



**HAL**  
open science

## **Multi-criteria and multi-stage environmental study of Pl@ntnet service for the year 2024**

Adrien Berthelot, Tiago da Silva Barros, Laurent Lefèvre, Anne-Laure Ligozat,  
Emeline Pegon

### ► **To cite this version:**

Adrien Berthelot, Tiago da Silva Barros, Laurent Lefèvre, Anne-Laure Ligozat, Emeline Pegon. Multi-criteria and multi-stage environmental study of Pl@ntnet service for the year 2024. Inria Lyon. 2026. ⟨hal-05448455v2⟩

**HAL Id: hal-05448455**

**<https://inria.hal.science/hal-05448455v2>**

Submitted on 4 Mar 2026

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Multi-criteria and multi-stage environmental study of Pl@ntnet service for the year 2024

Adrien Berthelot\*  
Tiago Da Silva Barros†  
Laurent Lefèvre‡  
Anne-Laure Ligozat§  
Emeline Pegon¶

March 4, 2026

## Acknowledgments

We thank all Pl@ntNet team members (Alexis Joly, Hugo Gresse, Jean-Christophe Lombardo, Pierre Bonnet) for providing us relevant information on architecture and usage of the Pl@ntNet service. We would also like to thank Gaël Guennebaud for his constructive feedback on the report. We also thank Université Côte d’Azur for funding this project. This work is also funded by the CANDISE research project funded by the UKRI-EPSC through award EP/X038858/1, and the Alt-Impact program of ADEME, CNRS and Inria. This project benefited from GENCI’s IT resources at IDRIS thanks to grant 2024- A0171011389 on the Jean Zay supercomputer, V100, A100 and H100 partitions.

---

\*University of Bristol, UK - adrien.berthelot@bristol.ac.uk

†Université Côte d’Azur, Sophia-Antipolis, France - tiago.da-silva-barros@univ-cotedazur.fr

‡Avalon Team, LIP, UCBL, Inria, ENS de Lyon, CNRS, LIP - Lyon, France - laurent.lefevre@ens-lyon.fr

§Université Paris-Saclay, CNRS, ENSIIE, LISN, Orsay, France - anne-laure.ligozat@lisn.fr

¶IRIT, CNRS, Avalon Team, LIP, ENS de Lyon - emeline.pegon@ens-lyon.fr

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>The Pl@ntNet Service</b>	<b>4</b>
2.1	Identification Request . . . . .	5
2.2	Pl@ntNet Infrastructure . . . . .	6
2.2.1	Training . . . . .	6
2.2.2	Inference . . . . .	6
2.2.3	Webhosting . . . . .	7
<b>3</b>	<b>Life Cycle Analysis</b>	<b>8</b>
3.1	The stages of Life Cycle Assessment . . . . .	8
3.2	The Impacts assessment . . . . .	9
3.3	Results interpretation and recommendations . . . . .	10
<b>4</b>	<b>Methodology</b>	<b>10</b>
4.1	Study goal, scope, and perimeter considered . . . . .	10
4.2	LCA modeling . . . . .	13
4.2.1	End User Terminals . . . . .	13
4.2.2	Network . . . . .	14
4.2.3	Training . . . . .	14
4.2.4	Inference . . . . .	16
4.2.5	Webhosting . . . . .	16
4.3	Parameters and sources . . . . .	18
<b>5</b>	<b>Results</b>	<b>20</b>
5.1	Global environmental impacts . . . . .	20
5.2	Detailed assessment of Pl@ntNet impacts . . . . .	21
5.2.1	<i>End Users</i> . . . . .	21
5.2.2	Networks . . . . .	25
5.2.3	Training . . . . .	26
5.2.4	Inference . . . . .	27
5.2.5	Webhosting . . . . .	29
<b>6</b>	<b>Limits and discussion</b>	<b>30</b>
6.1	Limits due to data quality . . . . .	30
6.2	Limits due to methodological choices . . . . .	31
6.3	How do the results address environmental concerns about Pl@ntNet? . . . . .	32
<b>7</b>	<b>Conclusion</b>	<b>33</b>

# 1 Introduction

In recent years, the mobile application market has been experiencing an exponential growth, leveraging both the number of users and revenue generated, and this trend shows no signs of slowing. In 2024, according to the annual State of Mobile 2025 report [28] produced by SensorTower, the mobile application market generated approximately 150 billion dollars in global consumer spending on in-app purchases and subscriptions (with an increasing rate of +13% per year), achieving an annual 136 billion of annual number of downloads. These impressive results underscore the importance of mobile apps in today’s digital economy. For businesses and developers, this represents a considerable opportunity, but also significant challenges.

The integration of technologies such as Augmented Reality (AR) and Artificial Intelligence (AI) into mobile application is a notable trend. These technologies offer enriched and personalized user experiences, such as augmented reality filters in social networks or AI-powered virtual assistants, experiencing a significant growth in the last years. The AI adoption is expected to grow exponentially in the coming years. Consumers have spent nearly 1.1 billion dollars (+200% year-on-year) on applications such as *ChatGPT*, *Google Gemini*, and *Doubao* and the time spent on *CharacterAI* or *ChatGPT* has reached 7.7 billion hours. [28]

In view of these figures and emerging trends, sustainable development and environmental impact have become key issues for stakeholders for guiding eco-design strategies for digital services. Artificial Intelligence is responsible for a wide range of environmental impacts. Some important environmental impacts of AI usage include *water consumption*, *human toxicity*, and other abiotic natural resources, such as *aluminum*, *cobalt*, *copper*, *glass*, *gold*, *tin*, *lithium*, *zinc*, and *plastic*. [17, 6].

The multiple impacts associated to computing industry may occur at different phases of ML-based services development. In order to evaluate them, researchers have tapped into the Life Cycle Assessment (LCA) methodology [2]. The resulting research field is to investigate societal and environmental impacts of computing systems and infrastructures at different stages of product life cycle.

In literature, several works [17, 25, 26, 9, 8] were proposed in order to investigate digital services life cycle impacts. Researchers focus their analysis on two types of impacts: (i) *Embodied impacts*, corresponding to the impacts associated to the hardware *production*, *transport* and *end of life*; and (ii) *Operational impacts*, corresponding impacts associated to the computing systems usage (idle and dynamic), mostly related to the energy consumption for powering the computing devices.

In this study, we focus our investigation on Pl@ntNet [23], a citizen science platform, which relies on Artificial Intelligence (AI) models to identify plant species. Pl@ntNet provides a large-scale infrastructure supporting millions of users in over 200 countries. At this stage of deployment, and with years of experience developing the platform, Pl@ntNet is committed to understanding the environmental impacts of its identification service and contributing to the search for reduction opportunities.

Our investigations assess the associated environmental impacts of Pl@ntNet for the year 2024. We based our approach on multi-criteria LCA, considering multiple impact type and the different life-cycle phases.

This study is based on an attributional LCA. Only direct impacts have been taken into consideration. Indirect impacts, both positive and negative (such as rebound effects, substitution, structural changes), are not explored and the consequential effects of the impacts of the use and activities of the service were not the subject of the work. (see Section 3.3 for more details).

We divide the Pl@ntNet service into multiple sections: End Users Terminals, Networks,

Training, Inference and Webhosting. For each section, we assess the *Abiotic Depletion Potential* (ADP) [32], the *Global Warming Potential* (GWP) [31] and the *Primary Energy* (PE) [16] for both *embodied* and *operational* impacts.

Our findings point that the Pl@ntNet application was responsible for emitting around **78 tonnes of CO<sub>2</sub> equivalent in 2024**. The end users terminals are responsible for a relevant share of the above emissions, an equivalent of 62% of the global impacts. Our contributions may be summarized as follows:

- We conduct an LCA on the Pl@ntNet application for the 2024 year. We base our analysis and methodology on [9] assessing multiple categories of environmental impacts.
- We investigate multiple sections concerning Pl@ntNet service: *end users*, *networks*, *training*, *inference* and *webhosting*. For each section, we use the associated equipments statistics for estimating the environmental impacts.
- We derive the embodied and operational impacts, thus comparing the multiple phases of service life cycle.
- We quantify the environmental impacts, identifying the application parts that mostly contribute to the global impacts and pointing optimization opportunities.

This study is structured in the following way. In section 2 we describe the Pl@ntNet service and the identification request path, from the moment when the user takes the picture until the results display. In Section 3, we discuss the Life Cycle Analysis (LCA) methodology highlighting its key components. In Section 4, we present our methodology. We describe how we estimate the embodied and operational results of each section of Pl@ntNet service. In section 5, we provide our results of the environments impacts estimation and we compare the different service sections. In section 6, we discuss the results and explain the limits of this study. Finally, in Section 7, we draw the main conclusions and we point the main future perspectives.

## 2 The Pl@ntNet Service

The aim of the Pl@ntNet project is to enable both novices and experienced naturalists to identify and list the plants they encounter in the field, and to share their observations.

Launched in 2009, Pl@ntNet is a scientific project bringing together the French National Institute for Research in Computer Science and Control (INRIA), the French Institute for Development Research (IRD), the Centre for International Cooperation in Agronomic Research for Development (CIRAD), the French National Institute for Research in Agriculture, Food and the Environment (INRAE) and the Tela Botanica network.

The initial aim of the project was to inventory biodiversity for scientific research using image identification. In 2011, the project was transformed into participatory science with the launch of a collaborative web platform, followed by a smartphone application in 2013 (Figure 1).

Specialists, amateurs and the curious connect in a collaborative network. The most qualified users contribute to the most expert knowledge and validate observations. In April 2025, the application was able to identify 75,200 species, had 7.8 million registered users and offered 32 million observable images. In addition to plant identification, which remains the main use of the Pl@ntNet service, the platform offers a geolocation system to map the species listed, a community network and specific services dedicated to professional structures (private, public and scientific research).



Figure 1: Pl@ntNet service, figure extracted from the website

Since 2020, Pl@ntNet data has been integrated into the Global Biodiversity Information Facility (GBIF), helping to mobilize and use data on biodiversity trends, as well as promoting cooperation and knowledge sharing.

The platform is now one of the largest biodiversity observatories in the world, with several million contributors in over 200 countries. Table 2 describes the top-10 countries with most usage and their respective ratio of the total sessions using Pl@ntNet. The application is available in 52 languages thanks to the work of the developer and translator communities.

Several thousand scientific publications have made use of Pl@ntNet data, and the project is regularly included in other global research projects and programs.

## 2.1 Identification Request

The object of our study will be linked to the main service of Pl@ntNet: the recognition of a plant or species from a photograph transmitted via the mobile application. When the user requests for a plant identification, the request goes through different steps for obtaining the results. The request path is described in Fig. 2. First, the users opens the application and takes a picture of the plant, which will serve as input for identifying the species. The picture is compressed and transmitted to the CIRAD datacenter where it will be processed. The request is forwarded by the API and stored by the Pl@ntnet database system. Then, the identification is transferred to the identification servers, where the ML model performs an image classification task in a GPU card outputting the plant species. Next a similarity search is performed in order to get the most likely species and finally, the API transmits the results back to the user. The results consists of a set containing the most likely species of the plant identified, together with some small images of each of the plant possibilities.

Each image in the database is represented by a number of visual attributes that distinguish one species from another. To select the most discriminating attributes, the AI based techniques Visual Transformers [34] are utilized. They can characterize both fairly general traits such as flower color or the general shape of the leaf, as well as subtle details such as the shape of the flower petal tips or the presence of slight hairiness on the leaf margins.

All the above characteristics are used by the ML models for distinguishing different species

when identifying a plant species based on the picture. The performance of the ML models used by Pl@ntNet are strongly correlated with the model training, which is done in Jean Zay supercomputer using Pl@ntNet dataset.

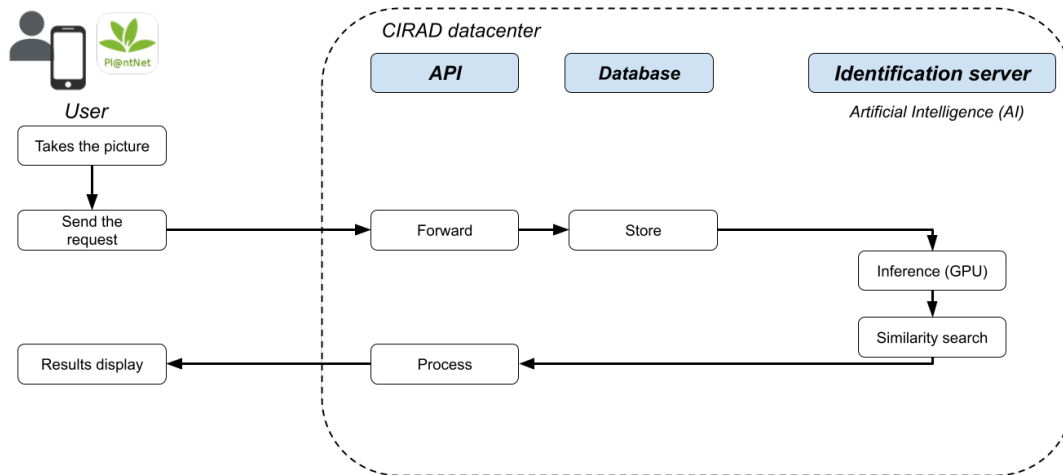


Figure 2: Plant identification path using Pl@ntNet service.

## 2.2 Pl@ntNet Infrastructure

Here, we provide the infrastructure used for hosting the Pl@ntnet service. The infrastructure can be divided in 3 sections, each relying on specific equipment and location: training, inference and webhosting. The first section, training, concerns only machine learning process done to deliver an AI model using large dataset. The second section, inference, concerns the process of identification described in the last section, responding to requests using the previously trained model. Lastly, the webhosting part concerns all the other servers and equipment managed by the Pl@ntnet team to the benefit of the Pl@ntnet service.

### 2.2.1 Training

During training, Pl@ntNet uses the Jean Zay supercomputer, using the graphical cards Nvidia Tesla V100 and Nvidia A100, for performing training. For the node containing V100 graphical card, we considered the v100-32g node as it is the most common node in Jean Zay. The node specification provides: 2 processors Intel Cascade Lake 6248, 192 GB RAM, 4 GPU Nvidia Tesla V100 SXM2 32 GB.

For the node containing A100 graphical card, we considered the p5 partition <sup>1</sup>, which contains: 2 processors AMD Milan EPYC 7543, 512 GB RAM, 8 GPU Nvidia A100 SXM4 80 GB.

### 2.2.2 Inference

The inference requests using Pl@ntnet service are performed in the CIRAD datacenter. A total of 8 fully dedicated machines are used for inference, named as following: Capcir, Orxois, Chamois, Devois, Rajol, Huchet, Senart and Sourduin. The servers present the following specifications:

<sup>1</sup><http://www.idris.fr/jean-zay/cpu/jean-zay-cpu-hw.html>

- The servers CAPCIR and ORXOIS present 2 CPU Intel Gold 5118, a GPU Nvidia Quadro M4000, 62 GB of RAM and a HDD with 2TB of storage capacity.
- The servers RAJOL and HUCHET present 2 CPU Intel Gold 5118, a GPU Nvidia RTX 8000, 92 GB of RAM and an SSD with 4TB of storage capacity.
- The servers CHAMOIS, DEVOIS, SENART and SOURDUN present 2 CPU Intel Gold 5118, a GPU Nvidia A40-48GB, 128 GB of RAM and an SSD with 4TB of storage capacity.

### 2.2.3 Webhosting

The Pl@ntNet webhosting section is deployed in the CIRAD datacenter. The infrastructure may be divided in two parts. The API service is deployed through Virtual Machines (VMs). The physical infrastructure used for this section uses 12 blades containing the following specifications: 2 CPU Intel Gold 5118 and 512 GB of RAM.

Since CIRAD cluster does not offer a power meter providing measures about the power consumption, we estimated the power by considering a similar server in Grid5000 platform [5]. We selected the node `roazhon-9`, which contains 4 instances with the following specifications: 2 Intel Xeon Gold 5118 and 512 GB RAM. We extracted the average power consumption in `roazhon-9` node for idle and charge loads, as described in Table 1.

Node	Idle Power (W)	Charge Power (W)
roazhon-9-1	50	332
roazhon-9-2	56	256
roazhon-9-3	53	253
roazhon-9-4	53	303
<b>Average</b>	<b>53.75</b>	<b>291.25</b>

Table 1: Power measurements in idle and charge mode in Roazhon-9 node.

For the database service, two machines are used by Pl@ntNet in this section: (i) BRIVADOIS, machine provided with 4 CPU Intel Xeon Gold 5118, 256 GB of RAM memory and 7TB of SSD Storage; and (ii) GRAY, provided with 4 CPU Intel Xeon Gold 5118, 256 GB of RAM memory and 9TB of SSD Storage. Furthermore, the storage section utilizes the storage equipment `mesoPl@ntnet`, provided with a SSD with 270 TB of capacity.

Country	Ratio
France	25.1%
Germany	12.0%
USA	11.1%
Italy	6.5 %
United Kingdom	5.0%
Spain	4.1 %
Netherlands	3.8 %
Switzerland	2.5%
Brazil	2.4%
Canada	2.0%

Table 2: Top-10 countries using Pl@ntNet service in world and their respective ratio of total sessions.

### 3 Life Cycle Analysis

The study we are conducting is based on a Life Cycle Analysis methodology adapted to digital services [7].

Life cycle assessment (LCA) is used to evaluate the potential environmental impact of products, services or organisations. There are other methods for assessing environmental impacts, such as carbon footprint and impact studies. But LCA has specific features that make it a unique systemic approach. Used regularly since the end of the 1990s and in different industrial sectors [18] and standardised in the ISO norms 14040 [20] and 14044 [21] series, this method proposes to establish the ecological footprint of a product or service based on several key concepts [10]:

- **Multicriteria:** Several environmental indicators are systematically considered, including global warming potential, depletion of abiotic resources, photochemical ozone creation, water, air and soil pollution, human ecotoxicity and biodiversity. The list of indicators is not fixed and depends on the sector of activity.
- **Life cycle:** in order to integrate the impacts generated during all stages of the life cycle of products or services, from the extraction of natural resources to the production of waste, including energy consumption during the use phase, transport phase, etc.
- **Quantitative:** each indicator is quantified so that all the externalities of a product or service can be put on the same scale and objective decisions can be taken.
- **Functional:** the object of study is defined by the function it fulfils, so that different technical solutions can be compared.

Although LCA was initially applied more to the field of products, its scope has been extended in recent years to the study of global digital services. Firstly, through adaptation of methodological standard by the ETSI [3] and the ITU [22], and also through production of environmental impact data of ICT products and infrastructures done by the ADEME, the NegaOctet consortium or by the Boavizta association [29]. In this study, we are assessing Pl@ntNet not as a product but as a service. We will be following recommendation from the ITU L1410 [22] and from recent academic work on LCA of digital services [7]. As a digital service, we will study Pl@ntNet through its use of 3 types of system: end-user device, network infrastructure and datacenter, adopting an end-to-end perspective.

#### 3.1 The stages of Life Cycle Assessment

LCA is based on 4 stages, summarized in Figure 3:

**1 - Objectives definition** This first step consists of defining and presenting the framework of the study. It allows determining the objectives of the evaluation, the final goal, presenting the recipients of the study and establishing the boundaries of the system studied and defining the functional unit used as the reference unit of the study.

**2 - Life Cycle Inventory** This consists of collecting data and using calculation procedures to quantify the inputs and outputs of the system under study. The life cycle inventory (LCI) identifies the basic flows inherent in the system. The data collected may be measured on site, calculated or estimated. These types of data collection make it possible to evaluate and establish a quality value for the LCA data. The analysis of the LCI is sometimes considered to be the heart

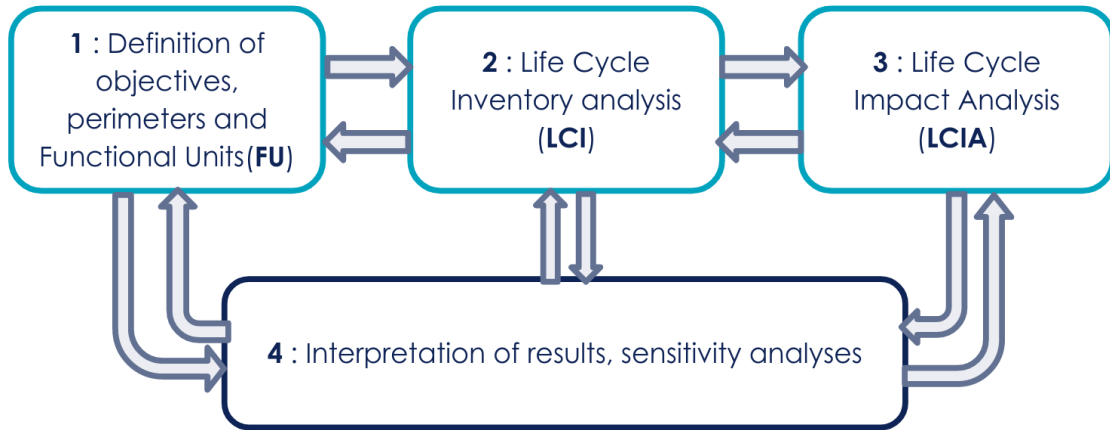


Figure 3: LCA main phases

of the LCA study. This stage is central to listing all the equipment, flows and infrastructures used, as well as the intensity of their use. In the case of LCA of digital services, the modelling of the digital service is conventionally divided into 3 main categories as indicated in methodologies such as the Digital Services Product Category Rules (PCR) [13].

1. *User Terminals*: This category includes the terminals used by the end users.
2. *Networks*: This category includes the network infrastructures for data exchanges between end-user terminals and data centres. Networks may include the fixed network, mobile networks, backbone networks, satellite networks, submarine cable networks, etc.
3. *Datacenter*: This category includes equipment and infrastructure for hosting, processing and calculating data (servers, disks, network equipment, etc.)

**3 - Impacts assessment** The goal here is to convert the collected data into potential environmental impacts. These conversions can be done using professional software. Specifically, the collected data flows are multiplied by what are called characterization factors to obtain an environmental impact. Impact categories are selected based on the LCA objectives or as indicated by established standards; ISO 14040, ISO 14044, ITU L-1410 or PCRs.

**4 - Results Interpretation** The results obtained must be interpreted in accordance with the defined objectives. This analysis at the level of the life cycle phases, the different flows, the components or the processes used allows the salient points of the environmental impact to emerge and opens up discussions on improvement solutions.

### 3.2 The Impacts assessment

The impact characterisation phase aims to assess the significance of potential environmental impacts based on the inventory results. This process involves selecting impact categories and associating inventory data with impact categories (e.g. climate change). To date, many studies are limited to assessing global warming potential (GWP) and electricity consumption. To simplify matters, this consumption is often presented as the energy footprint, even though the latter, in the form of cumulative primary energy demand, covers a much wider area. Other lesser-known

categories such as the depletion of metals and minerals and the antimony equivalent footprint (ADP), the depletion of water resources (Wu), Land Use (Lu) and ionising radiation (Ir) can be significant for the digital sector. The PEF standard recognises more than 16 categories of environmental impact.

### 3.3 Results interpretation and recommendations

Interpretation is the final phase of the LCA, and includes the results of the inventory and assessment, which are illustrated, summarised and discussed. This part is used by the recipients of the study as a basis for conclusions, recommendations and decision-making in accordance with the objective and the established scope.

Considering that part of the data collection may be based on working hypotheses or secondary data from the literature, which may be uncertain. In order to qualify the order of magnitude of the variation in the results, a sensitivity and uncertainty analysis must be carried out.

## 4 Methodology

In this section, we provide the methodology used for evaluating the environmental impacts related to Pl@ntNet service. We first discuss the scope and the metrics assessed by the current study. Then, we provide the model used to estimate the environmental impacts associated with Pl@ntNet service. All notation used for our analysis is described in Table 3.

### 4.1 Study goal, scope, and perimeter considered

This study was conducted on the initiative of the Pl@ntNet team to explore the environmental footprint of their service. The aim is to assess the environmental impacts, especially to identify where these impacts occur within the service. Particular attention must be paid to the machine learning, or *training*, phase, as it is a crucial component of the service.

In this work, we estimate the environmental impacts associated with the Pl@ntNet service for identifying plant species by performing an attributional LCA. Only direct impacts will be taken into consideration. Indirect impacts, both positive and negative (such as rebound effects, substitution, structural changes), are not explored and the consequential effects of the impacts of the use and activities of the service were not the subject of the work.

We define the Functional Unit (FU) of the study as follows: *Provision of the Pl@ntNet plant identification service over a one-year period (2024), on a global scale.*

The temporal scope for inference activities (i.e., actual plant identification requests) is limited to the year 2024. For the *training*, however, we explore three distinct scenarios to account for different temporal scopes, reflecting the cumulative nature of model development:

- *Final Models* Scenario: This scenario accounts for the environmental impacts associated with training the final versions of the models that are deployed to users. Based on information from Pl@ntNet, we assume that a new model is trained approximately every two months, each requiring approximately 40 hours of computation on 64 nodes, with each node equipped with four NVIDIA V100 GPUs.
- *Standard* Scenario: Here, we consider all model training activities that occurred during the year 2024. This reflects the training footprint within the same reference period as the inference phase.

Notation	Description
$C$	Set of countries
$N$	Number of identification requests in 2024
$U$	Number of users using Pl@ntNet in 2024
$alloc_e$	Allocation factor of equipment $e$ for Pl@ntnet
$I_{capex}^e$	Impact of manufacturing one unit of equipment $e$
$k_e$	Ratio of inference requests performed on equipment $e$
$k_c$	Ratio of inference requests originated in country $c$
$t_e$	Inference time on equipment $e$ , <b>in seconds</b>
$P_e$	Power of equipment $e$ , <b>in KW</b>
$C_e^c$	Energy consumption of equipment $e$ in country $c$ , <b>in kWh</b>
$L_e$	Lifespan of equipment $e$ , <b>in years</b>
$u_e$	Usage ratio of equipment $e$
$H_e$	Training time of Pl@ntnet service in equipment $e$ , <b>in hours</b>
$nGPU_e$	Number of GPUs in equipment $e$
$\rho$	Multiplicative factor used
$d$	Duration of a request using Pl@ntNet, <b>in minutes</b>
$timePerDay_e$	Total time of equipment $e$ usage in a day, <b>in minutes</b>
$energyPerYear_e$	Total energy consumption in equipment $e$ in a year, <b>in kWh</b>
$T$	Types of networks used for transmitting data
$k_t$	Ratio of type $t$ used given all types
$S$	Total data exchanged between the user and the server, <b>in Bytes</b>
$I^t$	Impact of network type $t$ for GB transmitted
$nCPU$	Number of threads on machine
$ram$	RAM memory capacity, <b>in GB</b>
$sto$	Storage capacity, <b>in GB</b>
$\eta$	Number of blades managed by the hypervisor in physical machine
$a_{vm}^{cpu,ram,sto,rest}$	Share of resources (CPU, RAM, storage or rest) allocated to $vm$
$P_e^{idle,charge}$	Power of equipment $e$ in idle (or charge) mode, <b>in W</b>
$s_{vm,e}^{idle,charge}$	Share of time that $vm$ (or equipment $e$ ) is in idle (or charge) mode
$EGM_c$	Electricity grid Mix (EGM) of country $c$ in kgO <sub>2</sub> eq., kgSbeq. or MJ

Table 3: Notation used in the environmental study

- *Legacy Scenario*: This scenario takes into account all cumulative training activities from 2020 to 2024. This is based on the understanding that the current model is the product of iterative improvements over multiple years. By evaluating the legacy training efforts, we aim to capture the long-term environmental cost embedded in the service’s current capabilities.

**LCA Phases** In our analysis, we propose to investigate the *embodied* and *operational* impacts associated with Pl@ntNet service. Fig. 4 illustrates the structure of our Life Cycle Analysis (LCA). The impact evaluation may be divided into five main sections: *end users*, *networks*, *training*, *inference* and *webhosting*.

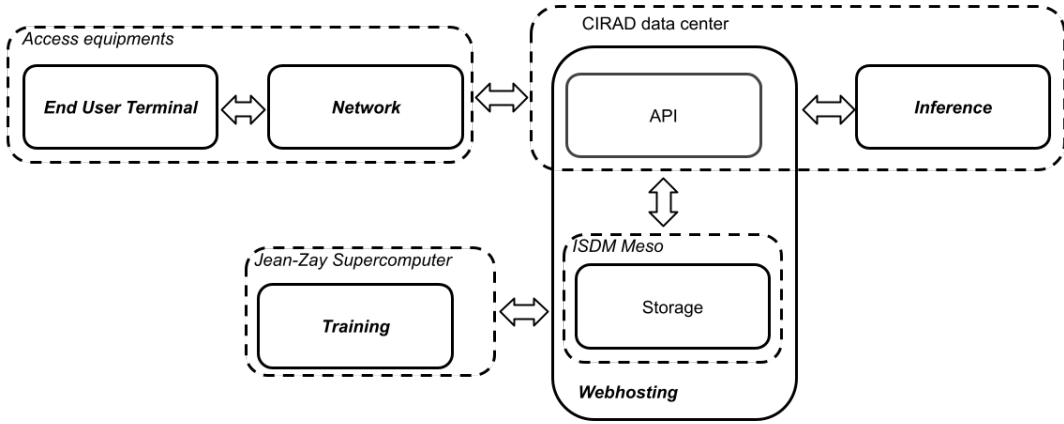


Figure 4: Pl@ntNet service architecture. We divide the service into five sectors: *End Users*, *Network*, *Training*, *Inference*, and *Webhosting* (API and storage).

**Criteria evaluated** Our investigation proposes a multi-phase and multi-criteria approach. We evaluate three metrics indicating the impacts:

- *Abiotic Depletion Potential (ADP)*, assessing the contribution to the depletion of available mineral and metal resources caused by their extraction. This indicator is estimated in kilograms of extracted antimony equivalent (kgSbeq).
- *Global Warming Potential (GWP)*, assessing the contribution to global warming. This indicator is estimated in kilograms of carbon dioxide equivalent emissions (kgCO<sub>2</sub>eq).
- *Primary energy use (PE)*, assessing the cumulative demand for primary energy. This is not strictly speaking an environmental impact category, but it is a relevant indicator of resource demand. It is expressed in MegaJoules (MJ).

In line with the recommendations of the RCP Digital Services, the ITU-T L.1410<sup>2</sup> standard and due to the alignment with the scientific literature [9] and the existing tools that evaluate environmental impacts, we have selected two environmental impact indicators (ADP and GWP) and one flow indicator (PE) to reflect the most significant results of the system under study.

<sup>2</sup>ITU-T L.1410: <https://www.itu.int/rec/T-REC-L.1410-202411-I/en>

**Exclusions** The following stages and data are excluded from the environmental assessment:

- flows linked to employee transport.
- flows linked to administrative and communication services.
- construction and maintenance of the building housing the equipment.
- spare parts related to equipment maintenance.
- use of equipment for software development.

These exclusions from the scope of the study are justified by the fact that the study aims to focus on the IT aspects of the service. More managerial issues, i.e. sustainability of the workplace and other support function, may influence the impacts associated with the service, particularly during its development phase. However, as workers are spread across different geographical locations and data on the development history is difficult to collect, it was decided to exclude this issue from the study.

## 4.2 LCA modeling

In our analysis, we consider the *embodied* and the *operational* impacts associated with five Pl@ntNet architecture sections: *End User Terminals*, *Networks*, *Training*, *Inference*, and *Web-hosting*. We describe here the model used to estimate each section of the Pl@ntNet service.

### 4.2.1 End User Terminals

The *End User Terminals* impact refers to the impacts caused by the user’s terminal when accessing the Pl@ntnet service.

Let  $N$  represent the number of identification requests. Each device used to access the service is referred to as equipment  $e$ , with a lifespan of  $L_e$  years and an average daily usage time of  $\text{timePerDay}_e$  minutes. A single identification session has an average duration of  $d$  minutes.

**Embodied Impacts** To estimate the end users’ embodied impacts, we first calculate, for each equipment  $e$ , the allocation factor using the Pl@ntNet service,  $\text{alloc}_e$ . It corresponds to the time spent by the Pl@ntNet service divided by the total time used by the equipment. For the time spent in Pl@ntNet, we consider it to be the product of the session duration  $d$  and the total number of requests  $N$ . The total time utilizing an equipment  $e$  is given by the usage time per day  $\text{timePerDay}_e$  (in minutes) multiplied by 365 days and the lifespan  $L_e$  (in years). The allocation factor computation for each equipment is given by the Equation below.

$$\text{alloc}_e = \frac{d}{365 \times \text{timePerDay}_e \times L_e} \times N$$

Then, we compute the environmental impact for the total number of terminals by computing the sum for all equipment of the product between the impact of manufacturing one equipment  $I_{\text{capex}}^e$  and the allocation factor  $\text{alloc}_e$ , as described in the equation below. Note that the notation  $I_*^*$  is used in this section to represent any impact metric, where the superscript and subscript vary depending on the equipment and the LCA stage. These values may correspond to one of the evaluated environmental metrics (ADP in kgSbeq, GWP in kgCO<sub>2</sub>eq, or PE in MJ).

$$I_{\text{endUser}}^{\text{capex}} = \sum_e \text{alloc}_e \times I_{\text{capex}}^e$$

**Operational Impacts** To estimate the operational impacts of terminal usage, we first calculate the equipment power,  $P_e$  (in kW), by dividing the equipment’s annual energy consumption, denoted as  $\text{energyPerYear}_e$ , by the total equipment usage time over one year, computed by multiplying  $\text{timePerDay}_e$  in minutes by 365 days. Then, it is divided by 60 to convert to hours. The computation is described in the equation below.

$$P_e(kW) = \frac{\text{energyPerYear}_e}{365 \times \text{timePerDay}_e \times 1/60}$$

Then, we estimate the total energy consumption for equipment  $e$  in a country  $c$ , denoted as  $C_e^c$  (in kWh), during one year due to Pl@ntNet usage. The equation below describes the energy computation. The usage time is computed by multiplying the session duration  $d$  in minutes, the total number of requests  $N$ , and the terms  $k_e$  and  $k_c$ , which refer, respectively, to the proportion of requests allocated to equipment  $e$  and originated in country  $c$ . Additionally, we divide the time by 60 to convert to hours.

$$C_e^c(kWh) = \left( d \times N \times k_e \times k_c \times \frac{1}{60} \right) \times P_e$$

Finally, we compute the operational environmental impacts by multiplying the energy consumption and Electricity Grid Mix (EGM) for all countries and equipment types. The EGM provides the environmental impact per unit of energy consumed (in kWh), and its value varies by country. It is important to note that the unit of measurement for EGM depends on the specific impact being assessed (e.g., ADP, GWP, or PE). The following equation summarizes the computation of operational impacts.

$$I_{\text{endUser}}^{\text{opex}} = \sum_{c \in C} \sum_{e \in E} C_e^c \times \text{EGM}_c$$

#### 4.2.2 Network

For network environmental impacts in Pl@ntNet, the data exchanged between the terminal equipment and the Pl@ntnet server is considered when the user makes an identification request.

When the user requests image identification, it transmits the image to the Pl@ntNet data center. In this process, the image is automatically compressed by the application. Then, after the server processes the request, it retransmits the answer to the user. The data exchanged between the user and the data center has a total volume  $S$  (in Bytes). The network link between the end-user terminal and the server may have different types, depending on the equipment used (e.g., mobile networks, Wi-Fi). We denote by  $T$  the set of network types used by the end user terminals. We denote the ratio of requests using  $t$  as  $k_t$ . For each type  $t$ , an environmental impact  $I^t$  is associated per unit of transmitted data (i.e., byte). The equation below describes the computation of environmental impacts for the *network* section:

$$I_{\text{network}} = \sum_{t \in T} k_t \times S \times I^t$$

#### 4.2.3 Training

The training impact is associated with the effects resulting from the model training and development employed by the Pl@ntNet service. We consider that each equipment  $e$  used for training contains  $nGPU_e$  GPU graphic cards.

**Embodied Impacts** Here, we describe the estimation of the *embodied* impacts associated with the *training* section. For equipment  $e$ , we denote as  $\text{alloc}_e$  the ratio between the time used for training Pl@ntNet models in equipment  $e$ , denoted as  $H_e$  (in hours), and the total time used by that equipment. The total usage time of the equipment in a year is estimated by the product of the number of GPUs in the equipment, referred to as  $nGPU_e$ , the equipment lifespan (in years),  $L_e$ , and the utilization ratio, represented by  $u_e$ . The following equation describes the computation of the allocation factor.

$$\text{alloc}_e = \frac{H_e}{365 \times 24 \times u_e \times L_e \times nGPU_e}$$

Then, we compute the *embodied* impacts during *training* as the product between the allocation factor and the impact for manufacturing the equipment  $e$ , denoted as  $I_{\text{capex}}^e$ .

$$I_{\text{training}}^{\text{capex}} = \sum_{e \in E} \text{alloc}_e \times I_{\text{capex}}^e$$

**Operational Impacts** The operational environmental impact refers to the energy consumption that occurs during the training process of machine learning models. A significant portion of this impact comes from the electricity usage of data centers, which includes not only the energy consumed by servers but also that required for supporting infrastructure, such as cooling systems.

To account for the full scope of this energy usage, many studies introduce a multiplicative factor (denoted here as  $\rho$ ) when estimating operational impact. The most commonly used metric is the *Power Usage Effectiveness (PUE)*, which quantifies the ratio between the total energy consumed by a data center and the energy used specifically by the servers. While widely adopted, PUE often provides a simplified view by primarily focusing on the energy required during active production periods.

To address the limitations of PUE, recent work by Luccioni et al. [25] and Morand et al. [26] proposes using the *Dynamic Ratio* as a more precise alternative.

This metric captures a broader range of operational states by considering three distinct modes: (i) Production: nodes are powered on and running jobs; (ii) Idle: nodes are powered on, but no jobs are running; and (iii) Infrastructure: nodes are powered off, while the rest of the infrastructure remains operational.

The *Dynamic Ratio* is computed as the ratio of the total energy consumption across all three operating modes to the energy consumption in the production mode alone, in contrast to PUE, which typically includes only the production and infrastructure modes.

Our goal is to test different multiplicative factors when evaluating the operational impacts. For this, we first compute the equipment energy consumption as the product of the equipment TDP power,  $P_e$ , the number of training hours,  $H_e$ , and the multiplicative factor to account for the server infrastructure consumption, denoted as  $\rho$  (PUE or *Dynamic Ratio*). Furthermore, we convert to kWh by multiplying by 0.001, as described in the equation below.

$$C_e(\text{kWh}) = P_e \times H_e \times \rho \times 0.001$$

Then, we estimate the operational impact for training by multiplying the energy consumption  $C_e$  and the electricity grid mix (EGM), which converts the total energy assessed to one of the evaluated metrics (ADP, GWP, and PE). The following equation describes the computation of operational impacts.

$$I_{\text{training}}^{\text{opex}} = \sum_{e \in E} C_e \times \text{EGM}$$

#### 4.2.4 Inference

During inference, the models are asked to predict the image classification labels based on the input containing the raw image data. We consider that for each equipment  $e$ , a ratio  $k_e$  of the total  $N$  inference requests is directed to the equipment. Furthermore, each piece of equipment performs the inference request within time  $t_e$ .

**Embodied Impacts** The eight inference servers operated at CIRAD are entirely dedicated to the Pl@ntNet service. Then, the embodied environmental impact of the inference equipment is simply evaluated using the equation below, where  $I_{\text{capex}}^e$  refers to the environmental impact of manufacturing the inference equipment  $e$  and  $L_e$  to the equipment lifespan in years.

$$I_{\text{inference}}^{\text{capex}} = \sum_{e \in E} \frac{I_{\text{capex}}^e}{L_e}$$

**Operational Impacts** For estimating the operational environmental impacts, we first compute the total energy consumption by multiplying the equipment power  $P_e$ , the number of requests directed to the equipment  $e$  (obtained by the product of  $N$  and  $k_e$ ) and the single request inference time  $t_e$  to compute dynamic consumption. To this dynamic consumption, we add the *idle* consumption,  $P_e^{\text{idle}}$ , estimated by the technical characteristics of servers using methods from [29]. This baseline power is associated with  $T_e^{\text{idle}}$ , a duration ranging from total time of active compute ( $N \times k_e \times t_e$ ) to a full year (conservative hypothesis)<sup>3</sup>. Furthermore, we multiply by 0.001 to convert to kWh and by the multiplicative factor  $\rho_e$ , as described in the following equation.

$$C_e(\text{Wh}) = ((P_e \times N \times k_e \times t_e) + (P_e^{\text{idle}} \times T_e^{\text{idle}})) \times \rho_e \times 0.001$$

Then, we estimate the operational environmental impacts during inference, in a similar way to training, as described in the equation below:

$$I_{\text{inference}}^{\text{opex}} = \sum_{e \in E} C_e \times \text{EGM}$$

#### 4.2.5 Webhosting

The webhosting section comprises the servers used for storing and managing the application, processing user requests, and providing the graphical interface. This section also covers the database management and the storage of all Pl@ntnet service data. To handle this section, we will use the methods from [8] to handle a group of webservices. Therefore, we divide the analysis into two modules: (i) API, which consists of Virtual Machines (VMs); and (ii) Storage, comprising physical nodes responsible for managing the Pl@ntNet database.

For the API deployment, several Virtual Machines (VMs) are deployed along a physical machine. The VMs may have different requirements for computational resources, such as CPU

---

<sup>3</sup>As the minimalistic scenario is at list when the equipment is powered only during treating inference workload and at the other end the equipment is always powered. We used in our study the conservative hypothesis of a full year.

and RAM. In our analysis, we assess the environmental impacts considering four different resources: CPU, RAM, Storage, and other resources (which comprise other components, such as case, motherboard, assembly, and power).

We denote the set of virtual machines deployed to implement Pl@ntnet services as  $V$ . Each virtual machine  $vm \in V$  presents an allocation factor  $a_v$ , which represents the ratio of resources allocated over the total resources provided by the physical server. The equations below describe the allocation factor for CPU, RAM, and storage.

$$a_{vm}^{cpu} = \frac{nCPU_{vm}}{nCPU_{phy}}; a_{vm}^{ram} = \frac{ram_{vm}}{ram_{phy}}; a_{vm}^{sto} = \frac{sto_{vm}}{sto_{phy}}$$

For the *rest* allocation factor, due to the difficulty in precisely quantifying the resource allocation for each virtual machine (VM), we define it as the average of the CPU, RAM, and storage allocation factors. This average serves as a representative approximation for the allocation of other measurable components.

$$a_{vm}^{rest} = \frac{a_{vm}^{cpu} + a_{vm}^{ram} + a_{vm}^{sto}}{3}$$

**Embodied Impacts** The embodied environmental impacts of *webhosting* section are estimated in distinct ways depending on the evaluated services. *For evaluating the API service, which uses Virtual Machines (VMs)*, the embodied impact is given by the product of the physical server manufacturing impact and the allocation factor of each VM for the different resources, as described in the equation below.  $I_{capex}^{cpu}$ ,  $I_{capex}^{ram}$ ,  $I_{capex}^{sto}$ , and  $I_{capex}^{rest}$  refer to the environmental impact for manufacturing, respectively, the CPU, RAM, storage, and restresources in the webhosting server. Furthermore, the estimation of the impacts also considers the equipment lifetime,  $L_e$  and utilization rate,  $u_e$ .

$$I_{API}^{capex} = \frac{\sum_{vm \in V} a_{vm}^{cpu} \times I_{capex}^{cpu} + a_{vm}^{ram} \times I_{capex}^{ram} + a_{vm}^{sto} \times I_{capex}^{sto} + a_{vm}^{rest} \times I_{capex}^{rest}}{u_e \times L_e}$$

*For the storage service*, which uses physical machines for deployment, we consider that all the computation resources from the machines are allocated to the Pl@ntNet service. We denote as  $I_{capex}^e$  the environmental impact of manufacturing the physical machines hosting the database. The equation below describes the estimation of embodied impacts for the database service.

$$I_{storage}^{capex} = \sum_e \frac{I_{capex}^e}{u_e \times L_e}$$

**Operational Impacts** The operational environmental impacts consider that machines (physical or virtual) operate in two different modes: *idle* (when no process is running) and *charge* (when processes are being executed).

The annual operational impact on physical machines for each operation mode can be computed using the equation below, where  $\eta_e$  represents the number of blades on the physical machine and  $P_e^{idle,charge}$  corresponds to the blade power during idle or charge mode.

In the equation below, we compute the total energy over a year, in kWh, based on a blade power  $P_e$ , the total number of blades,  $\eta_e$ , and the usage time (in hours), obtained by multiplying the number of days in a year (365 days) and the number of hours in a day (24 hours). Then, the energy is converted to kWh by multiplying by 0.001.

$$C_e(kWh) = \eta_e \times P_e^{idle,charge} \times 365 \times 24 \times 0.001$$

Then, we estimate the physical impacts by multiplying the energy consumption ( $C_e$ ) and the Electricity Grid Mix (EGM), as shown in the equation below.

$$I_e^{\text{idle,charge}} = C_e \times \text{EGM}$$

For the API service, we consider that the VMs may be in idle or charge mode. When the virtual machine is in charge, we estimate the impact on the VM as the product of the time-share in charge mode,  $s_{\text{vm}}^{\text{charge}}$ , the impact of the physical machine  $I_{\text{charge}}^{\text{phy}}$ , and the CPU allocation factor,  $a_{\text{vm}}^{\text{cpu}}$ , since the CPU corresponds to the highest consuming source.

On the other hand, when the machine is in idle mode, the environmental impact is estimated as the product of the time-share in idle mode,  $s_{\text{vm}}^{\text{idle}}$ , the operational impact in idle mode,  $I_{\text{idle}}^{\text{phy}}$ , and the allocation factor of rest,  $a_{\text{vm}}^{\text{rest}}$ , since in idle mode, the CPU, RAM and storage are not being effectively used.

The equation below describes the operational environmental impacts of the API service.

$$I_{\text{API}}^{\text{opex}} = \sum_{\text{vm} \in V} s_{\text{vm}}^{\text{idle}} \times a_{\text{vm}}^{\text{rest}} \times I_e^{\text{idle}} + s_{\text{vm}}^{\text{charge}} \times a_{\text{vm}}^{\text{cpu}} \times I_e^{\text{charge}}$$

In the same way, for estimating the usage impacts for the database service, we utilize the physical machine impact in the idle and charge mode, respectively  $I_e^{\text{idle}}$  and  $I_e^{\text{charge}}$ , and share of time that the equipments are in idle and charge mode, respectively  $s_e^{\text{idle}}$  and  $s_e^{\text{charge}}$ . The following equation illustrates the computation.

$$I_{\text{storage}}^{\text{opex}} = \sum_e s_e^{\text{idle}} \times I_e^{\text{idle}} + s_e^{\text{charge}} \times I_e^{\text{charge}}$$

### 4.3 Parameters and sources

During our Life Cycle Assessment (LCA), we use the following parameters to evaluate the environmental impacts associated with Pl@ntNet usage. Most of the data of the analysis comes from direct measurements and experiments. In the case of work on assumptions and sensitivity analyses, we used data from studies. The detail of parameters for our analysis are presented in Table 4.

- **Session duration:** Based on Pl@ntNet statistics, the service accounted for over 220M of identification requests in 2024. We consider that a session for identifying the plant species has a duration  $d$  of 30sec. We estimate this value based on exchanges with Pl@ntNet engineers and a practical evaluation of the time required, which includes taking the photo, requesting the identification, and obtaining and visualizing the identification results. Furthermore, we consider that the exchanged packets between the Pl@ntNet data center and the application have an average size of 367 kB, based on discussions with Pl@ntNet engineers.
- **Considered terminal equipments:** Pl@ntNet users access to the service with PC, tablets, smartphones, and smart TVs. As smartphone users accounted for more than 99.1% of the 220M identification requests in 2024, we focus only on this type of equipment in the current study. For smartphone specifications, we assume the average daily time spent on a smartphone is 223 minutes, according to the Global Overview Report [24]. Furthermore, the ADEME [1] reports that the average lifespan is 3 years, and it consumes, on average, 4 kWh per year.

- **Training infrastructure:** According to Jean Zay, as reported in Morand et al. [26], a supercomputer used for training has a lifespan of 6 years, with an 85% utilization ratio. Furthermore, in our analysis, we tested several multiplicative factors: (i) No Factor, when  $\rho = 1.0$ , (ii) PUE ( $\rho = 1.21$ ), and (iii) dynamic ratio ( $\rho = 1.86$ ), described in Jean Zay specifications and reported in Morand et al. [26].
- **Datacenters parameters:** For the *Inference* and *Webhosting* data centers, we consider the average PUE in France to be 1.46, according to the ARCEP report [4]. We assume that the servers are used exclusively for Pl@ntNet, have all a lifespan of 6 years and are always powered, i.e. either treating a workload or on *idle* mode.
- **Networks:** The *Network* impacts were extracted from the European Digital Technology report [12]. The report estimates the environmental impacts of mobile data networking in Europe.
- **Electricity Grid Mix (EGM) :** The EGM impact for all the countries was retrieved from ADEME database [1]. The France EGM was used to estimate the impact of the *Training*, *Inference*, and *Webhosting* sections. For End Users, we used the EGM of all evaluated countries.
- **Manufacturing impacts: Manufacturing impacts of servers** For computing the environmental impacts associated with manufacturing the servers used by Pl@ntnet, we utilized the Boavizta tool [11]. The tool estimates the ADP, GWP, and PE impacts based on the specifications of the hardware components. This estimation was performed for the *Training*, *Inference*, and *Webhosting* sections.

**Manufacturing impacts of smartphones:** For the embodied impacts of smartphones, the ADEME report [1] estimates the ADP, GWP, and PE impacts for manufacturing the equipment.

Parameter	Value	Source
$N$	220M	Pl@ntNet
$d$	30 sec	Pl@ntNet
$S$	367 KB	Pl@ntNet
timePerDay (smartphone)	223	GWI [24]
$L_e$ (smartphone)	3 years	ADEME [1]
energyPerYear	4 kWh	ADEME [1]
$u_e$ (training)	85%	Jean Zay [19]
$u_e$ (inference)	100%	Pl@ntNet
$u_e$ (webhosting)	100%	Pl@ntNet
$L_e$ (training)	6 years	Jean Zay [19]
$L_e$ (inference)	6 years	Pl@ntNet
$L_e$ (webhosting)	6 years	Pl@ntNet
$\rho$ (training)	[1.0,1.21,1.86]	Jean Zay [19]
$\rho$ (inference)	1.46	ARCEP [4]

Table 4: List of parameters used in Pl@ntNet Life Cycle Assessment with their respective values and sources.

- **Training times:** Table 5 shows the total GPU hours used for training Pl@ntNet models across three scenarios: *Final Models*, *Standard*, and *Legacy*. These values reflect usage on V100 and A100 GPUs in the Jean Zay supercomputer.

Scenario	V100 hours	A100 hours	Source
<i>Final Models (2024)</i>	61,440	0	PI@ntNet
<i>Standard (2024)</i>	83,704	96,507	PI@ntNet
<i>Legacy(2020-2024)</i>	292,065	96,507	PI@ntNet

Table 5: Training time (in number of hours) of PI@ntNet service models for each scenario in Jean Zay supercomputer. We display the time for each type of node used.

## 5 Results

In this section, we present the results of the Life Cycle Assessment (LCA) of the PI@ntNet service. We conducted a multi-criteria, multi-stage evaluation of environmental impacts using the methodology described in Section 4. All the parameters used are described in Table 4. First, we discuss the global impacts of the PI@ntNet service. We identify the sections of the PI@ntNet architecture that most contributed to environmental impacts, and we compare the *embodied* and *operational* impacts.

Then, we describe in detail the environmental impacts obtained for each PI@ntNet section: *End Users*, *Networks*, *Training*, *Inference*, and *Webhosting*. During the analysis, we discuss the main impacts and identify optimization opportunities for the PI@ntNet service.

### 5.1 Global environmental impacts

In this section, we analyze the environmental impacts of the PI@ntnet service. Table 6 summarizes the environmental impacts caused by PI@ntNet in 2024, based on an average request session time of 30 seconds. The assessment also accounts for *training* under the *standard* and *legacy* scenarios and applies the *PUE* as a multiplicative factor (See Table 9) In the *standard* scenario, it was responsible for emitting 71.9 tCO<sub>2</sub>eq. Furthermore, its impact on primary energy was  $2.45 \times 10^6$  MJ, and its ADP impact was 1.83 kgSbeq.

Section	ADP (kgSbeq)	GWP ( $\times 10^3$ kgCO <sub>2</sub> eq)	PE ( $\times 10^6$ MJ)
End Users	1.14	49.2	0.69
Networks	0.07	8.0	0.4
Training ( <b>standard</b> )	<b>0.2</b>	<b>7.9</b>	<b>0.9</b>
Training ( <i>legacy</i> )	<i>0.5</i>	<i>15.1</i>	<i>1.7</i>
Inference	0.22	3.75	0.37
Webhosting	0.19	2.97	0.08
<b>Total (standard)</b>	<b>1.83</b>	<b>71.9</b>	<b>2.45</b>
<i>Total (legacy)</i>	<i>2.1</i>	<i>79.0</i>	<i>3.25</i>

Table 6: PI@ntNet environmental impacts per section considering **standard** and *legacy* scenarios.

Fig. 5 illustrates the contribution of each service section to the total impact according to *standard* (Fig 5a) and *legacy* (Fig. 5b) scenarios. We observe that, for Abiotic Depletion Potential (ADP), the *End Users* section presents the highest impact. The section was responsible for extracting 1.14 kgSbeq. Similarly, the *End Users* section was responsible for the highest GWP impact. This section emitted 49.2 tCO<sub>2</sub>eq, equivalent to 68.4% of the total GWP impact (*standard* scenario). This is attributed to the resource-intensive manufacturing process for producing the devices (smartphones) used to access the PI@ntNet service. During this process, a significant amount of raw materials is extracted, and a large quantity of greenhouse gases is emitted.

On the other hand, for Primary Energy (PE), the *Training* component had the greatest impact. The section was responsible for consuming  $0.9 \times 10^6$  MJ (and  $1.7 \times 10^6$  MJ in *legacy* scenario), equivalent to 36.7% of the total energy. The Machine Learning training process requires substantial computational resources, thereby increasing energy consumption.

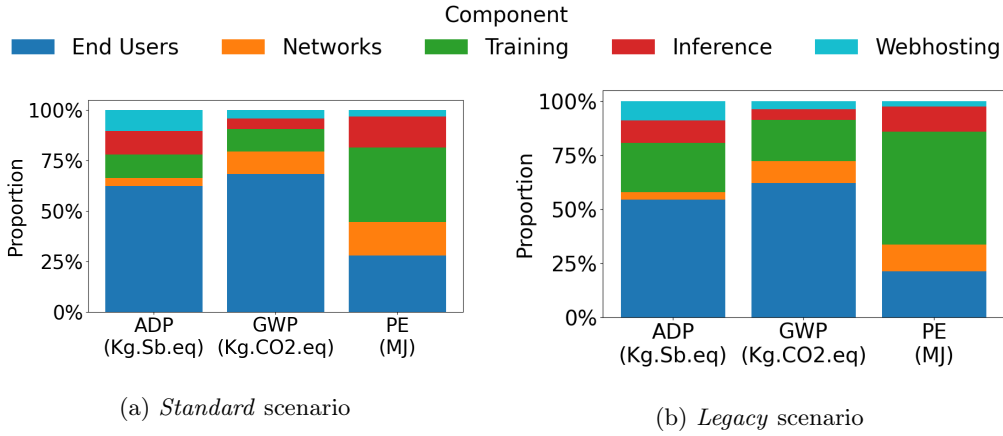


Figure 5: Embodied and operational environmental impacts proportions over different sections in Pl@ntNet using two training scenarios: (a) *standard* and (b) *legacy*.

**Embodied and Operational Impacts** Figs. 6a and 6b describe, respectively, the *embodied* (capex) and *operational* (opex) impacts for each section, according to *standard* scenario. The *Networks* category is excluded from this analysis, as it does not decompose into capex and opex impacts. For the *embodied* impacts (Fig. 6a), we observe that the *end users* section presents a high impact. The section was responsible for emitting 46 tCO<sub>2</sub>eq for manufacturing, which reflects the application widespread usage. In 2024 alone, more than 20 million users used Pl@ntNet, resulting in more than 220 million identification requests.

For the *operational* impacts (Fig. 6b), a significant portion of the environmental impacts is attributed to *training*. In *standard* scenario, datacenter usage for *training* accounted for 5.77 tCO<sub>2</sub>eq (and 10.8 tCO<sub>2</sub>eq in *legacy* scenario), representing 53.5% (and 68.2% in *legacy* scenario) of the total operational GWP impact associated with Pl@ntNet (excluding *network*).

On the other hand, other sectors presented a lower *operational* impact. The *webhosting* section, which manages the API and the Pl@ntnet database, presented a GWP impact of 304 kgCO<sub>2</sub>eq, which is significantly lower when compared to the *training* section. When comparing training and inference, considering *standard* scenario, the GWP impact of *training* is approximately 2.4 times that of *inference*.

## 5.2 Detailed assessment of Pl@ntNet impacts

Following, we provide and discuss LCA results and the environmental impacts associated to each of the evaluated Pl@ntNet sections.

### 5.2.1 End Users

Table 7 describes the environmental impacts for the *End Users* section in Pl@ntNet service. Analyzing the total Global Warming Potential (GWP), the *End Users* section emitted 49.4 tCO<sub>2</sub>eq.

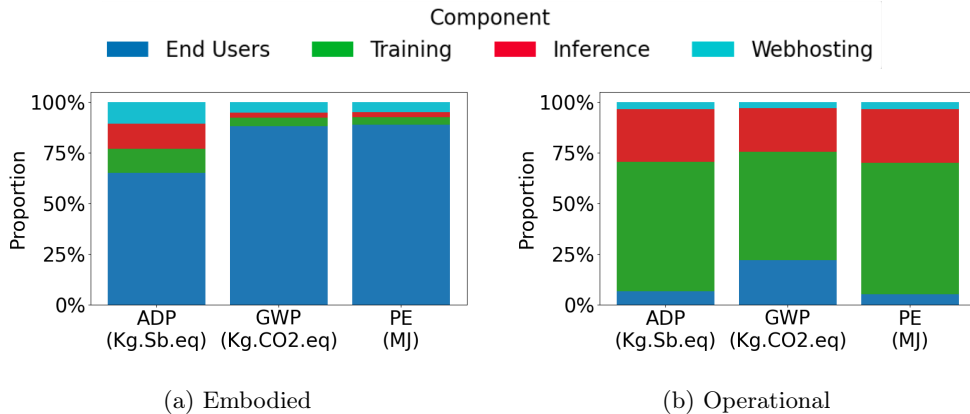


Figure 6: Embodied and operational environmental impacts proportions over different sections in Pl@ntNet using *standard* scenario.

From the total emissions, the embodied stage is responsible for the largest amount of impacts, emitting 46.9 tCO<sub>2</sub>eq, equivalent to 95.3% of the total section emissions.

The observed result aligns with that of Ercan et al. [15]. Performing a Life Cycle Assessment (LCA) of a Sony smartphone, the authors highlight a high Global Warming Potential (GWP) attributed to the production phase. According to them, the integrated circuits, mainly, and display manufacturing are responsible for the main environmental impacts of the evaluated device.

	ADP (kgSbeq)	GWP ( $\times 10^3$ kgCO <sub>2</sub> eq)	PE ( $\times 10^4$ MJ)
Embodied	1.14	46.9	62
Operational	0.0003	2.3	7.0
<b>Total</b>	1.14	49.2	69

Table 7: *End users* environmental impacts for embodied (capex) and operational (opex) stages.

**Comparing countries impacts** We also analyzed the impact of *End Users* during the *operational* phase, depending on the location of requests. Fig. 7 describes the GWP impact during the *operational* phase for each evaluated country. The displayed countries are responsible for a large majority of Pl@ntNet sessions, accounting for over 95%.

The impacts in each may vary depending on Pl@ntNet usage patterns (the number of requests) and the energy grid mix (EGM) of each country. For instance, when comparing the impacts between France and Germany, we observe that France was responsible for 25% of the total sessions, more than twice that of Germany, as observed in Table 2. However, the French GWP impact was 160 kgCO<sub>2</sub>eq, approximately 4× lower than Germany’s GWP impact of 624 kgCO<sub>2</sub>eq. The observed gap occurs since, although France presents more identification requests, the low-carbon sources are larger in the electricity grid mix, leading to attenuated environmental impacts.

On the other hand, some countries, although accounting for a small share of Pl@ntNet sessions, have a significant GWP impact. For instance, Czechia and Poland account for, respectively, 2% and 1.8% of the sessions, but contribute 103.7 and 117 kgCO<sub>2</sub>eq, respectively.

**Sensitivity analysis** To better understand the environmental impacts associated with the *End User Terminals* section, we performed a *sensitivity analysis* using usage statistics from *Google*

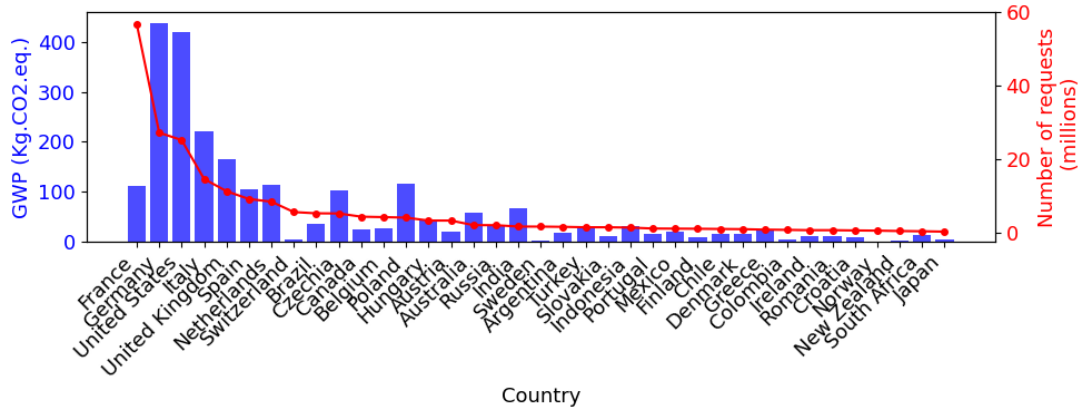


Figure 7: Impact for *End Users* section during usage (OPEX) for different countries. On left, we represent the GWP impact (in kgCO<sub>2</sub>eq) and on right, we show the number of identification requests for each country in 2024. The countries are ordered by number of requests in decreasing order.

*Analytics*, a web and mobile platform that tracks and reports user engagement metrics.

Then, we estimated the impacts under three alternative scenarios based on usage statistics and compared them with our baseline scenario, which assumes a fixed session duration of 30 seconds. Figure 8 shows the GWP impact of each alternative scenario.

In the first scenario, we consider the total user engagement time provided by *Google Analytics* across the evaluated countries. In 2024, the platform recorded that users spent 14 billion seconds in Pl@ntNet application. This engagement time resulted in a total GWP impact of 102tCO<sub>2</sub>eq, approximately 2× our initial estimate based on a session-duration hypothesis.

However, a substantial amount of the user engagement time is not identified or is associated with an unlabeled page. To assess the influence of this uncertainty, we defined a second scenario excluding the unidentified share (which corresponds to 61% of total user engagement time). In this scenario, the *End User Terminals* were responsible for 38.9tCO<sub>2</sub>eq.

Finally, in a third scenario, we estimated the impacts by focusing exclusively on the screens that "create a new observation". Under this more restrictive assumption, the *End User Terminals* GWP was further reduced to 25.4tCO<sub>2</sub>eq, which remained below our baseline hypothesis based on the session duration, which accounted for a GWP impact of 49.2tCO<sub>2</sub>eq.

**Session time analysis** Users may present wide behaviors when using Pl@ntNet services. Actions such as taking the photo and consulting the results may exhibit significant variation. Thus, we performed a sensitivity analysis by varying the session duration from 30s to 2min30s with a 15s interval.

In Figs. 9a-9c, we display the ADP, GWP and PE impacts, respectively, as a function of the request duration time. We observe that the environmental impacts scale linearly with the session duration. For GWP impact, *End Users* section emissions vary from a total of 49 tCO<sub>2</sub>eq to 250 tCO<sub>2</sub>eq. In the same way, the ADPe and the PE varies from 1.14 to 5.71 kgSbeq and from  $691 \times 10^3$  to  $3.4 \times 10^6$  MJ.

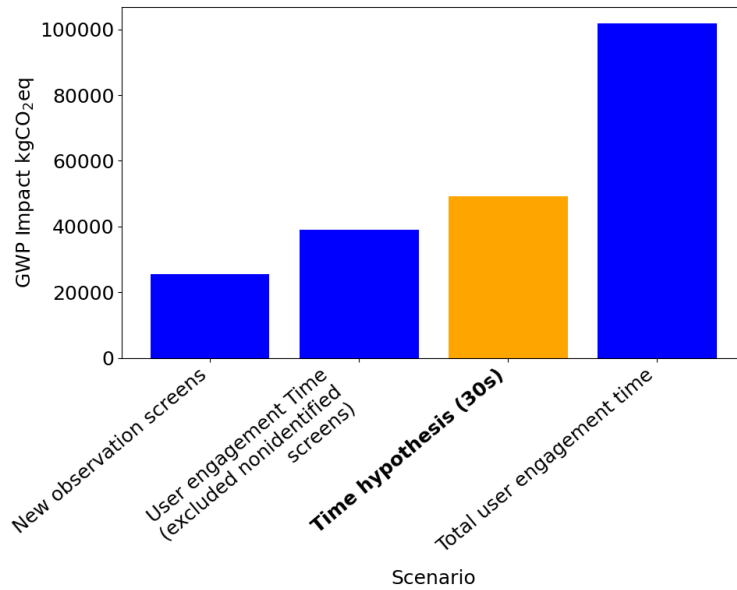


Figure 8: GWP Impact (kgCO<sub>2</sub>eq) of *End user terminals* in Pl@ntnet application according to the different scenarios using *Google Analytics* statistics and time hypothesis. We provide four scenarios, considering: (i) exclusively the engagement time for the screens concerning *new observation*; (ii) total user engagement time (excluding non-identified screens); (iii) an hypothesis for the session duration of 30s; and (iv) the total user engagement time.

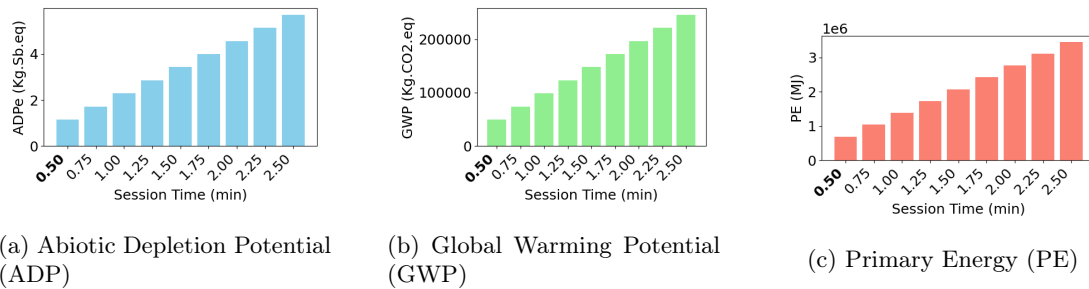


Figure 9: Sensitivity analysis over the identification request duration. We varied the duration in a range from 0.5min to 2.5min and we display the (a) ADP, (b) GWP and (c) PE impacts associated with *End Users Terminals*. For the following report, the session duration  $d = 0.5$  (**in bold**) is used.

**Conclusions regarding End user terminals** Estimating the impact of user terminals on service is subject to the influence of several parameters that are difficult to control. Session duration greatly affects the footprint allocated to terminals, as shown in Figure 8 and 9c. Looking at the data displayed in Figure 8, the total amount of estimated screen time for Pl@ntNet users varies 51% to 203% times the baseline scenario estimated by multiplying the number of identification processes by 30 seconds, i.e.  $6.81 \times 10^9$ .

However, a second parameter has an equally significant influence on terminal footprint allocation and is perhaps even more difficult to estimate: the total usage time of the user’s smartphone, i.e. in our model the  $\text{timePerDay}_e$  parameter. We used a figure from the ADEME dataset associated with smartphone footprints as data for this parameter, i.e. 227 minutes per day on average. However, a recent large-scale study [33] has published new empirical data on this same parameter, announcing an average of 391 minutes per day. This increases the total smartphone usage time by 175% and therefore significantly reduces the CAPEX footprint of the end user terminals. The addition of this uncertainty makes the footprint of end user terminals even more uncertain. This is leading to a variation of 31% to 207%, between min. and max. scenario, of the carbon footprint of end users terminals compare to the baseline scenario. In conclusion, while it remains important to include the footprint of terminals in the assessment of the service, considering 1) the uncertainties surrounding their footprint, and 2) the reduced ability of service managers to influence this impact, it may be interesting to observe the footprint of the service without including terminals. If the objective is to reduce the service’s footprint by the Pl@ntNet team, Figure 10 may be a more effective tool for communicating about the service’s footprint.

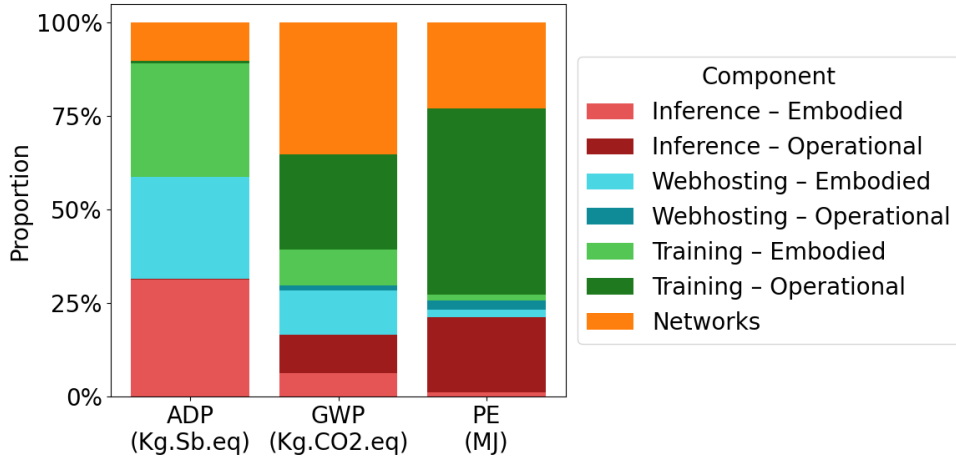


Figure 10: Distribution of the environmental impact of Pl@ntNet service for the *standard* scenario, excluding End-user Terminals

### 5.2.2 Networks

We display in Table 8 the environmental impacts associated with Networks section of Pl@ntNet service. The network impacts methodology do not differentiate between *embodied* (capex) and *operational* (opex), considering them as an aggregated value (See Section 4).

The *Network* section was responsible for 8tCO<sub>2</sub>eq in 2024, equivalent to 10.2% of total emissions. Although we consider that mobile data links are used, the image compression during

transmission between devices and the datacenters leads to resource optimization, attenuating the impacts.

	ADP (kgSbeq)	GWP (kgCO <sub>2</sub> eq)	PE (MJ)
<i>Networks</i> Impact	$7.1 \times 10^{-2}$	$8.0 \times 10^3$	$4.6 \times 10^5$

Table 8: *Networks* environmental for Pl@ntNet service in 2024. The data exchanged in the request and answer is considered in this analysis.

### 5.2.3 Training

We also evaluated the environmental impacts caused by the Pl@ntNet models *training*.

Then, in our evaluation, we vary the multiplicative factor, considering three different possibilities: (i) No multiplicative factor ( $\rho = 1.0$ ); (ii) PUE ( $\rho = 1.21$ ); and (iii) dynamic ratio ( $\rho = 1.83$ ).

Furthermore, we considered three distinct scenarios for the training temporal window: (i) *Legacy*, when the total training time from 2020 until 2024 is considered; (ii) *Standard*, when the training time performed in 2024 is considered; and (iii) *Final Models*, when we consider the training time only for the final models made available for the users.

Table 9 describes the environmental impact of the *training* section for each evaluated scenario (*Legacy*, *Standard* and *Final Models*) and each multiplicative factor considered in our analysis.

**Training impacts** The high-impact behavior observed in the *training* section can be attributed, firstly, to the continuous emergence of new species and observations, which require frequent model retraining to meet users’ needs for better and more accurate models. For instance, between 2020 and 2024, the number of species identified by Pl@ntnet increased from 30798 to 75415, representing an increase of 145%.

Furthermore, the development of related products, such as models for disease detection and agricultural diagnostics, has expanded. These new initiatives require additional training efforts, thus leveraging environmental impacts due to the increased computational demand.

Finally, Pl@ntNet researchers have recently focused on optimizing the inference process in order to reduce inference time. While this leads to lower environmental impacts during inference, it introduces a trade-off between inference efficiency and training efficiency. To achieve a low inference time, researchers need to submit their models to longer and more exhaustive training rounds, which demands a longer and more expensive training process, potentially increasing environmental impacts.

**Impacts scaling up in 2024** We also observe a considerable impact associated with the *Training* section in 2024. In the *Legacy* scenario, which accounts for the impacts from 2020 to 2024, the GWP impact varies in the range between 13.1 tCO<sub>2</sub>eq and 20.6 tCO<sub>2</sub>eq. On the other hand, in the *Standard* scenario (2024 year), the GWP impact varies between 6.94 tCO<sub>2</sub>eq and 10.9 tCO<sub>2</sub>eq, representing, on average, 50% of the total GWP impact observed for *Training* between 2020 and 2024.

The above observation is aligned with the number of training hours used by Pl@ntNet. The training service consumed 208,000 hours of model training from 2020 to 2023, during four years. On the other hand, in 2024, during just one year, Pl@ntNet consumed 180,000 hours, almost the same amount as the previous four years. The new demands of the Pl@ntNet service and the increased availability of data have significantly scaled up the impacts over the last year, posing a substantial impact on the training section due to the Pl@ntNet service.

**Final Models vs Standard scenarios** We also compared the impacts of the *Final Models* (made available to users) with the *Standard* scenario, which includes developing and testing models. The results described in Table 9 illustrate that, in 2024, the *Final Models* accounted for between 1.85 and 2.87 tCO<sub>2</sub>eq. On the other hand, the total training in 2024 (*Standard* scenario) ranged from 76.94 to 10.9 tCO<sub>2</sub>eq. The *Standard* scenario represents, thus, on average, 4× of the *Final Models* impact.

This illustrates that the impacts may scale up beyond the final model’s training. Developing a model to achieve good accuracy may require intense training, leveraging its impacts. Therefore, it is vital that researchers focus on the ML impacts, not only for the *Final Models*, but also for the model development.

**Embodied vs Operational impacts** Fig. 11 describes the share of GWP impact for the *embodied* and *operational* stages in *Legacy* scenario under different multiplicative factors. We observe that *operational* impacts represent the dominant share of the total GWP emissions during *training*. Depending on the multiplicative factor used, between 68% and 79% of the GWP is due to the *operational* phase. This highlights that the emissions generated while training the models, due to electricity usage and the duration of training, constitute the most significant portion of the environmental impact.

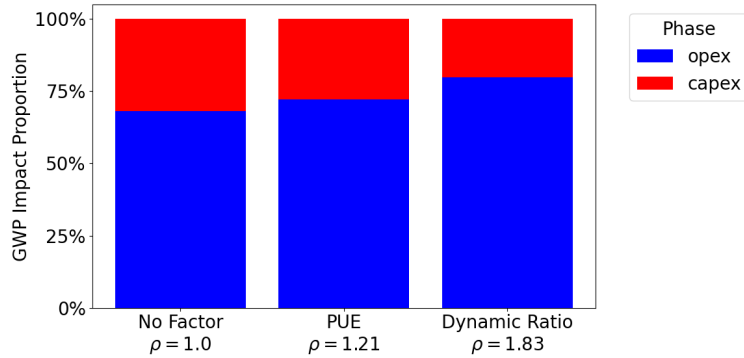


Figure 11: *Training* GWP Impact proportion (in percentage) between *embodied* and *operational* stages. The results consider the *legacy* scenario and test the three multiplicative factors: (i) *No Factor*, (ii) *PUE*, and (iii) *Dynamic Ratio*.

#### 5.2.4 Inference

When a plant identification is requested, the model inference is executed on one of the dedicated servers. All the inference requests are performed using GPUs.

**Comparing inference machines** For each inference server, we collected the average GPU processing time and the proportion of identification requests handled by each machine. All details are described in Table 10.

We observed that only a small share of requests were processed in *capcir* and *orxois* machines. This occurs because these machines are older and dedicated to serving requests from legacy devices, which are no longer compatible with updated versions of the PI@ntNet application. The

Scenario	Multiplicative Factor	ADP ( $\times 10^{-1}$ kgSbeq)	GWP ( $\times 10^3$ kgCO <sub>2</sub> eq)	PE ( $\times 10^6$ MJ)
Final Models (2024)	No Factor $\rho = 1.0$	0.78	1.85	0.19
	<b>PUE</b> $\rho = 1.21$	<b>0.78</b>	<b>2.1</b>	<b>0.23</b>
	<i>Dynamic Ratio</i> $\rho = 1.83$	<i>0.79</i>	<i>2.87</i>	<i>0.35</i>
Standard (2024)	No Factor $\rho = 1.0$	2.12	6.94	0.76
	<b>PUE</b> $\rho = 1.21$	<b>2.13</b>	<b>7.94</b>	<b>0.90</b>
	<i>Dynamic Ratio</i> $\rho = 1.83$	<i>2.15</i>	<i>10.9</i>	<i>1.36</i>
Legacy (2020-2024)	No factor $\rho = 1.0$	4.72	13.2	1.41
	<b>PUE</b> $\rho = 1.21$	<b>4.78</b>	<b>15.1</b>	<b>1.70</b>
	<i>Dynamic Ratio</i> $\rho = 1.83$	<i>4.82</i>	<i>20.6</i>	<i>2.55</i>

Table 9: Environmental impacts of training in three different scenarios: *Final models*, *standard* and *legacy*, considering several multiplicative factors.

Server	GPU	Purchase year	Processing time (ms)	Ratio (%)
<i>capcir</i>	Quadro M4000	2016	103.7	2
<i>orxois</i>	Quadro M4000	2016	122.2	3
<i>huchet</i>	RTX 8000	2020	52.2	1
<i>rajol</i>	RTX 8000	2020	3125.6	25
<i>chamois</i>	A40	2022	158.4	37
<i>devois</i>	A40	2022	62.12	24
<i>senart</i>	A40	2024	50,9	3
<i>sourdun</i>	A40	2024	70.4	5

Table 10: Average inference processing time (in milliseconds) and percentage of total requests handled by each server.

*senart* and *sourdun* machines also accounted for a small share of requests, as they have only been operational since September 2024, and thus have not yet handled a significant volume of traffic.

Fig. 12 describes the embodied and operational environmental impacts for each inference machine. We observe a high operational impact (OPEX) for the *rajol* machine. In 2024, this machine generated 1.77 tCO<sub>2</sub>eq, accounting for 47% of the total inference impact. This increased impact is attributable to the machine’s use in multi-species inference, which involves more complex models and requires significantly longer inference durations. The energy consumption and environmental impacts, therefore, increase substantially.

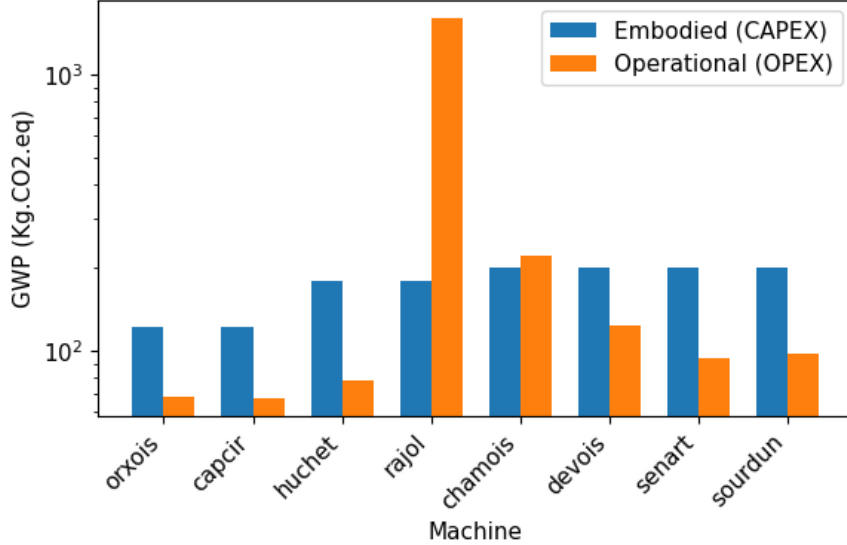


Figure 12: *Inference* GWP impacts (in log scale) per machine during 2024. For each machine, we display (i) embodied (in blue) and (ii) operational stages impacts. The *capcir* and *orxois* machines use a Nvidia Quadro M4000; *huchet* and *rajol* use a NVIDIA RTX 8000; and *chamois*, *devois*, *senart* and *sourdun* use a NVIDIA A40 GPU.

### 5.2.5 Webhosting

Here, we analyze the main impacts associated with the *webhosting* section. We remind that the analysis is divided into two modules: (i) *API*, consisting of Virtual Machines (VMs); and (ii) *Storage*, composed of physical nodes responsible for managing the PI@ntNet database.

Table 11 summarizes the embodied (capex) and operational (opex) environmental impacts for the *Webhosting* section, divided into *API* and *storage* modules.

We first observe that the ADP impact is, as expected, significantly higher in the *embodied* phase than in the *operational* phase for both modules. This is consistent with the fact that ADP primarily accounts for the extraction of raw metal resources, which is directly linked to equipment manufacturing processes.

For the *API* module, the GWP is also higher in the embodied stage, totaling 0.44 tCO<sub>2</sub>eq, which represents 68.75 % of the total GWP for this module. This is mainly due to two factors: (i) the manufacturing of servers is a carbon-intensive process, significantly increasing the embodied

GWP, and (ii) the servers operate in France, where the electricity grid has a low carbon intensity, thus reducing the operational GWP.

In contrast, the PE impact is predominantly concentrated in the operational stage, accounting for  $3.07 \times 10^4$  MJ, or 84.3% of the total PE impact. This is expected, given that the servers run continuously, resulting in sustained energy consumption throughout their operational lifetime.

Regarding the *storage* module, we observe that the *embodied* stage dominates all three impact categories. The GWP and PE impacts in this phase are 2.23 tCO<sub>2</sub>eq and  $2.78 \times 10^4$  MJ, accounting for 95.7% and 64% of the total module impacts, respectively. This is primarily due to the substantial environmental burden associated with the manufacturing of high-capacity SSDs used for the Pl@ntNet application, which involves resource and energy-intensive processes.

<i>API (VMs)</i>			
	ADP (kgSbeq)	GWP (tCO <sub>2</sub> eq)	PE ( $\times 10^4$ MJ)
Embodied	0.08	0.44	0.57
Operational	0.00012	0.20	3.07
<i>Storage (Physical)</i>			
	ADP (kgSbeq)	GWP (tCO <sub>2</sub> eq)	PE ( $\times 10^4$ MJ)
Embodied	0.11	2.23	2.78
Operational	0.00006	0.1	1.56
<b>Total</b>	<b>0.19</b>	<b>2.97</b>	<b>7.98</b>

Table 11: *Webhosting* section environmental impacts for the *APIs* (virtual machines) and the *storage* (physical machines) modules.

## 6 Limits and discussion

It is important to highlight the limitations of this study. These limitations are due, on the one hand, to the quality of the data that could be used for the study and, on the other hand, to the methodological choices made.

### 6.1 Limits due to data quality

The quality of the data affects the results of the quantitative model used in this study. This quality is limited by two factors. First, the quality of the environmental footprint data for the digital equipment and infrastructure used in the model. Second, the collection of field data to establish an inventory of the equipment hosting Pl@ntNet and to measure the intensity of use of each piece of equipment.

With regard to the data used, which was mainly sourced from the ADEME and Boavizta databases, the following limitations should be noted:

- The same data was used for the embodied footprint of all smartphones used. This could greatly influence the footprint of the devices if, on average, smartphones have a longer or shorter lifespan.
- The network footprint used the results of a European LCA study and made an allocation based on the total volume of data transported by this network at the European level to allocate the average footprint of one gigabyte transported. This data was used in the absence of a better alternative for quantifying the full life cycle footprint of data transfer via international networks.

- The data used to calculate the embodied footprint of the web hosting and inference part does not take into account the transport and end-of-life phases. This may contribute to a slight underestimation of the capex footprint of this part.
- In the absence of a complete LCA, the footprints of GPUs were estimated using a method normally used to estimate the footprint of CPUs. This may contribute to an underestimation of the capex footprint of the inference and machine learning parts.

With regard to the difficulties associated with collecting ground data, the following limitations should be noted:

- The capabilities of CIRAD’s hypervisor, which hosts the web hosting VMs, are known, and a satisfactory estimate of the hypervisor’s footprint has been calculated. However, it has not been possible to collect data on the hypervisor’s activity outside the capacities allocated to the Pl@ntNet service’s VMs. This means that it is not possible to know whether, outside the Pl@ntNet VMs, the hypervisor is underutilised or not. The embodied footprint of the hypervisor’s VMs allocated to the Pl@ntNet service would be greater if the hypervisor was oversized in relation to needs. This may contribute to an underestimation of the footprint of the webhosting part.
- It was not possible to measure the power consumption of the equipment on site directly using a wattmeter. This may contribute to a slight underestimate of the opex footprint of the webhosting, inference and machine learning part.
- The average duration of user sessions was not measured empirically but estimated. This may contribute to an underestimation of the footprint of the user terminal part.
- In the absence of direct measurement on end-user devices, it was also not possible to take into account and measure the time spent by users using the application outside a typical session defined by the usage scenario. This may contribute to an underestimation of the footprint of the user terminal part.
- In the absence of direct energy measurement on inference servers, the energy consumption of inference workload were estimated through the use of TDP data of equipment. Also, it was not possible to determine the amount of time when the servers were shutdown and a conservative hypothesis was taken were the server are always at list on *idle* mode. This may contribute to a small uncertainty on the footprint of the inference part.
- Internal network equipment at sites linked to Pl@ntNet was not taken into account, nor were network flows between hosting sites. This may contribute to an underestimate of the footprint of the webhosting, inference and machine learning part.
- It was not possible to ascertain the technical resources specifically implemented for data storage at the ISDM-MESO site. The same model was used to assess the IT resource footprint of both CIRAD and ISDM-MESO. It was not possible to take into account the specific characteristics of this site, which specialises in data storage, e.g. more efficient storage equipment or a quality of service requiring more resources, particularly through data replication. This could influence the footprint of the webhosting part.

## 6.2 Limits due to methodological choices

The first main limitation of the study is related to the **attributional LCA framework**. The aim here was not to calculate the footprint generated as a direct result of using the Pl@ntNet

service for one year, but to estimate the footprint for which such a functional unit would be responsible. What environmental footprint can be attributed to using Pl@ntNet for one year? Therefore, the notion of "cost of a single use" of Pl@ntNet to identify one species is more a representation rather than a direct consequence. Like many digital services, this one relies on widely shared equipment and infrastructure. These resources are designed to be used collectively and would not be economically or technically viable to create and maintain for just one user, so the scenario is purely hypothetical

In this framework, the capex footprint of equipment and infrastructure is distributed by the model throughout their usage phase, based on assumptions about the duration and characteristics of this phase. The manufacturing footprint of this equipment exists and cannot be reduced in itself, but it can be distributed over a longer or wider usage period through Pl@ntNet or other services.

It is therefore important not to rely solely on this model when considering the purchase of new equipment. Like most attributional models, this model mathematically favours the use of larger equipment in order to achieve economies of scale. But the model, to allocate the capex footprint, relies on hypothesis around the past and future lifespan and use of these equipments without integrating dynamic phenomena like rebound or induction effect.

The second methodological limitation of this study relates to the choices made regarding the scope defined.

The study chose to focus on the technical architecture of Pl@ntNet, i.e. how the service is operated and functions over the course of a year. Thus, the study takes into account the cost of AI model training in Pl@ntNet's footprint, but not the costs of Pl@ntNet's software development. For reasons related to the study's objectives and a lack of data on the traceability of time and resources devoted to Pl@ntNet's software development, this cost was not included. However, the question of the scope of inclusion of past AI model training phases remains a limitation of the study, even though it is explored in section 5.2.3.

### **6.3 How do the results address environmental concerns about Pl@ntNet?**

What should be thought of the study's results? What do the results reveal about the environmental impact of the Pl@ntNet service?

This study used an LCA based on direct sources to explore the footprint of an AI service with millions of users. This makes it a rare source that allows us to see the dynamics of footprint distribution for a large-scale service.

Within Pl@ntNet, the footprint is mainly carried by the user terminals, followed by the machine learning component. This distribution of the footprint distinguishes Pl@ntNet from more recent AI services based on generative AI models, where the footprint associated with the inference phase tends to be one of the most significant. It is important to note that, although Pl@ntNet users come from many countries, the training, inference and web hosting components are located in France, powered by a largely decarbonized electricity mix, significantly reducing the opex carbon footprint of these parts.

The footprint associated with end users is mainly due to the manufacture of smartphones. This footprint has been allocated to the Pl@ntNet service based on the usage time of these devices. This footprint represents Pl@ntNet's contribution, on its own scale, to a more digitalized society. The model represents this indirect link between the availability of services such as Pl@ntNet and the incentive to own a smartphone. To reduce the footprint of this part, possible solutions include reducing smartphone usage time or enabling Pl@ntNet to run on older smartphones, thereby reducing the need to purchase new devices.

The significance of the machine learning footprint is due to the use of the Jean Zay data centre,

despite its relatively efficient infrastructure powered by low-carbon energy, which highlights an extensive use. Levers to reduce the footprint of this part may be found in its use, as explored in Table 9, which shows that the final training cost is significantly lower than other uses in the training phase. It could be interesting to explore the relationship between model quality gains and computational cost.

More broadly, the environmental impact model proposed by the study could be used to examine the growth dynamics of the Pl@ntNet service. The service is experiencing growth, albeit slowing, in the number of users, but still sustained growth in content and usage<sup>4</sup>. This growth is also reflected in the cost of machine learning. According to Table 9, over the period 2020-2024, the year 2024 alone accounts for more than half of the impacts.

Finally, questioning the environmental impact of Pl@ntNet requires consideration of the purpose of the service. Pl@ntNet is an ambitious participatory science project that has made a positive contribution to academic research and raised awareness among a wider audience. The direct environmental impacts measured during this study can be discussed in relation to the objectives of the service and its secondary impacts. The positive impact of Pl@ntNet on research will be difficult to measure quantitatively, but it can nevertheless be assessed qualitatively.

It could also be interesting to explore in future work the potential consequences of Pl@ntNet on the work of professionals linked to the service's areas of application. For example, has Pl@ntNet made it possible to obtain field data, thereby saving teams from having to travel? Or, on the contrary, does Pl@ntNet's design encourage more travel and more use of the application? Pl@ntNet is also involved in the development of new tools for detecting invasive plants or plant diseases. These issues may be studied in the future through consequential LCA and prudent net impact studies [27, 14].

Exploring the direct environmental impacts of Pl@ntNet and its ecological cost is primarily a first step that should then be used to establish the best possible links between the project's objectives and its means.

## 7 Conclusion

Given the widespread adoption of Artificial Intelligence (AI) models and their integration into several applications, it has become vital to study and quantify the environmental impacts of AI-based services. In this study, we focus on the Pl@ntNet service, an image recognition application designed to identify plant species from photographs. We proposed a Life Cycle Analysis (LCA) of the Pl@ntNet service in order to evaluate the environmental impacts.

Our analysis adopts a multi-criteria and a multi-stage approach. We investigated the Abiotic Depletion Potential equivalent (ADPe), the Global Warming Potential (GWP) and the Primary Energy (PE) impacts associated to the evaluated application. These impacts were evaluated across both embodied phase (production, transport and end of life) and operational stages of Pl@ntNet service.

Pl@ntNet accounted for 71.9 tCO<sub>2</sub>eq emissions in 2024 (and 79 tCO<sub>2</sub>eq considering *legacy* scenario). We structured our analysis in five key sections of the service: *End User*, *Networks*, *Training*, *Inference*, and *Webhosting*. From the total amount, the *End Users* section accounted for the most relevant share, contributing 49.2 tCO<sub>2</sub>eq. to 68.4% of the total GWP. We can also see this section's predominant impact on the depletion of abiotic mineral and metal resources.

The *Training* section also exhibited a substantial environmental impact. Depending on the considered scenario, the *Training* emissions ranged from 2.1 tCO<sub>2</sub>eq (considering the released in

---

<sup>4</sup><https://identify.plantnet.org/fr/stats>

2024) to 15.1 tCO<sub>2</sub>eq (for all training activities from 2020 to 2024). These results highlight the energy intensity associated with *Training*, making it a critical target for impact mitigation.

In the future, we emphasize the importance of improving optimization techniques to reduce the environmental footprint. These advances, especially in training, are essential for mitigating the impacts of AI-based services such as Pl@ntNet.

In this work, we focused on assessing the service’s direct environmental impacts. We are aware that this approach is single-focused and could be complemented by two other evaluation areas: *indirect* and *social* impacts. These assessments would improve the systemic understanding of service impacts, shed light on the diversity of uses across contexts, and uncover risks of human and infrastructure-related social harm.

In the case of the development of a globally deployed service incorporating a three-way collaboration between scientists, hobbyists, and AI, we believe that this multi-criteria proposed analysis of the service provides subtle nuances in the interpretation of results and new perspectives for impact reduction. Assessing these areas remains underrepresented in the literature, and few methodologies are available [30].

## References

- [1] ADEME. Base Impacts: Environmental Impact Database, 2025. Accessed May 2025.
- [2] ADEME. Référentiel par catégorie de produit (rcp) des services numériques. <https://librairie.ademe.fr/industrie-et-production-durable/6022-referentiel-par-categorie-de-produit-rcp-des-services-numeriques.html>, 2025.
- [3] Andrae Anders and Pagnozzi Marcello. Environmental engineering (EE); methodology for environmental life cycle assessment (LCA) of information and communication technology (ICT) goods, networks and services. Technical Report RES/EE-EEPS14, ETSI, 12 2014. [https://www.etsi.org/deliver/etsi\\_es/203100\\_203199/203199/01.03.00\\_50/es\\_203199v010300m.pdf](https://www.etsi.org/deliver/etsi_es/203100_203199/203199/01.03.00_50/es_203199v010300m.pdf).
- [4] ARCEP. Enquête annuelle "pour un numérique soutenable" – édition 2025 (données 2023), 2025. Accessed May 2025.
- [5] Daniel Balouek, Alexandra Carpen Amarie, Ghislain Charrier, Frédéric Desprez, Emmanuel Jeannot, Emmanuel Jeanvoine, Adrien Lèbre, David Margery, Nicolas Niclausse, Lucas Nussbaum, Olivier Richard, Christian Pérez, Flavien Quesnel, Cyril Rohr, and Luc Sarzyniec. Adding virtualization capabilities to the Grid'5000 testbed. In Ivan I. Ivanov, Marten van Sinderen, Frank Leymann, and Tony Shan, editors, Cloud Computing and Services Science, volume 367 of Communications in Computer and Information Science, pages 3–20. Springer International Publishing, 2013.
- [6] Nesrine Bannour, Sahar Ghannay, Aurélie Névéol, and Anne-Laure 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, pages 11–21, 2021.
- [7] Adrien Berthelot. L’empreinte environnementale complète d’un usage numérique : contribution à l’analyse de cycle de vie de services numériques. Theses, Ecole normale supérieure de lyon - ENS LYON, November 2024. <https://theses.hal.science/tel-04874694>.
- [8] Adrien Berthelot, Eddy Caron, Romain De Laage, Laurent Lefèvre, and Alexis Nicolas. Fine-grained methodology to assess environmental impact of a set of digital services, 2024. *Preprint*: <https://hal.science/hal-04928998/>.
- [9] Adrien Berthelot, Eddy Caron, Mathilde Jay, and Laurent Lefèvre. Estimating the environmental impact of generative-ai services using an lca-based methodology. Procedia CIRP, 122:707–712, 2024.
- [10] Anders Bjørn, Mikołaj Owsianiak, Christine Molin, and Alexis Laurent. Main Characteristics of LCA, pages 9–16. Springer International Publishing, Cham, 2018. [https://doi.org/10.1007/978-3-319-56475-3\\_2](https://doi.org/10.1007/978-3-319-56475-3_2).
- [11] Boavizta. <https://github.com/Boavizta/boaviztapi>, 2025. Accessed May 2025.
- [12] Frederic Bordage, Lorraine de Montenay, Sofia Benqassem, Julie Delmas-Orgelet, Firmin Domon, Damien Prunel, Caroline Vateau, and E Lees Perasso. Digital technologies in europe: an environmental life cycle approach. The Greens/EFA, Brussels, 2021.

- [13] BUREAU VERITAS, APL - datacenter, DDemain, and ADEME. Référentiel méthodologique d'évaluation environnementale des services numériques, July 2021. <https://librairie.ademe.fr/produire-autrement/6022-referentiel-methodologique-d-evaluation-environnementale-des-services-numeriques.html>.
- [14] Joseph Cook, Romain Jacob, Jo Lindsay Walton, Adrien Berthelot, Asim Hussain, and Daniel Schien. Beyond counting carbon: Ai environmental assessments struggle to inform net impact decisions. Preprint, November 2025. <https://research-information.bris.ac.uk/en/publications/beyond-counting-carbon-ai-environmental-assessments-struggle-to-i/>.
- [15] Mine Ercan, Jens Malmodin, Pernilla Bergmark, Emma Kimfalk, and Ellinor Nilsson. Life cycle assessment of a smartphone. In ICT for Sustainability 2016, pages 124–133. Atlantis Press, 2016.
- [16] Rolf Frischknecht, Franziska Wyss, Sybille Büsser Knöpfel, Thomas Lützkendorf, and Maria Balouktsi. Cumulative energy demand in LCA: the energy harvested approach. The International Journal of Life Cycle Assessment, 20(7):957–969, 2015.
- [17] Udit Gupta, Young Geun Kim, Sylvia Lee, Jordan Tse, Hsien-Hsin S Lee, Gu-Yeon Wei, David Brooks, and Carole-Jean Wu. Chasing carbon: The elusive environmental footprint of computing. In 2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA), pages 854–867. IEEE, 2021.
- [18] Robert G. Hunt, William E. Franklin, and R. G. Hunt. LCA — how it came about: — personal reflections on the origin and the development of LCA in the USA. The International Journal of Life Cycle Assessment, 1(1):4–7, 1996.
- [19] IDRIS - CNRS. Jean zay supercomputer. <https://www.idris.fr/eng/jean-zay/>, 2019. Accessed: 2025-06-22.
- [20] ISO/TC 207/SC 5. Environmental management — life cycle assessment — principles and framework. Standard ISO 14040:2006, International Organization for Standardization, Geneva, CH, 07 2006. <https://www.iso.org/standard/37456.html>.
- [21] ISO/TC 207/SC 5. Environmental management — life cycle assessment — requirements and guidelines. Standard ISO 14044:2006, International Organization for Standardization, Geneva, CH, 07 2006. <https://www.iso.org/standard/38498.html>.
- [22] ITU. ITU 11410 : Methodology for environmental life cycle assessments of information and communication technology goods, networks and services, 2014. [https://www.itu.int/rec/dologin\\_pub.asp?lang=f&id=T-REC-L.1410-201412-I!!PDF-E&type=items](https://www.itu.int/rec/dologin_pub.asp?lang=f&id=T-REC-L.1410-201412-I!!PDF-E&type=items).
- [23] Alexis Joly, Hervé Goëau, Hervé Glotin, Concetto Spampinato, Pierre Bonnet, Vittorio Vignoli, Robert Planqué, Salvatore Palazzo, and Henning Müller. Interactive plant identification based on social image data. Ecological Informatics, 23:22–34, 2014.
- [24] Simon Kemp. Digital 2022: Global overview report. <https://datareportal.com/reports/digital-2022-global-overview-report>, 2022. Accessed: 2025-05-12.
- [25] Alexandra Sasha Luccioni, Sylvain Viguier, and Anne-Laure Ligozat. Estimating the carbon footprint of bloom, a 176b parameter language model. Journal of Machine Learning Research, 24(253):1–15, 2023.

- [26] Clément Morand, Anne-Laure Ligozat, and Aurélie Névéol. Mlca: a tool for machine learning life cycle assessment. In 2024 10th International Conference on ICT for Sustainability (ICT4S), pages 227–238. IEEE, 2024.
- [27] Aina Rasoldier, Jacques Combaz, Alain Girault, Kevin Marquet, and Sophie Quinton. How realistic are claims about the benefits of using digital technologies for GHG emissions mitigation? In Eighth Workshop on Computing within Limits 2022, page 14. PubPub, 2022. Backup Publisher: LIMITS.
- [28] Sensor Tower. State of mobile 2025. <https://sensortower.com/state-of-mobile-2025>, 2025. Accessed: 2025-06-22.
- [29] Thibault Simon, David Ekchajzer, Adrien Berthelot, Eric Fourboul, Samuel Rince, and Romain Rouvoy. BoaviztAPI: a bottom-up model to assess the environmental impacts of cloud services. In HotCarbon’24. Workshop on Sustainable Computer Systems, Santa Cruz, United States, July 2024.
- [30] Irene Solaiman, Zeerak Talat, William Agnew, Lama Ahmad, Dylan Baker, Su Lin Blodgett, Canyu Chen, Hal Daumé III, Jesse Dodge, Isabella Duan, et al. Evaluating the social impact of generative ai systems in systems and society. arXiv preprint arXiv:2306.05949, 2023.
- [31] Thomas F Stocker, Dahe Qin, Gian-Kasper Plattner, M Tignor, Simon K Allen, Judith Boschung, Alexander Nauels, Yu Xia, Vincent Bex, and Pauline M Midgley. Climate change 2013: The physical science basis. Technical report, Intergovernmental Panel on Climate Change, Working Group I Contribution to the IPCC Fifth Assessment Report (AR5)(Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.), 2013. 1535 pp.
- [32] Laurant van Oers, Jeroen B. Guinée, and Reinout Heijungs. Abiotic resource depletion potentials (ADPs) for elements revisited—updating ultimate reserve estimates and introducing time series for production data. The International Journal of Life Cycle Assessment, 25(2):294–308, 2020.
- [33] Ari Winbush, Daniel McDuff, John Hernandez, Andrew Barakat, Allen Jiang, Conor Heneghan, Benjamin W. Nelson, and Nicholas B. Allen. Smartphone use in a large us adult population: Temporal associations between objective measures of usage and mental well-being. Proceedings of the National Academy of Sciences, 122(43):e2427311122, 2025.
- [34] Bichen Wu, Chenfeng Xu, Xiaoliang Dai, Alvin Wan, Peizhao Zhang, Zhicheng Yan, Masayoshi Tomizuka, Joseph Gonzalez, Kurt Keutzer, and Peter Vajda. Visual transformers: Token-based image representation and processing for computer vision, 2020.