter grid searches, particularly for applications requiring high predictive precision.

Overall, these results emphasize the importance of balancing accuracy and sustainability in the design of AI pipeline. While the 336-h configuration delivers the highest classification performance, its environmental cost may be prohibitive in many scenarios. The 168-h configuration emerges as a practical alternative, offering nearly equivalent accuracy at a substantially lower environmental impact. In contrast, the 48-h configuration demonstrates that while shorter grid searches can minimize environmental costs, they may come at the expense of significant accuracy losses, which could compromise model utility in critical applications.

In conclusion, the proposed method allows for the measurement of the environmental sustainability of an AI pipeline under specified operational conditions. While the assessment is grounded in observed and configuration-specific data, the resulting metrics and associated uncertainties can also inform future implementations, reconfigurations, or repeated executions of the same pipeline, thereby supporting informed and evidence-based decision-making.

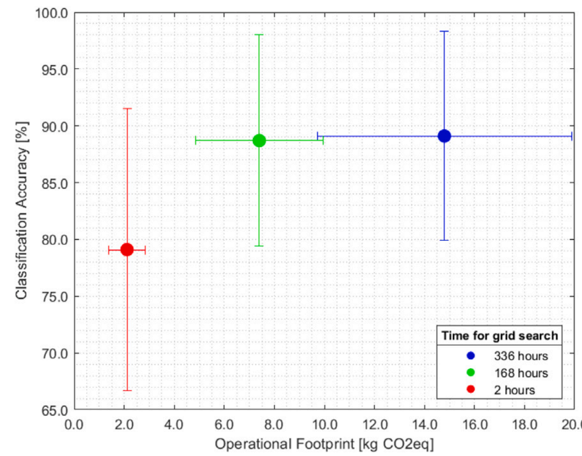

Fig. 3. Graphical representation of the relationship between environmental sustainability (operational footprint) and performance of the AI model. The results are expressed at 99 % coverage level.

6. Conclusions and future work

This paper addresses the measurement of environmental sustainability associated with the use of AI pipelines from a *user-centered* perspective, which means considering all the typical operations performed by end users, thereby filling an existing gap in the current state of the art. A systematic approach, aligning with ISO standards and Supplement 1 of the GUM, was followed: for each stage of the pipeline, a checklist of the items was produced. Therefore, for each item, both embodied and operational impacts were obtained in terms of PDFs through LCA operation. The obtained PDFs were propagated with MCM simulations based on the assumption that the overall impact is the sum of the individual impacts. Finally, a coverage interval was derived, hence providing the measure of the environmental impact of the overall AI pipeline. The considered case study on biosignal processing highlighted the significant role of *Model Validation* stage in influencing environmental sustainability. In light of this, strategies to mitigate the impact were considered to reassess the AI pipeline from a more sustainable standpoint. One strategy employed in this paper was the reduction of the grid search in the validation stage, which demonstrates an effective trade-off between environmental impact and model performance. Future work will address (i) the improvement in the precision of impact assessments, and (ii) the development of practical guidelines for sustainable AI pipelines without compromising performance. This will help integrate sustainable AI practices into real-world applications, aligning with the climate change goals of the United Nations 2030 Agenda.

CRediT authorship contribution statement

Leopoldo Angrisani: Writing – review & editing, Resources, Project administration, Funding acquisition, Conceptualization. **Mauro D'Arco:** Writing – review & editing, Supervision, Methodology, Formal analysis. **Egidio De Benedetto:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Luigi Duraccio:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Immacolata Esposito:** Writing – original draft, Visualization, Validation, Software, Investigation, Data curation. **Annarita Tedesco:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was financially supported by the Italian Ministry of University and Research (MUR) through the projects: "Made in Italy Circolare e Sostenibile", PNRR PE11 (CUP E63C22002130007) and PNRR ex D.M. 118/2023 – "Area Transizioni digitali e ambientale" (CUP E66E23001000002) .

Data availability

Data will be made available on request.

References

[1] R. Heinberg, D. Lerch, What is sustainability, Post Carbon Read. 11 (2010) 19.
[2] V. Ramanathan, Y. Feng, Air pollution, greenhouse gases and climate change: Global and regional perspectives, Atmos. Environ. 43 (1) (2009) 37–50.
[3] S. Popli, R.K. Jha, S. Jain, A comprehensive survey on green ICT with 5G-NB-IoT: Towards sustainable planet, Comput. Netw. 199 (2021) 108433.
[4] T. Jose, J.A. Mayan, LoRaWAN and artificial intelligence integrated smart acoustic sensor network for bird species identification and deterrence system for farm protection, Measurement (2025) 118452.
[5] M. Wieczorowski, D. Kucharski, P. Sniatala, P. Pawlus, G. Krolczyk, B. Gapinski, A novel approach to using artificial intelligence in coordinate metrology including nano scale, Measurement 217 (2023) 113051.
[6] A.I. Khan, Y.B. Abushark, F. Alsolami, A. Almalawi, M.M. Alam, P. Kshirsagar, R.A. Khan, Prediction of breast cancer based on computer vision and artificial intelligence techniques, Measurement 218 (2023) 113230.
[7] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S.D. Langhans, M. Tegmark, F. Fuso Nerini, The role of artificial intelligence in achieving the sustainable development goals, Nat. Commun. 11 (1) (2020) 1–10.
[8] G. Bonire, A. Gbenga-Ilori, Towards artificial intelligence-based reduction of greenhouse gas emissions in the telecommunications industry, Sci. Afr. 12 (2021) e00823, http://dx.doi.org/10.1016/j.sciaf.2021.e00823, URL: https://www.sciencedirect.com/science/article/pii/S2468227621001277.
[9] M. Dong, G. Wang, X. Han, Impacts of artificial intelligence on carbon emissions in China: in terms of artificial intelligence type and regional differences, Sustain. Cities Soc. 113 (2024) 105682, http://dx.doi.org/10.1016/j.scs.2024.105682, URL: https://www.sciencedirect.com/science/article/pii/S2210670724005079.
[10] C.-J. Wu, R. Raghavendra, U. Gupta, B. Acun, N. Ardalani, K. Maeng, G. Chang, F. Aga, J. Huang, C. Bai, et al., Sustainable ai: Environmental implications, challenges and opportunities, Proc. Mach. Learn. Syst. 4 (2022) 795–813.
[11] D. Patterson, J. Gonzalez, Q. Le, C. Liang, L.-M. Munguia, D. Rothchild, D. So, M. Texier, J. Dean, Carbon emissions and large neural network training, 2021, arXiv preprint arXiv:2104.10350.
[12] Z. Cao, X. Zhou, H. Hu, Z. Wang, Y. Wen, Toward a systematic survey for carbon neutral data centers, IEEE Commun. Surv. Tutor. 24 (2) (2022) 895–936, http://dx.doi.org/10.1109/COMST.2022.3161275.
[13] N. Bannour, S. Ghannay, A. Névéol, A.-L. Ligozat, Evaluating the carbon footprint of NLP methods: a survey and analysis of existing tools, in: Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing, 2021, pp. 11–21.
[14] A. Lacoste, A. Luccioni, V. Schmidt, T. Dandres, Quantifying the carbon emissions of machine learning, 2019, arXiv preprint arXiv:1910.09700.
[15] P. Henderson, J. Hu, J. Romoff, E. Brunskill, D. Jurafsky, J. Pineau, Towards the systematic reporting of the energy and carbon footprints of machine learning, J. Mach. Learn. Res. 21 (248) (2020) 1–43.
[16] L. Lannelongue, J. Grealey, M. Inouye, Green algorithms: quantifying the carbon footprint of computation, Adv. Sci. 8 (12) (2021) 2100707.
[17] L.F.W. Anthony, B. Kanding, R. Selvan, Carbontracker: Tracking and predicting the carbon footprint of training deep learning models, 2020, arXiv preprint arXiv:2007.03051.
[18] E. Masciari, E.V. Napolitano, An Effective Measure for Evaluating the Environmental Impact of Ai Tasks. Available At SSRN 4962273.
[19] A.A. Torres-García, C.A.R. Garcia, L. Villasenor-Pineda, O. Mendoza-Montoya, Biosignal Processing and Classification Using Computational Learning and Intelligence: Principles, Algorithms, and Applications, Academic Press, 2021.
[20] United Nations, Sustainable development goals, 2025, https://sdgs.un.org/goals. (Accessed 23 December 2025).
[21] The global goals - goal 13: Climate action, 2025, https://www.globalgoals.org/goals/13-climate-action/. (Accessed 23 December 2025).
[22] ISO 14040, environmental management - life cycle assessment — Principles and framework, 2006, https://www.iso.org/obp/ui/#iso:std:iso:14040:ed-2:v1:en. (Accessed 23 December 2025).
[23] ISO 14064, greenhouse gases - part 1: Specification with guidance at the organization level for quantification and reporting of greenhouse gas emissions and removals, 2018, https://www.iso.org/obp/ui/#iso:std:iso:14064:-1:ed-2:v1:en. (Accessed 23 December 2025).
[24] ITU L.1420: Methodology for energy consumption and greenhouse gas emissions impact assessment of information and communication technologies in organizations, 2025, https://www.itu.int/rec/T-REC-L.1420. (Accessed 23 December 2025).
[25] K. Demestichas, E. Daskalakis, Information and communication technology solutions for the circular economy, Sustainability 12 (18) (2020) 7272.
[26] L. Angrisani, M. D'Arco, E. De Benedetto, L. Duraccio, M. Imbò, A. Tedesco, Sustainability-aware measurement design: How to select the right measuring system, Measurement (2024) 115493, http://dx.doi.org/10.1016/j.measurement.2024.115493.
[27] L. Angrisani, G. Bonini, M. D'Arco, E. De Benedetto, L. Duraccio, I. Esposito, A. Tedesco, Sustainability balance for ICT infrastructures: A time-scale analysis of the dynamic fourth lane on the A4 Italian motorway, IEEE Access (2025) 1, http://dx.doi.org/10.1109/ACCESS.2025.3611801.
[28] JCGM Publications: Guides in Metrology - BIPM JCGM 101:2008, Evaluation of measurement data – supplement 1 to the guide to the expression of uncertainty in measurement – propagation of distributions using a Monte Carlo method, 2025, http://www.bipm.org/en/publications/guides/gum.html. (Accessed 23 December 2025).
[29] X. Chen, Z. Chen, S. Gao, X. Gao, A high-itr ssvep-based bci speller, Brain-Comput. Interfaces 1 (3–4) (2014) 181–191.
[30] L. Angrisani, P. Arpaia, E. De Benedetto, L. Duraccio, F.L. Regio, A. Tedesco, Wearable brain–computer interfaces based on steady-state visually evoked potentials and augmented reality: A review, IEEE Sens. J. 23 (15) (2023) 16501–16514, http://dx.doi.org/10.1109/JSEN.2023.3287983.
[31] X. Chen, Y. Wang, S. Gao, T.-P. Jung, X. Gao, Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface, J. Neural Eng. 12 (4) (2015) 046008.
[32] Y. Wang, X. Chen, X. Gao, S. Gao, A benchmark dataset for SSVEP-based brain–computer interfaces, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (10) (2016) 1746–1752.
[33] L. Angrisani, M. D'Arco, E. De Benedetto, L. Duraccio, F. Lo Regio, A. Tedesco, A novel measurement method for performance assessment of hands-free, XR-based human-machine interfaces, IEEE Sens. J. (2024) http://dx.doi.org/10.1109/JSEN.2024.3444472.
[34] S. Raschka, Model evaluation, model selection, and algorithm selection in machine learning, 2018, arXiv preprint arXiv:1811.12808.
[35] Unicorn hybrid black 8 channel wearable eeg headset, 2025, https://www.gtec.at/product/unicorn-hybrid-black. (Accessed 23 December 2025).
[36] HoloLens 2 - for precise, efficient hands-free work., 2025, https://www.microsoft.com/en-us/hololens. (Accessed 23 December 2025).
[37] GreenDelta, Openlca website, 2025, https://www.openlca.org/. (Accessed 23 December 2025).
[38] OpenLCA Nexus, OpenLCA Nexus: Your source for LCA and sustainability data., 2025, https://nexus.openlca.org/database. (Accessed 23 December 2025).
[39] MathWorks, Mathworks 2022: Climate action report, 2025, https://www.mathworks.com/content/dam/mathworks/brochure/mathworks-2022-climate-action-report.pdf. (Accessed 23 December 2025).
[40] Ookla, Speedtest global index – internet speed around the world, 2025, https://www.speedtest.net/global-index/italy. (Accessed 23 December 2025).
[41] G. Guennebaud, A. Bugeau, Energy consumption of data transfer: Intensity indicators versus absolute estimates, J. Ind. Ecol. 28 (4) (2024) 996–1008.
[42] Nowtricity, Real time electricity production emissions by country, 2025, https://www.nowtricity.com/. (Accessed 23 December 2025).
[43] IEEE Spectrum - April 2024, We need to decarbonize software - the green software movement is tackling the hidden environmental impact of today's code, 2025, https://spectrum.ieee.org/magazine/2024/april. (Accessed 23 December 2025).