

Mode d'emploi : Green Algorithms

Raison d'être

Les intentions :

Dans le cadre de l'appel à projet « démonstrateurs d'IA frugale dans les territoires pour la transition écologique », l'Ecolab demande l'utilisation de l'outil Green Algorithms afin de calculer l'empreinte environnementale des solutions d'IA. Cette proposition vise à proposer une solution simple d'utilisation avec des résultats interprétable facilement.

La frugalité d'une solution d'IA est défini par deux facteurs:

- **La frugalité en donnée** : absence d'emploi de données massives et uniquement la donnée nécessaire pour l'entraînement des solutions jusqu'à l'opérationnalité
- **La frugalité en énergie** : l'absence de surentraînement des solutions d'IA (entraînement jusqu'à l'opérationnalité des solutions d'IA dans le cadre de démonstrateurs). Le choix de localisation des centres serveurs, la consommation en énergie lors du calcul des solutions, en bref l'émission d'équivalent CO2 doit être estimée au mieux.

Extrait du cahier des charges:

« **De plus**, le dossier précisera la méthodologie utilisée pour parvenir aux résultats présentés dans le cadre de cette auto-évaluation. La méthodologie de calcul de l'impact de l'entraînement du système d'IA s'appuiera obligatoirement sur l'outil **Green Algorithms (GT)**, Lannelongue et al, <https://www.green-algorithms.org/> et précisera le pays de localisation des serveurs. »

Green Algorithms : Pas à Pas

L'outil en Open source Green Algorithms que nous vous suggérons d'utiliser pour votre calcul d'impact environnemental a été mis en place par :

- Auteurs : **Loïc Lannelongue, Jason Grealey et Michael Inouye**

Vous trouverez la méthodologie ici :

- Green Algorithms: Quantifying the Carbon Footprint of Computation
- Doi: <https://doi.org/10.1002/advs.202100707>

<https://onlinelibrary.wiley.com/doi/epdf/10.1002/advs.202100707>

Lien du site web du calculateur d'empreinte carbone :

<http://calculator.green-algorithms.org/>

Memento

La frugalité en énergie :

Le numérique représente en France près de 2,5% de l'empreinte carbone nationale. Dans un contexte d'usage croissant du numérique et, ici, de l'IA, cette solution de calcul d'impact vise à vous faire vous interroger sur votre approche des solutions d'IA. Les IA seront des facilitateurs et accélérateurs de la transition écologique bien qu'elles aient également un impact sur différents aspects tel que la consommation de différentes ressources, comme l'énergie. L'IA se doit d'être frugale ce qui se traduit par différents procédés :

- Frugalité en données, minimiser les données pour réduire la consommation énergétique.
- Utiliser des outils qui ont une puissance adaptée à ce que l'on calcule, comme par exemple utiliser des CPU à la place de GPU
- Utiliser du fine-tuning.
- Mettre en pratique le principe de parcimonie de codage (dans les solutions de deep learning par exemple).

L'objectif de la démarche, est de faire entrer le critère de frugalité énergétique dans les projets démonstrateurs. Ce critère doit être pris en compte en amont de la construction des projets et réfléchi au cours de son déploiement.

L'écran de Green Algorithms

Details about your algorithm

To understand how each parameter impacts your carbon footprint, check out the formula below and the [methods article](#)

Runtime (HHMM)

Type of cores

Number of cores

Model

Memory available (in GB)

Select the platform used for the computations

Select location

Do you know the real usage factor of your CPU?

☐ Yes ☒ No

Do you know the Power Usage Efficiency (PUE) of your local data centre?

☐ Yes ☒ No

Do you want to use a Pragmatic Scaling Factor?

☐ Yes ☒ No

[Reset](#)

[Change app version](#)



253.64 g CO₂e
Carbon footprint



2.28 kWh
Energy needed



0.28 tree-months
Carbon sequestration



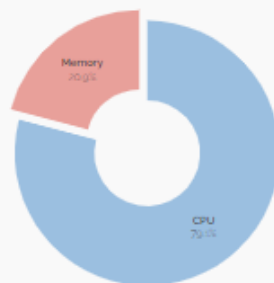
1.45 km
in a passenger car



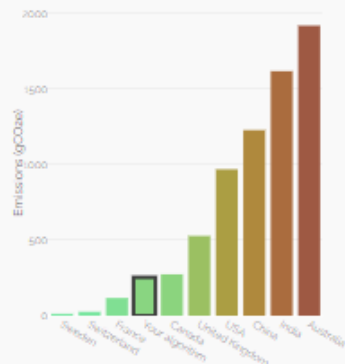
0.51%
of a flight Paris-London

Share your results with [this link!](#)

Computing cores VS Memory



How the location impacts your footprint



Mode d'emploi

Durée d'entraînement de la solution (heure:minute)

Type de processeur (CPU : processeur ; GPU : processeur graphique ; Both : les deux)

Nombre de cœurs

Modèle du processeur

Mémoire disponible (en GB)

Sélectionner la plateforme utilisée pour les calculs (local serveur : serveur local ; personal computer : ordinateur personnel ; cloud computing : calcul cloud)

Sélectionner le continent et pays de localisation des serveurs

Connaissez-vous le facteur d'utilisation réel de votre processeur ? (Oui / Non)

Connaissez-vous le facteur d'utilisation réel de votre processeur graphique ? (Oui / Non)

Connaissez-vous le facteur d'utilisation réel de votre PSF ? (Oui / Non)

Details about your algorithm

To understand how each parameter impacts your carbon footprint, check out the formula below and the [methods article](#)

Runtime (HHMM) 12 0

Type of cores CPU

Number of cores 12

Model Xeon E5-2683 v4

Memory available (in GB) 64

Select the platform used for the computations Local server

Select location Europe Austria

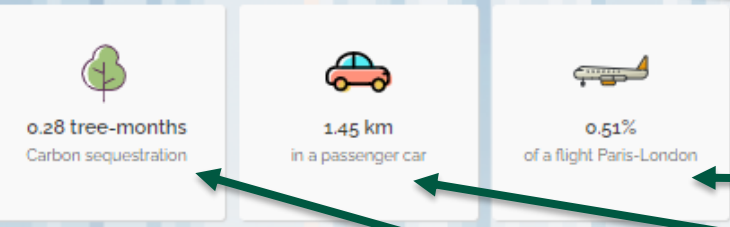
Do you know the real usage factor of your CPU? Yes No

Do you know the Power Usage Efficiency (PUE) of your local data centre? Yes No

Do you want to use a Pragmatic Scaling Factor? Yes No

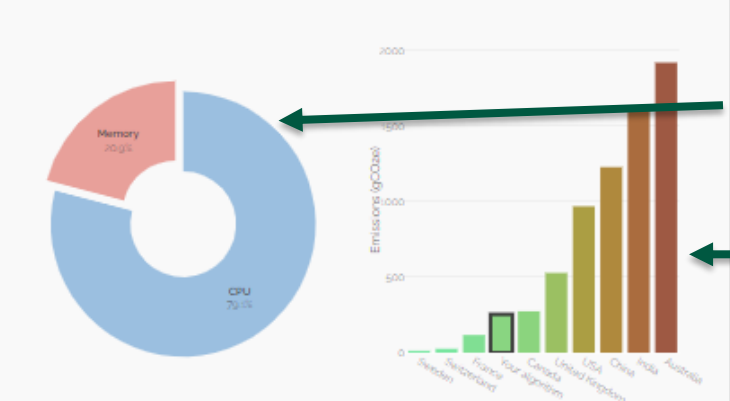
[Reset](#)

[Change app version](#)



Share your results with [this link](#)

Computing cores VS Memory



L'énergie nécessaire en KWh

L'empreinte carbone en Éq. gCO₂

Equivalence de la consommation de votre solution IA

L'équivalent en % d'un vol Paris - Londres

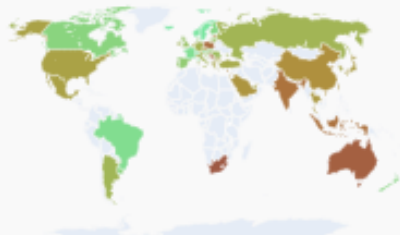
L'équivalent de km parcouru dans une voiture passer

L'équivalent en nombre de mois de la séquestration carbone par un arbre

Noyaux de calcul vs mémoire (répartition des consommation de chaque composante)

L'impact de la localisation des centres serveurs sur l'empreinte carbone (comparaison entre les moyennes pays)

Carbon Intensity across the world



About CO2e

'Carbon dioxide equivalent' (CO2e) measures the global warming potential of a mixture of greenhouse gases. **It represents the quantity of CO2 that would have the same impact on global warming as the mix of interest and is used as a standardised unit to assess the environmental impact of human activities.**

« L'équivalent dioxyde de carbone » (CO2e) mesure le potentiel global de mélange de gaz à effet de serre. Il représente la quantité de CO2 qui aurait le même impact que le mélange d'intérêt et est utilisé comme unité standardisée pour évaluer l'impact environnemental des activités humaines.

What is a tree-month?

It's the amount of CO2 sequestered by a tree in a month. **We use it to measure how long it would take to a mature tree to absorb the CO2 emitted by an algorithm.** We use the value of 11 kg CO2/year, which is roughly 1kg CO2/month.

Qu'est-ce qu'un arbre-mois ?

C'est la quantité de CO2 séquestrée par un arbre en un mois. Nous l'utilisons pour mesurer le temps qu'il faudrait à un arbre mature pour absorber le CO2 émis par un algorithme. Nous utilisons la valeur de 11 kg CO2/an, soit environ 1 kg CO2/mois.

What can you do about it?

The main factor impacting your footprint is the location of your servers: the same algorithm will emit **74 times more CO2e** if ran in Australia compared to Switzerland. Although it's not always the case, many cloud providers offer the option to select a data centre.

Memory power draw is a huge source of waste, because **the energy consumption depends on the memory available, not the actual usage**, only requesting the needed memory is a painless way to reduce greenhouse gas emissions.

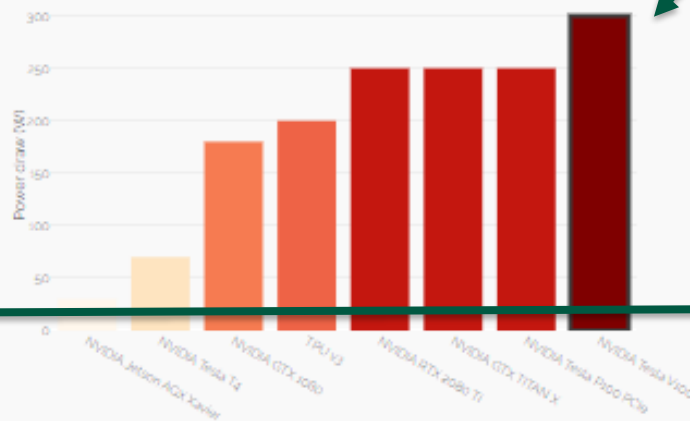
Generally, taking the time to write optimised code that runs faster with fewer resources saves both money and the planet.

And above all, **only run jobs that you need!**

Emission de CO2 dans le monde

Consommation électrique des différents type de modèle de processeur

Power draw of different processors



Que peut on y faire ?

Le principal facteur impactant votre empreinte est l'emplacement de vos serveurs : le même algorithme émettra 74 fois plus de CO2 s'il est exécuté en Australie qu'en Suisse. Bien que ce ne soit pas toujours le cas, de nombreux fournisseurs de cloud offrent la possibilité de sélectionner un data center. La consommation d'énergie de la mémoire est une source de « gaspillage », car la consommation d'énergie dépend de la mémoire disponible, et non de l'utilisation réelle. Demander uniquement la mémoire nécessaire est un moyen de réduire les émissions de GES. Généralement, prendre le temps d'écrire un code optimisé qui s'exécute plus rapidement avec moins de ressources permet de faire des économies.

La formule de calcul de l'impacte environnementale de vos solutions IA

The formula

The carbon footprint is calculated by estimating the energy draw of the algorithm and the carbon intensity of producing this energy at a given location:

$\text{carbon footprint} = \text{energy needed} * \text{carbon intensity}$

Where the energy needed is:

$\text{runtime} * (\text{power draw for cores} * \text{usage} + \text{power draw for memory}) * \text{PUE} * \text{PSF}$

The power draw for the computing cores depends on the model and number of cores, while the memory power draw only depends on the size of memory available. The usage factor corrects for the real core usage (default is 1, i.e. full usage). The PUE (Power Usage Effectiveness) measures how much extra energy is needed to operate the data centre (cooling, lighting etc.). The PSF (Pragmatic Scaling Factor) is used to take into account multiple identical runs (e.g. for testing or optimisation).

The Carbon Intensity depends on the location and the technologies used to produce electricity. But note that the "energy needed" indicated at the top of this page is independent of the location.

La formule

L'empreinte carbone est calculée en estimant la consommation d'énergie de l'algorithme et les émissions carbone des énergies utilisées pour produire de l'électricité dans une endroit donnée :

Empreinte carbone = énergie nécessaire * les émissions carbone

Où l'énergie nécessaire est :

Temps d'exécution * (consommation d'énergie par les cœurs * utilisation + consommation d'énergie pour la mémoire) * PUE (indicateur d'efficacité énergétique) * PSF (facteur d'échelle pragmatique)

La consommation électrique des cœurs de calcul dépend du modèle et du nombre de cœurs, tandis que la consommation électrique de la mémoire dépend uniquement de la taille de la mémoire disponible. Le facteur d'utilisation réelle du cœur (la valeur par défaut est 1, c'est-à-dire l'utilisation complète). L'indicateur d'efficacité énergétique (PUE) mesure la quantité d'énergie supplémentaire nécessaire pour faire fonctionner le data center (refroidissement, éclairage, etc.). Le PSF (facteur d'échelle pragmatique) est utilisé pour prendre en compte plusieurs exécutions identiques (par exemple pour les tests ou pour l'optimisation). L'intensité carbone dépend du lieu et des technologies utilisées pour produire de l'électricité.

Foire aux questions de Green Algorithms

Q1 Doit on indiquer le nombre de processeurs, le nombre de cœur ou le nombre de threads utilisés (unité d'exécution ? unité de traitement ?)

R1 En ce qui concerne l'unité centrale/processeur, le nombre de cœurs (les unités centrales ont généralement entre 4 et 12 cœurs par processeur). Pour ce qui est de l'unité de traitement graphique indiqué le nombre d'unité graphique. Si vous utilisez le multithreading (ou multifil) sur les processeurs (c'est-à-dire que vous utilisez plus de threads que de cœurs), indiquez tout de même le nombre de cœurs, mais sachez que vos émissions risquent d'être sous-estimées.

Q2 Que faire si mon processeur ne figure pas dans la liste des propositions ?

R2 Vous pouvez sélectionner « other » (pour autre) et trouver la valeur TDP (Thermal Design Power ou « enveloppe thermique ») sur le site du fabricant. Vous pouvez ajouter un commentaire sur la page de résolution des problèmes afin que les équipes de Green Algorithms puissent ajouter votre processeur à la liste.

<https://github.com/GreenAlgorithms/green-algorithms-tool/issues/1>

Q3 Que faire si mon pays ne figure pas dans la liste ?

R3 Rendez-vous sur la page de résolution des problèmes numéro 2 et ajouter un commentaire sur le pays de localisation (certains pays sont plus réservés quant au partage de leur bouquet énergétique). En parallèle, vous pouvez utiliser la moyenne mondiale, ou une approximation proche pour vos estimations.

<https://github.com/GreenAlgorithms/green-algorithms-tool/issues/2>

Foire aux questions de Green Algorithms

Q4 Puis-je comparer l'impact des algorithmes indépendamment du lieu ?

R4 Oui, il suffit d'utiliser « Energy Needed » (énergie nécessaire) en Watt affichée à côté de « carbone missions » (émissions carbone).

Q5 Comment puis-je connaître le facteur d'utilisation de mes processeurs ?

R5 Cela dépend de votre système. Par exemple, si vous utilisez SLURM, seff <job_id> vous donnera « CPU efficiency » (l'efficacité du processeur). Des commandes similaires existent pour les différents systèmes, et si vous ne les trouvez pas, vous pouvez simplement laisser la valeur par défaut de 1.

Q6 Comment estimer mon PSF « Pragmatic scaling factor » (le facteur d'échelle pragmatique) ?

R6 Essayez d'estimer le nombre de fois où vous devez effectuer votre analyse complète pour obtenir des résultats qui vous conviennent. Il peut s'agir d'essais et d'erreurs, d'optimisations des paramètres, de problèmes de mémoire etc.

Qu'est-ce que le PSF ?

De nombreuses analyses sont présentée comme une exécution unique d'un outil algorithmique ou d'un logiciel particulier. Cependant, les calculs sont rarement effectués une seule fois. Les algorithmes sont exécutés plusieurs fois, parfois même des centaines de fois, de manière systématique ou manuellement avec différents paramétrages. Les modèles statistiques peuvent comprendre un nombre quelconque de combinaisons de co-variables, de procédures ainsi que d'ajustements etc. Il est important d'inclure ces répétitions dans l'empreinte carbone. Pour prendre en compte le nombre de fois qu'un calcul est effectué en pratique, nous définissons le facteur d'échelle pragmatique (PSF), par ce même facteur les émissions de GES estimées sont multipliées. La valeur et les cause du PSF varient considérablement d'une tâche à l'autre. En apprentissage automatique (machine learning), l'ajustement des hyper paramètres d'un modèle nécessite des centaines, voire des milliers d'exécutions tandis que d'autres outils nécessitent moins d'ajustements et peuvent parfois être exécutés sur une période plus courte, ou encore certains outils peuvent nécessitaient de moins de réglages et peuvent parfois être exécutés un plus petit nombre de fois. Selon les travaux publiés ou l'expérience de l'utilisateur, le PSF doit être estimée pour toute tâche spécifique. Toutefois, dans Green Algorithms nous prévoyons et recommandons que le PSF soit estimée pour toute tâche spécifique et recommandons également que chaque utilisateur estime sa propre PSF.

A noter

Pour faire baisser le facteur d'échelle pragmatique :

- Limiter le nombre d'exécutions de votre algorithme au stricte nécessaire, en effet limiter les exécutions en particulier des solution qui sont les plus gourmandes en énergie est le moyen le plus simple de réduire l'empreinte carbone.
- Limiter la mise au point des paramètres au minimum nécessaire ainsi que d'exécuter des essais à petite échelle pour le débogage.

Prêtez attention à la localisation de vos centres serveurs :

- Lors du calcul de l'impact environnementale de votre solution IA, l'empreinte carbone des centres serveurs prend en compte le PUE (l'indicateur d'efficacité énergétique) qui dépend d'un centre à l'autre. Mais également le mix énergétique du pays de localisation pour la production d'électricité. Un pays avec un mix énergétique décarboné aura une empreinte carbone positive ce qui entraînera une meilleure balance environnementale de votre solution (à noter que Green Algorithms prend uniquement en compte l'empreinte carbone de l'entraînement des solutions et non l'empreinte à amont c'est-à-dire la fabrication des ordinateurs de calcul des centre serveurs).

Balance environnementale :

- Green Algorithms ne prend en compte dans son calcul de l'impact environnemental que l'empreinte carbone. D'autres indicateurs comme l'ACV (analyse du cycle de vie) ne sont quant à eux pas pris en compte.