Environmental assessment of projects involving AI methods

Laurent Lefèvre, Anne-Laure Ligozat, Denis Trystram, Sylvain Bouveret, Aurélie Bugeau, Jacques Combaz, Emmanuelle Frenoux, Gaël Guennebaud, Julien Lefèvre, Jean-Philippe Nicolaï, et al.

Proposal for a framework document

Environmental assessment of projects involving AI methods

Coordination for EcoInfo: Laurent Lefèvre (LIP, INRIA, CNRS), Anne-Laure Ligozat (Université Paris-Saclay, CNRS, ENSIIE, LISN), Denis Trystram (Université Grenoble Alpes, Grenoble INP, CNRS, Inria)
Contributions: Sylvain Bouveret (Univ. Grenoble-Alpes, CNRS, Grenoble INP, LIG), Aurélie Bugeau (Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI), Jacques Combaz (Univ. Grenoble Alpes, CNRS, Grenoble INP, VERIMAG), Emmanuelle Frenoux (Université Paris-Saclay, CNRS, Polytech Paris-Saclay, LISN), Gaël Guennebaud (Inria Bordeaux), Julien Lefèvre (Aix Marseille Univ., CNRS, INT, Inst Neurosci Timone), Jean-Philippe Nicolai (GAEL, Grenoble INP)
Translation to English: Karin Dassas and Anne-Laure Ligozat, rewritten from DeepL
Contact: anne-laure.ligozat@lisn.upsaclay.fr
Version: 15 March 2022

Introduction

The objective of this document is to provide criteria for assessing the environmental impacts of responses to calls for projects involving Artificial Intelligence (AI) methods. When proposing these criteria, we take into account, in addition to the general impacts of digital services, the specificities of the AI field and in particular of machine learning: impacts of the learning and inference phases, data collection...

This document is organized in two parts. The first part contains the list of questions proposed for respondents to the calls for proposals. In order to allow respondents to answer these questions in the most relevant way possible, and to allow evaluators to judge the relevance of these answers as well as possible and take them into account in their decision, we have written an explanatory note in the second part to accompany the questions asked to respondents.

This document was written in response to an explicit request from the Ministry of Ecological Transition to the EcoInfo collective.

1. Questions for respondents

In the following, we will refer to a proposal as a response to the call for proposals. We will try to evaluate the environmental impacts of the digital part of this proposal, which we will call *service*. We will focus on the impact of artificial intelligence methods.

In the following, respondents can answer N/A if a criterion is not eligible for their proposal.

Impacts of digital equipment in the proposal

In this section, we want to take into account the digital equipment used during the design of the service and during its use, both by the respondent, the service provider and the users of the service.

Embodied energy and end of life of digital equipment

Make an inventory of the digital equipment used to implement the proposal, taking into account the phases indicated in the scoping document. In the "End of Life Code" column, indicate, according to the code in the second table below, how the digital equipment used in the creation and then in the deployment of the proposal will be managed and, more specifically, what will happen at the end of life of the equipment.

Embodied energy

Type of digital equipment	Number of equipments	Estimated period of use (years)	Manufacturing GHG footprint according to Ecodiag (purchase/flow version, kgCO2e)	End of Life Code (see following table)

End of Life Code	1	2	3	4	5	6
	Reuse within the company	Resale for reuse	Resale for recycling	Approved recycling channel (specify)	Donation for reuse (specify)	Other (specify)

Energy in use phase

Consider the different phases of proposal creation and deployment, as indicated in the scoping document: data acquisition, processing, transport and storage, model learning and inference (testing and deployment)

Characterization of the proposal

Is there a forecast on the number of instances of the service sold?
What is the expected growth of this number?

These figures are to be taken into account in the rest of the document.

Take into account the different phases of creation and deployment of the proposal

Type of algorithm. If it is a Deep learning algorithm, specify the expected architecture (example: CNN, Transformer...)	
Number of model parameters (order of magnitude)	
Number of trainings expected over the life of the service (a range can be given)	
Expected number of runs of the trained model when testing and using the service (a range can be specified)	
Source of data used to create and deploy the model	
Type of data processing (cleaning, database management...) or data augmentation algorithm	
Possible data made available (specify which)	

Computation

Estimated number of hours	Type of equipment/ infrastructure (specify cloud provider if applicable)	PUE of the infrastructure if relevant and known	Location of the equipment or infrastructure	Footprint according to Green Algorithms[1] (kgCO2e)

CPU				
GPU or equivalent				

Data

Data type	Quantity of data stored (if creation of a data stream, indicate e.g. expected quantity / year)	Storage duration	Type of equipment/ infrastructure (specify cloud provider if applicable)	PUE of the infrastructure if relevant and known	Location of the equipment or infrastructure	Amount of data transferred / year

Rationale for the proposed method

Are the AI methods well adapted to the considered context? What are the expected societal and technological benefits (performance, quality of service, etc.) and environmental impacts compared to other methods?

What is the environmental resilience of the service, for example in the event of a network outage or a disruption in material or energy supply?

Environmental impacts related to behavioral, economic, or societal changes induced by the proposal

This section should **only be completed for AI services that have environmental benefits as an objective**. These benefits often do not take into account the behavioral, economic, or societal changes induced by the proposal and having environmental consequences.

Specify the reference scenario

The notions of environmental benefits but also of negative environmental impacts are based on the idea of a comparison between two situations: a reference scenario that does not involve AI and a scenario in which AI is deployed. Briefly specify these scenarios.

Reference scenario

Scénario involving IA

Identifying potential impacts

Below is a qualitative list of potential effects that can be expected from the proposal: obsolescence effects, direct or indirect rebound effects, societal transformations, etc. Here the term "potential effects" of the proposal is to be taken in a broad sense: they can be induced by the proposal itself or be related to the technology(ies) in which it is embedded.

Potential impacts

Nature of the effect	Short description of the effect	Types of environmental impact

Prevention of potential impacts

For each impact listed above, specify what, if any, countermeasures are planned to mitigate or reverse the impact and indicate the effectiveness that can be expected from the countermeasure. The term countermeasure is used when the impact is directly affected. Thus, any environmental compensation mechanism is not considered a countermeasure.

Countermeasures

Impact	Countermeasure	Expected efficiency

2. Explanatory note

Impacts of digital equipment in the proposal

The impacts of the digital equipment in the proposal are due to the life cycle of all the equipment needed to create and deploy the proposed project: raw material extraction, manufacturing, transportation and end of life on the one hand (embodied energy), and use on the other hand.

In the case of a proposal involving Artificial Intelligence methods, the respondents can also distinguish the impacts due to the different phases of creation and deployment of the proposed project: acquisition, processing, transport and storage of data, learning of the model and inferences (test and deployment).

The respondents will take into account the calculations and storage that are done on internal equipment or at the customer's site (known hardware), and those that are done on remote equipment, especially the cloud.

The "Embodied Energy" table aims at estimating the impact of raw material extraction, manufacturing, transportation and end of life of all the digital equipment needed for the proposal.

The "Characterization of the proposal" part provides an understanding of the type of service that will be implemented and aims to get an idea of the evolution scenarios of the proposed service. The questions about the expected number of instances of the service sold and the expected growth aim at estimating the sufficiency of a service, which is not only based on the efficiency of a single instance but on the product of the unit efficiency by the global usage.

The "Computation" table aims at estimating the impact that the associated calculations will have, both in the service creation phase (from data acquisition to model testing) and in the deployment phase (eventual learning and inference in the service implemented internally and at the clients'). Schematically, the carbon footprint related to the consumption of dynamic energy for training and inference of a model can be calculated as follows:

- $Etotal \ = \ PUE \ x \, \Sigma \, Eressource$ with Etotal the dynamic energy consumption from computation, PUE (power usage effectiveness) a metric of energy efficiency of the infrastructure when using a data center[1], and Eressource the energy consumption of a resource for the process considered, the resources taken into account being generally the following: CPUs, GPUs et DRAM
- then $CO2eq \ = \ Etotal \ x \ CI$ with CO2eq the carbon footprint and CI the carbon intensity of the electricity, which enables to convert electricity consumption into the corresponding greenhouse gas emissions, and therefore depends on the energy mix of the location where the computations are made and the period in which the program is launched

[1] Generally, the PUE is averaged over a year

The "Data" table aims to determine the carbon footprint related to the storage and transport of the data needed to create and deploy the service.

The "Rationale for the proposed method" section aims to estimate the necessity of the proposed service and to judge the cost/benefit ratio in environmental terms. For example, in the field of deep learning, some works have questioned the very limited performance gains with respect to the computation time and energy deployed for this. You can rely on a short bibliography (5 references maximum) to briefly describe the proposed method in comparison with previous or similar methods.
The question on resilience aims at understanding the capacities of the service to operate in a degraded hardware context (network cuts, power cuts, hardware downgrading).

Environmental impacts related to behavioral, economic, or societal changes induced by the proposal

Some proposals may advance environmental benefits. For example, it is generally considered that intelligent buildings will reduce energy consumption. Most of the time, these environmental benefits are estimated by making the simplifying assumption that the AI-based service only replaces or optimizes a reference service, and that it has no other effect on habits, behaviors, economic balances, or societal dynamics. This section aims to study the effects related to behavioral, economic, or societal changes induced by the proposal, in particular those that claim environmental benefits.

The "Potential impacts" table is intended to list potential effects related to the proposal. As these effects are much more difficult to quantify than those due to the life cycle of the equipment, the objective here is not to make an overall environmental assessment (which would require exhaustive quantification), but rather to qualitatively point out the potential effects. Here the term "potential impacts" of the proposal is to be taken in a broad sense: they can be induced by the proposal itself or be related to the technology or technologies in which it is embedded. Let's take the example of the autonomous vehicle, which uses AI massively to ensure its operation. A "platooning" service can, among other things, reduce the consumption per kilometer of vehicles. The potential impacts concern the platooning service itself, but also those of the autonomous vehicle insofar as platooning requires autonomy. Examples of potential effects of platooning and autonomous vehicles are given below.

Examples of potential impacts

Nature of the effect	Description of the effect	Types of environmental impacts
Obsolescence	accelerated renewal of the vehicle fleet due to the very rapid evolution of the performance of autonomous modes from one vehicle generation to the next	related to the manufacture of vehicles: climate change, resource depletion...
Direct rebound	increase of the distances traveled due to the	climate change

	decrease of the costs of use ("platooning"), the valorization of the travel time (thanks to the activities carried out during the travel: work, rest, leisure...), the greater accessibility (children, old people, and even objects will have access to the car), the self-parking...	(increase in greenhouse gas emissions)
Indirect rebound	the economic gains due to "platooning" are re-spent by users on products or services with a high environmental impact (air travel...)	function of the products/services
Societal changes	mobility leads to changes in travel habits, urban sprawl	climate change (greenhouse gas emissions), loss of biodiversity due to artificialization and sprawl