Empreinte environnementale des modèles de langue

ETAL
septembre 2025

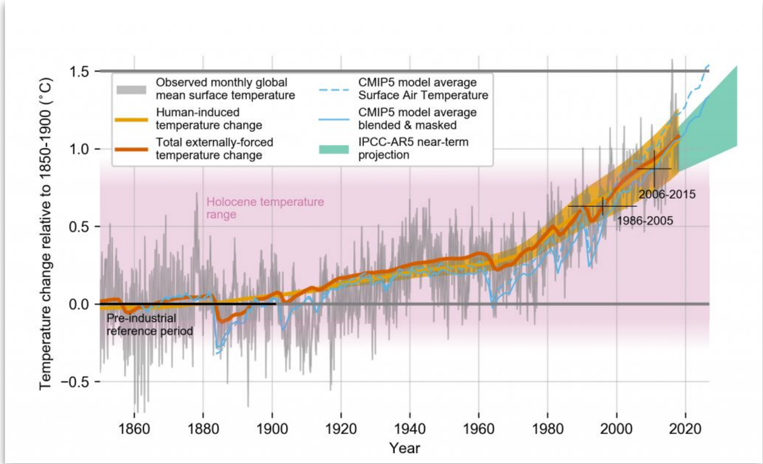
Anne-Laure Ligozat

ensIIE
Ecole Nationale Supérieure d'Informatique pour l'Industrie et l'Entreprise

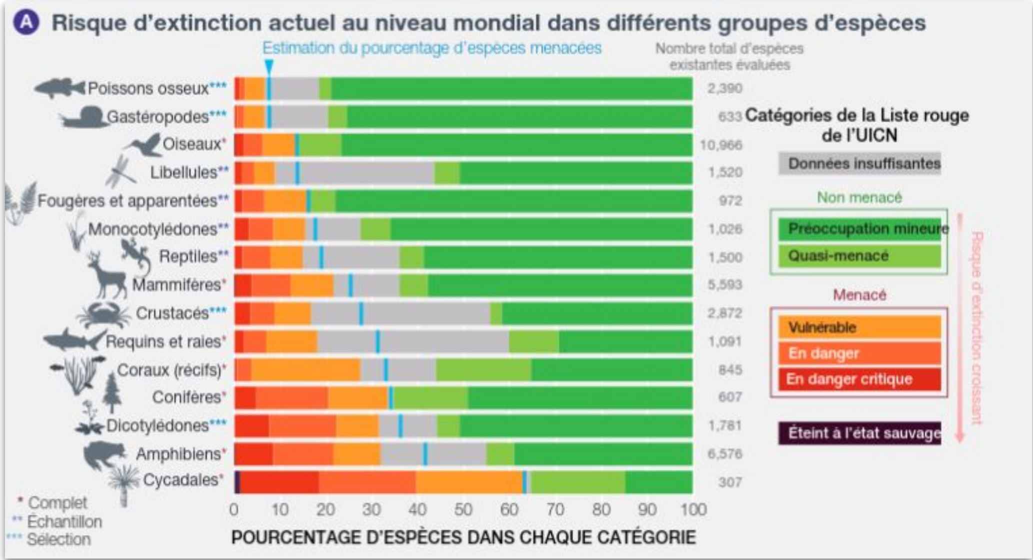
LISN
LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

Context

Environmental context

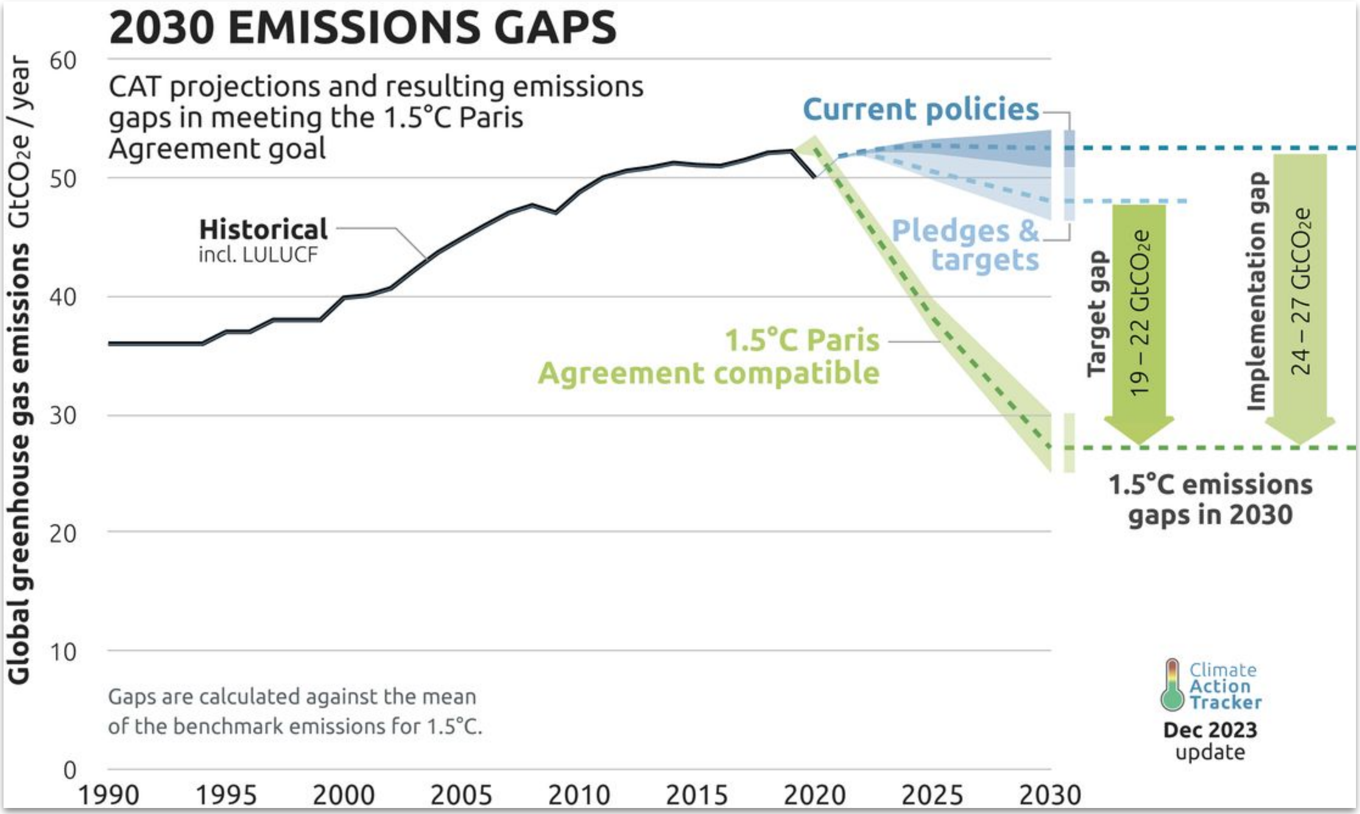


Source: IPCC



Source: IPBES

Environmental context



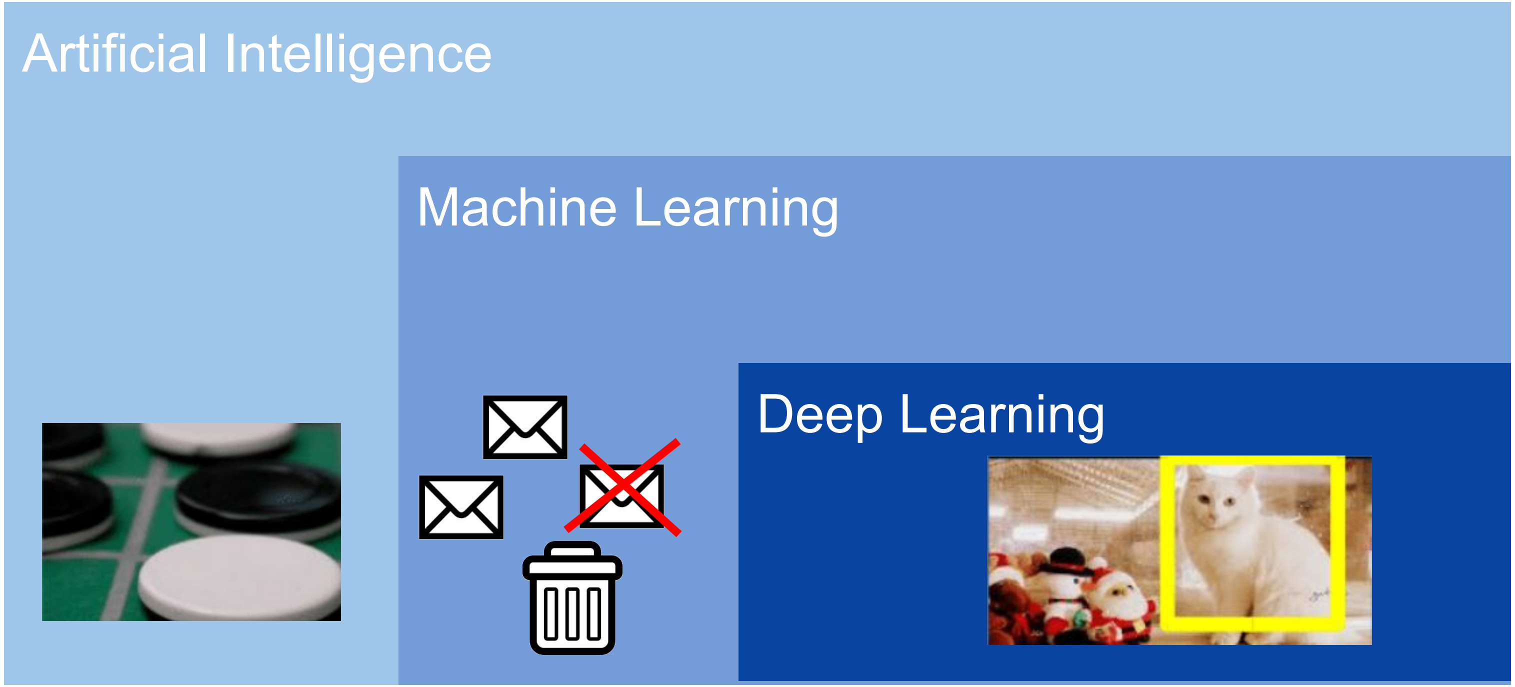
Environmental impacts of NLP?

stays within planetary boundaries?

helps other sectors stay within planetary boundaries?

Is AI itself on a sustainable path?

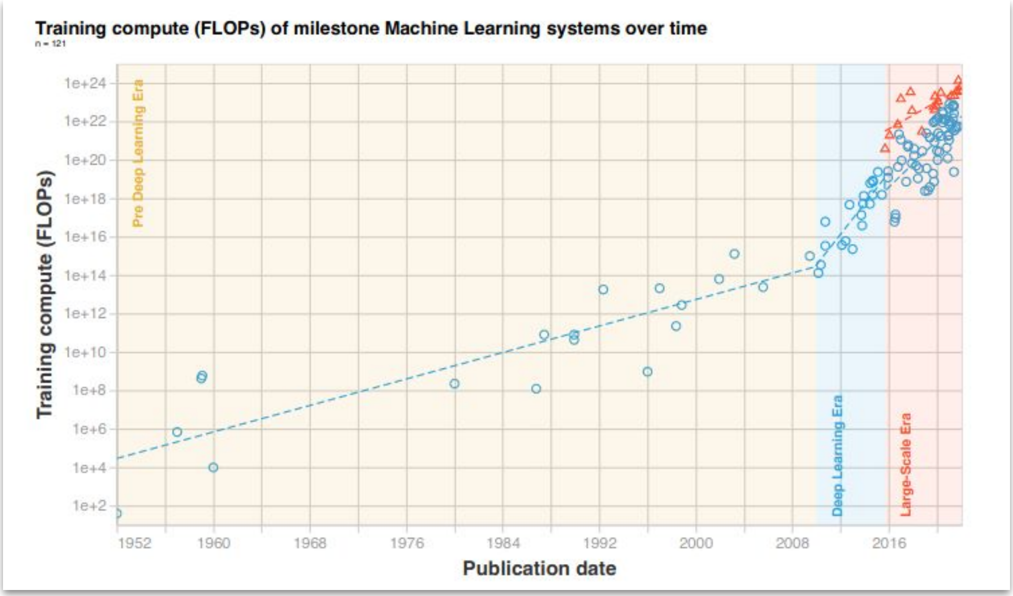
AI?



Potential high environmental impacts

potential high environmental impacts:

- massive **data**
- **computation** demand



(Sevilla et al., 2022)

How can we measure these impacts?

Bottom-up approach

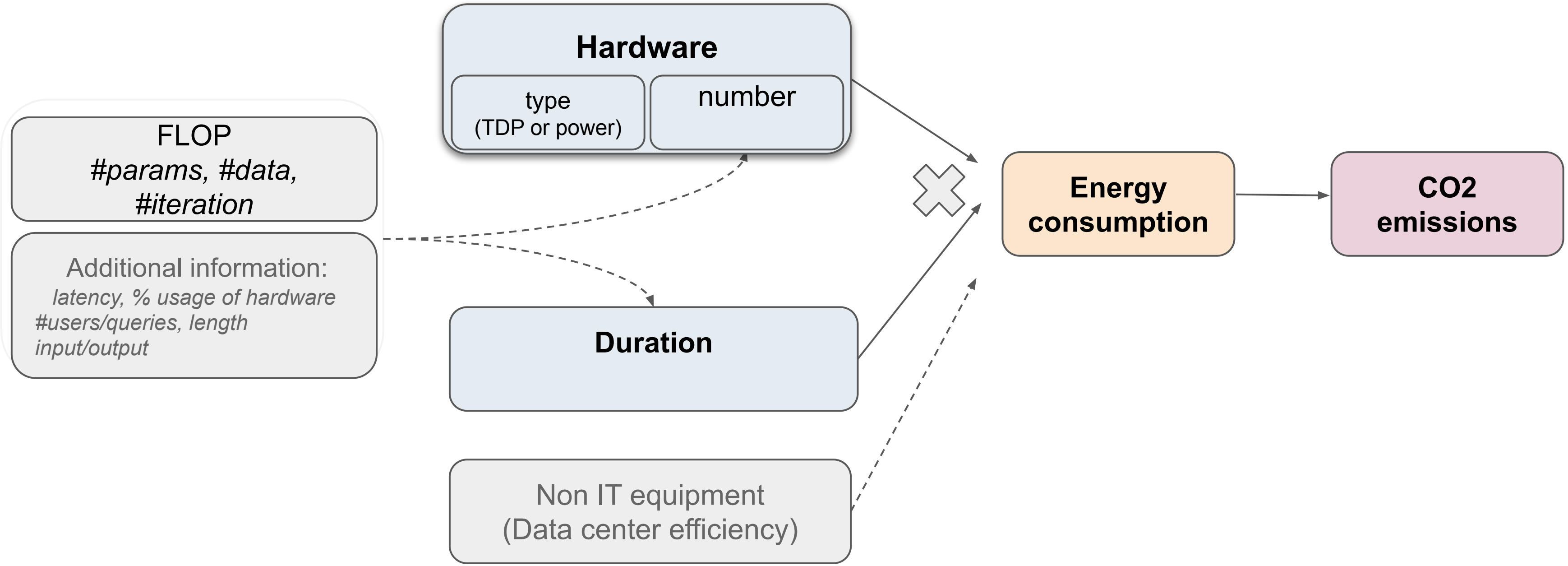
on a server, what is the additional energy use due to the AI program running:

- processor
- GPU
- memory…

$$\Rightarrow \text{footprint}_1 = \sum (\text{use}_{\text{resource}}) \times \text{electricity carbon intensity}$$

kWh → kg CO_2e

Estimating energy consumption

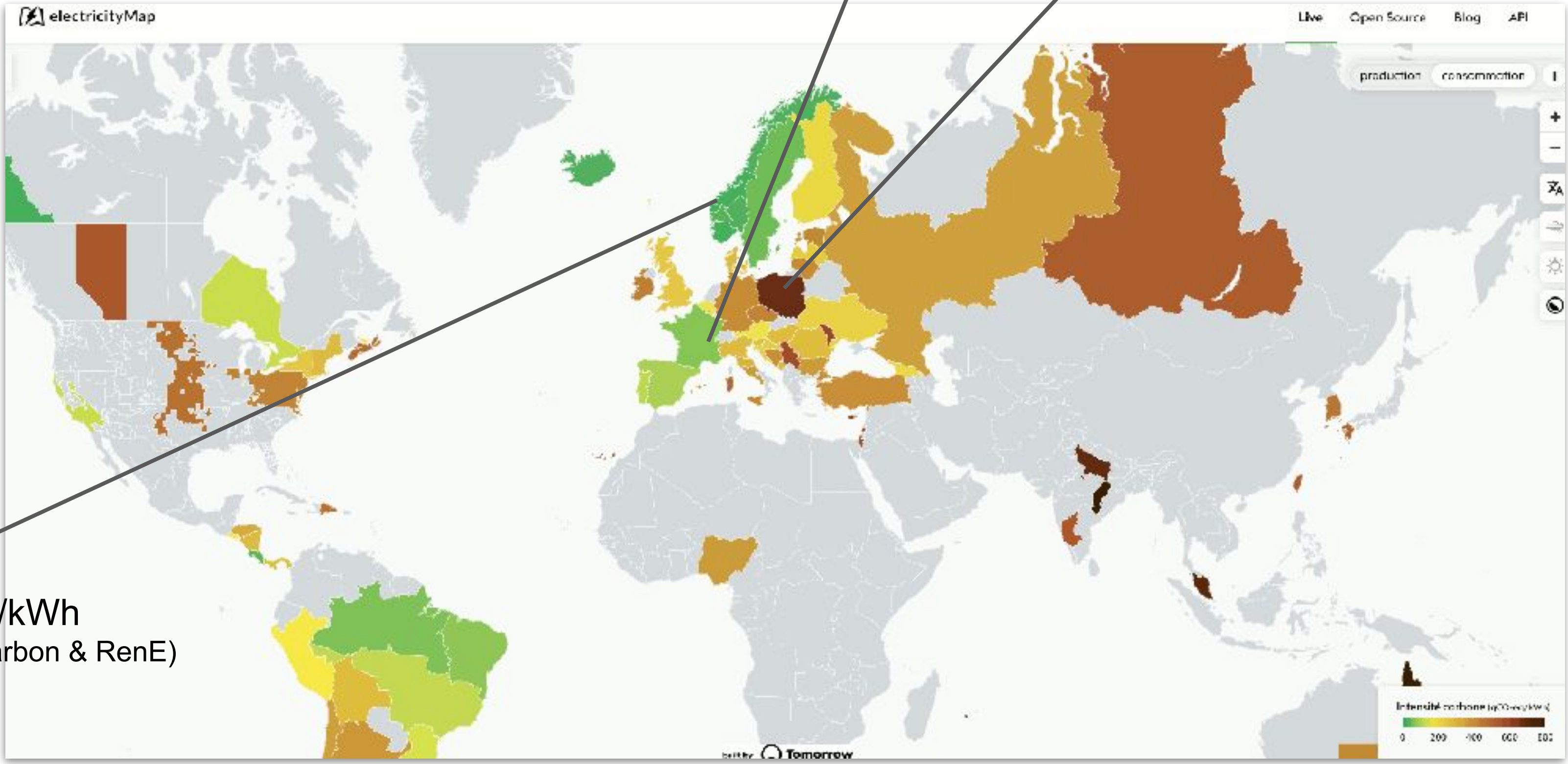


source: A. Bugeau & A.-L. Ligozat

Carbon intensity of electricity

France: 101g CO_2e/kWh
(86% low carbon, 13% RenE)

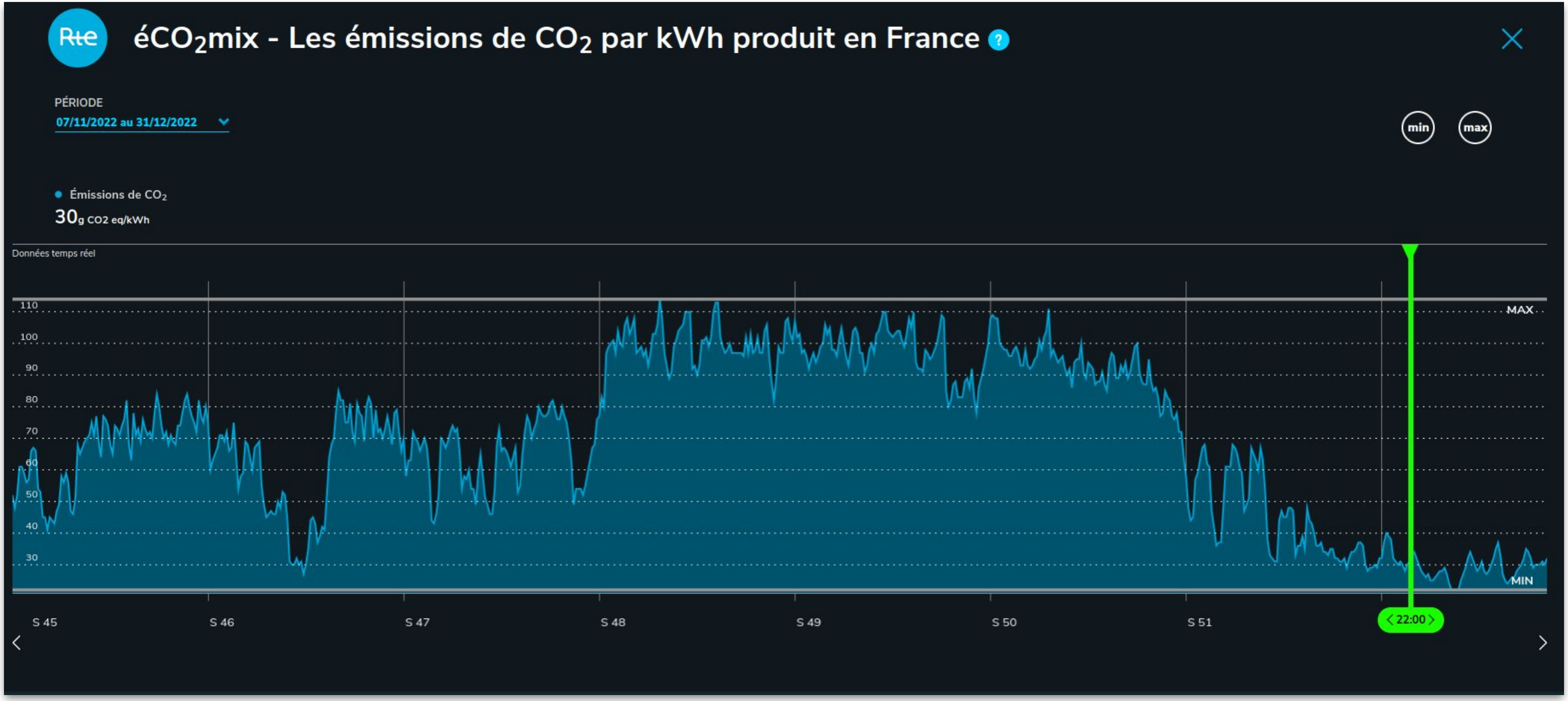
Poland: 927g CO_2e/kWh
(13% low carbon, 13% RenE)



Norway:
22g CO_2e/kWh
(100% low carbon & RenE)

source: [electricityMap]

Temporal evolution of the carbon intensity



Influence of the carbon intensity on the operational carbon footprint

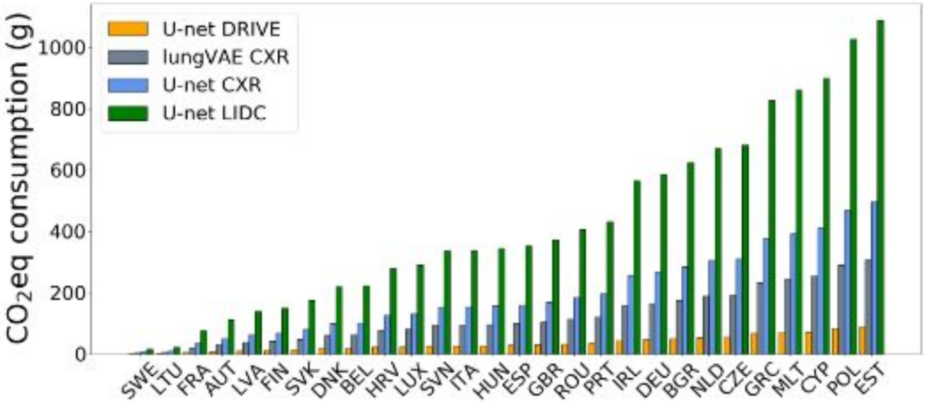


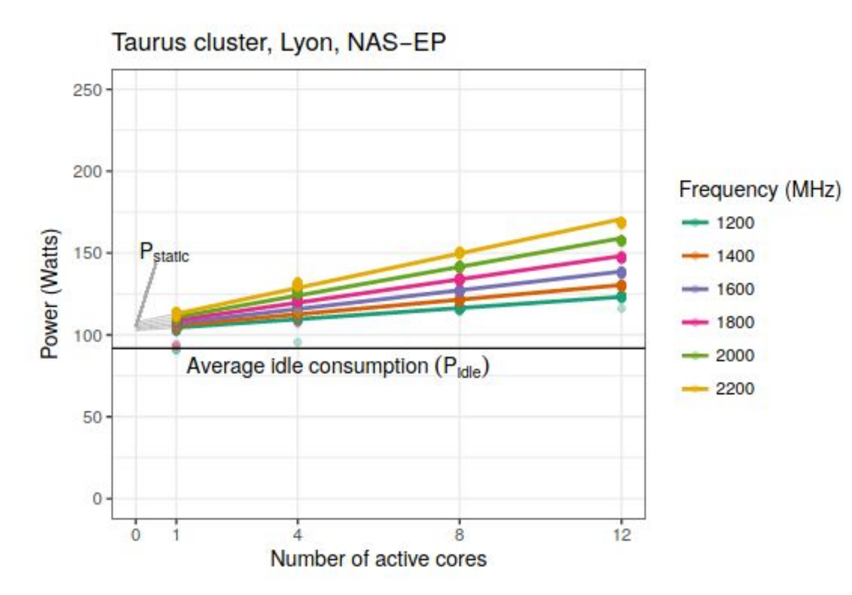
Figure 4. Estimated carbon emissions (gCO_2eq) of training our models (see Appendix B) in different EU-28 countries. The calculations are based on the average carbon intensities from 2016 (see Figure 8 in Appendix).

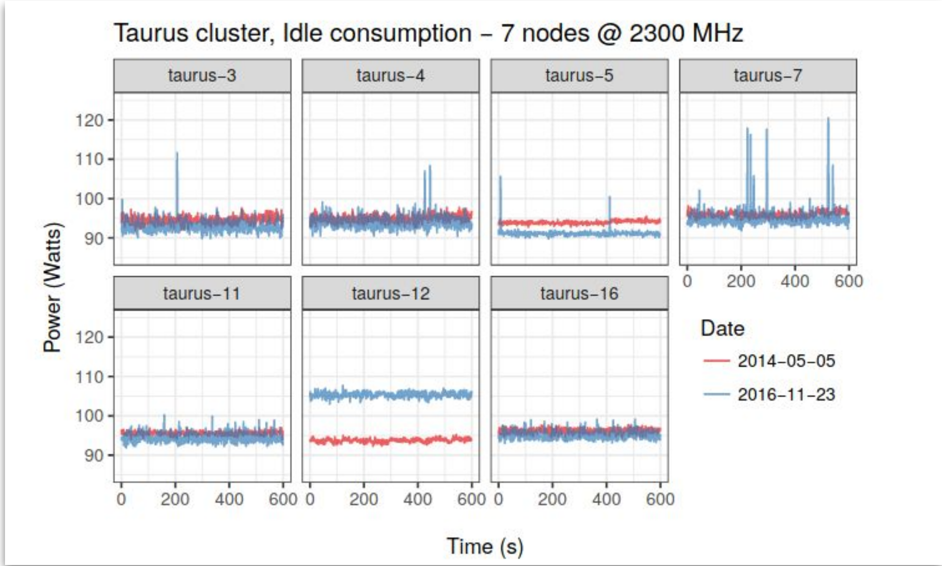
(Anthony et al., 2020)

Serveur energy use

not proportional to the charge

variation in time, with models…





Source: (Heinrich et al., 2017)

How to measure energy use?

hardware

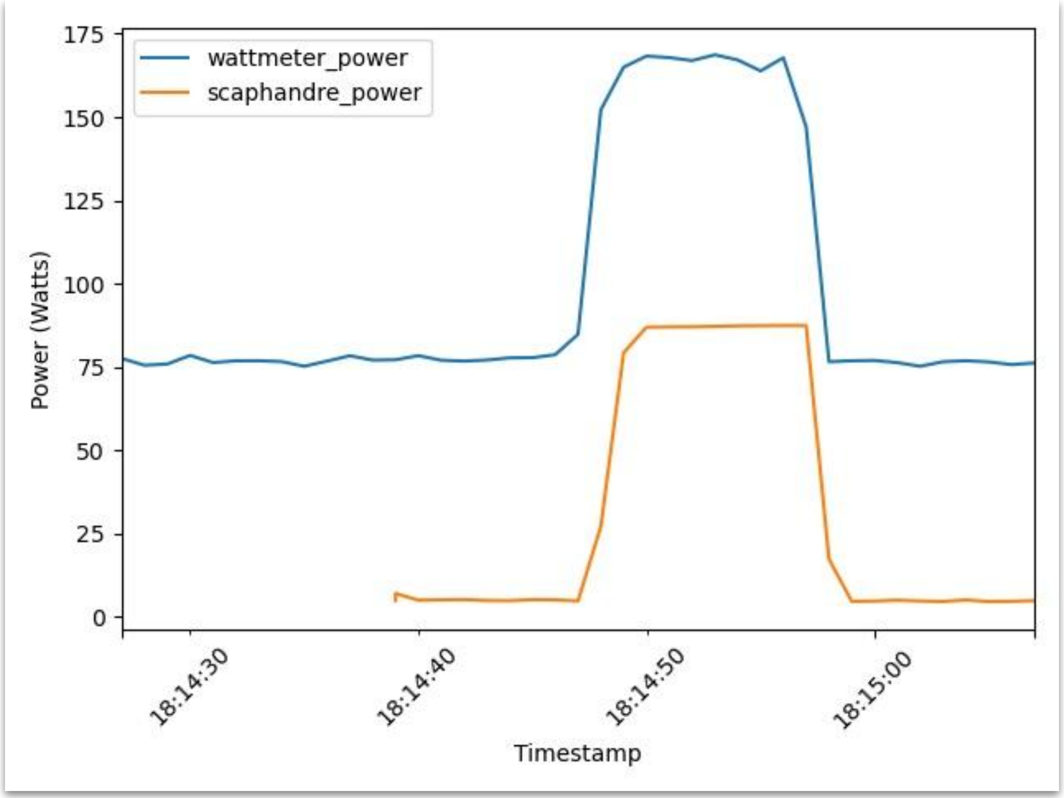
software

CODE CARBON

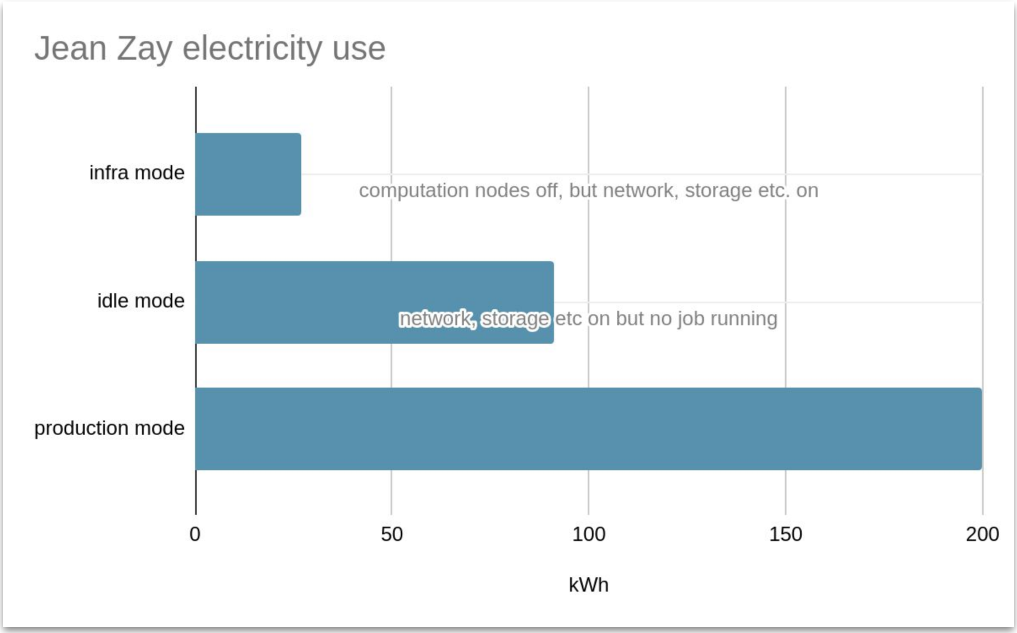
Green Algorithms

Towards environmentally sustainable computational science

Hardware vs software

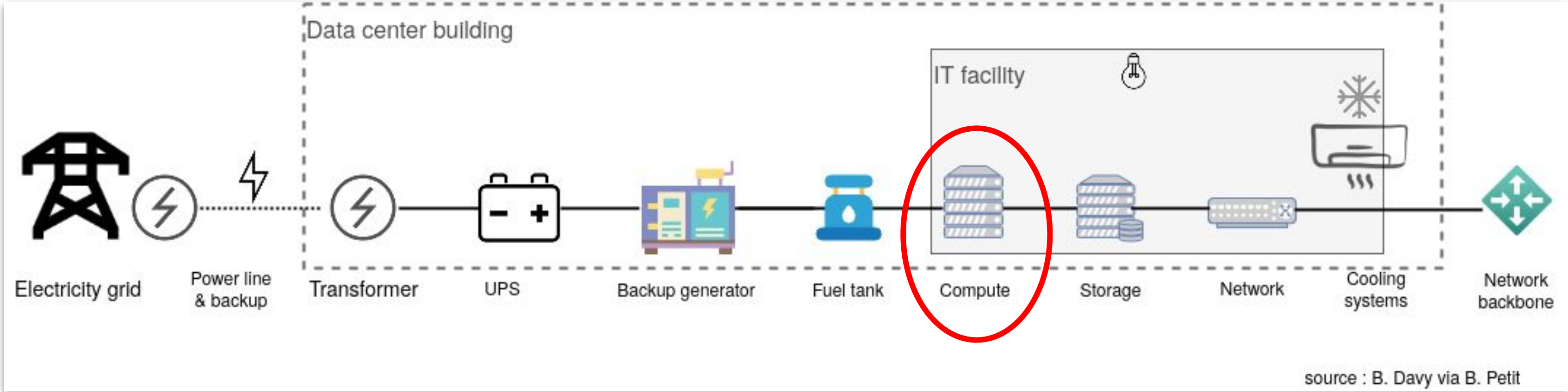


Electricity consumption in Jean Zay

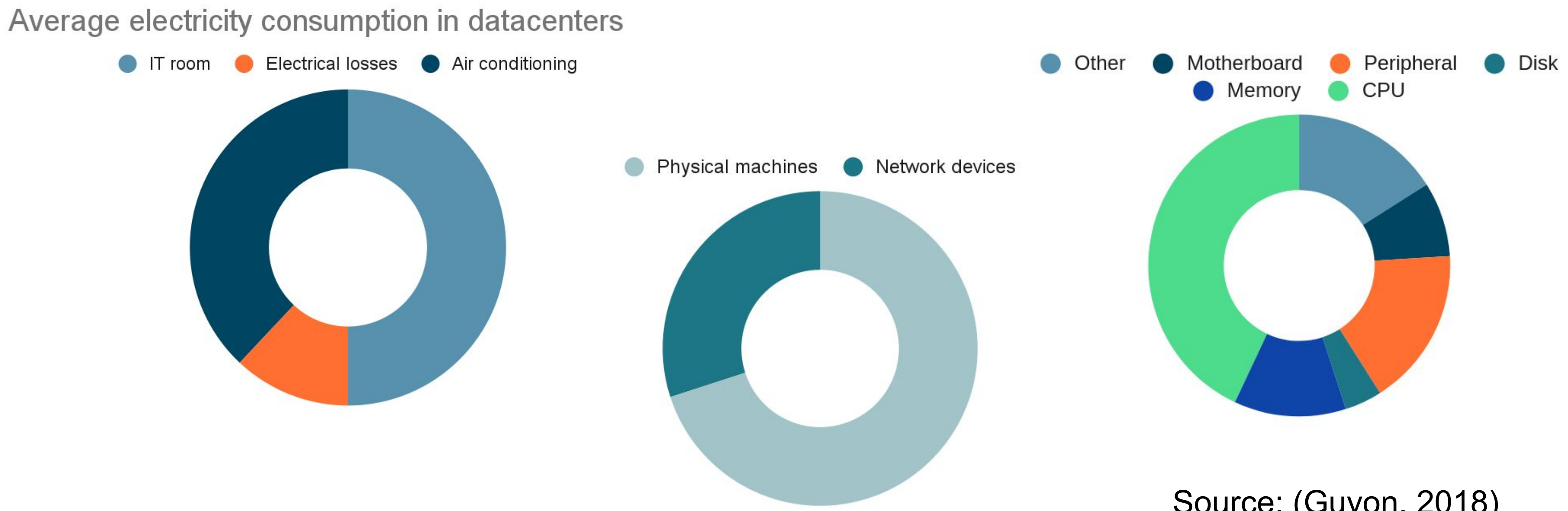


Evaluating the carbon footprint of an AI service

Which equipment?



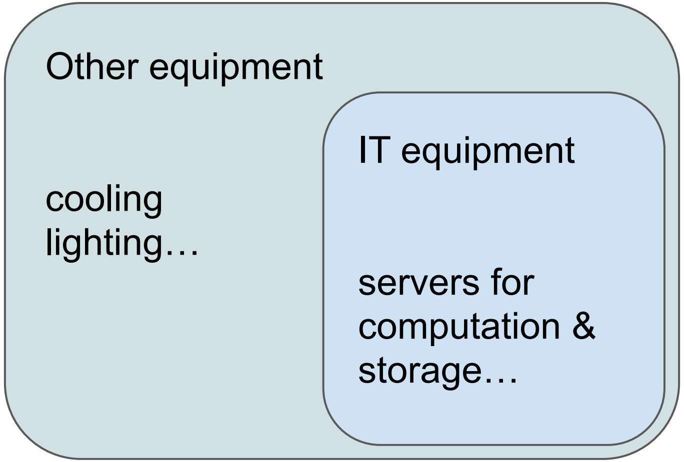
Other energy use



Source: (Guyon, 2018)

Efficiency of the facility

$$\text{PUE} = \frac{\text{total facility energy}}{\text{IT equipment energy}}$$



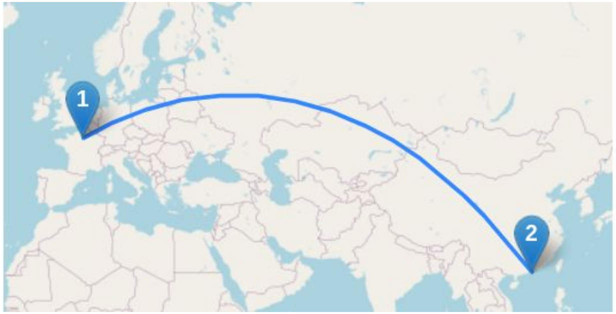
Carbon footprint of NLP models (Strubell et al, 2019)

variety of state-of-the-art NLP models

software-based energy measurement

Training

- 12 hours to several weeks
- emissions: between 18kg CO_2e and 284 t CO_2e
- most used model: 652 kg CO_2e, or
  - one one-way flight from Paris to Hong Kong
  - or 2 500km by car

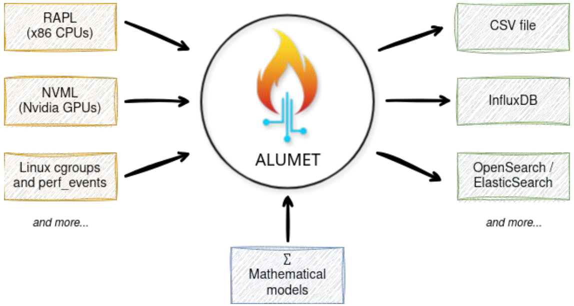


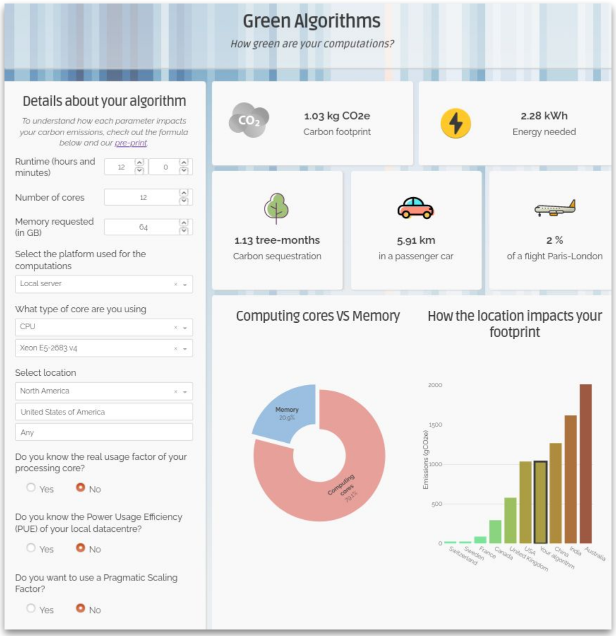
sum GPU time ~ 60 GPU during 6 months

Tools for carbon footprint estimation

Many factors influence the carbon footprint of this phase

- model, data…
- energy efficiency of the data center
- carbon intensity of the electricity





Comparison of several tools

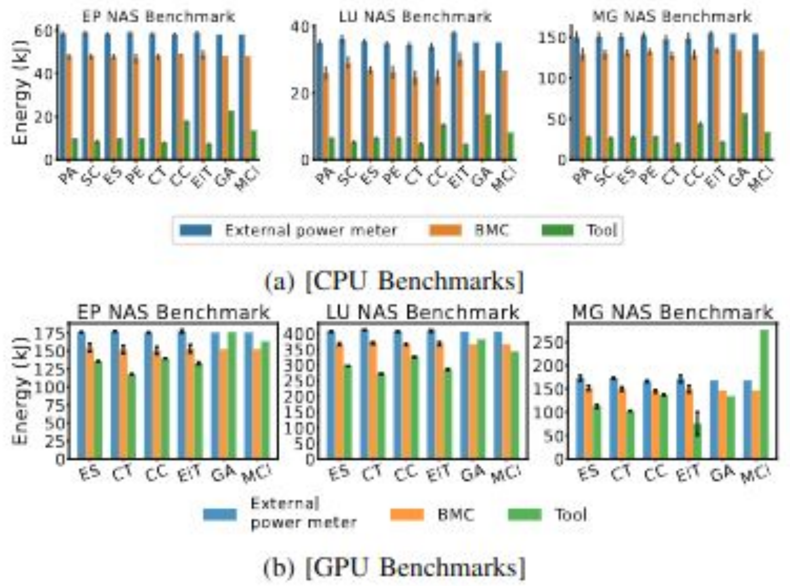
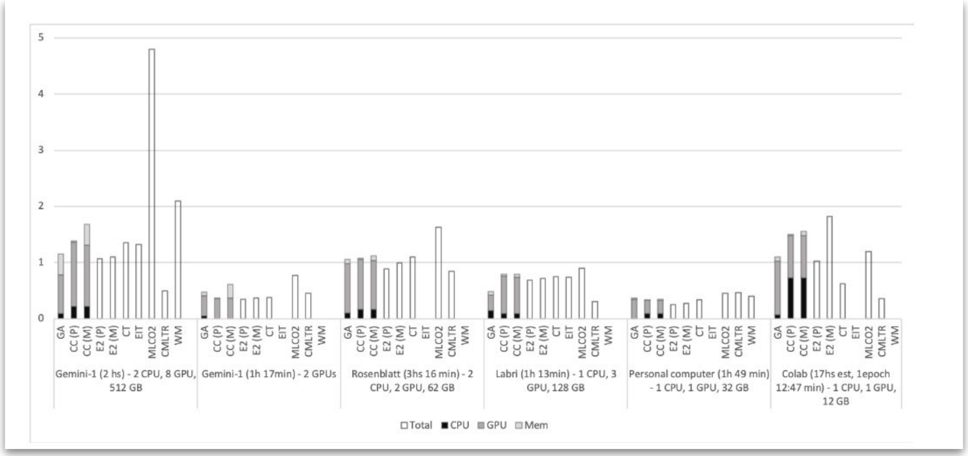


Fig. 2: Total energy consumed by the benchmarks as reported by the power meters. Tools: PowerAPI (PA), Scaphandre (SC), Energy Scope (ES), Perf (PE), Carbon Tracker (CT), Code Carbon (CC), Experiment Impact Tracker (EIT), Green Algorithm (GA), ML CO2 Impact (MCI)

source: (Jay et al., 2023)



source : (Bouza et al., 2023)

Let’s practice!

https://calculator.green-algorithms.org/

La consommation est calculée avec la formule suivante :

$$C_{total} = runtime \times \left(\sum_{c \in cores} (P_c \times usage_c) + P_{memory} \right) \times PUE \times PSF$$

où :

- $runtime$ est la durée d'exécution en $heures$
- P_c est la puissance consommée par un coeur c (CPU ou GPU) en $Watt$. La puissance réellement consommée n'étant pas connue, Green Algorithms utilise le TDP du CPU ou GPU qui est considéré comme une estimation de la puissance moyenne de l'équipement.
- $usage_c$ est le taux d'utilisation du coeur c, entre 0 et 1
- P_{memory} est la puissance consommée par la mémoire
- PUE est l'indicateur d'efficacité électrique du datacenter, supérieur à 1. Le PUE par défaut est le PUE moyen mondial : 1,67.
- PSF est le nombre de fois où cet entraînement a été effectué, par défaut 1

Puis l'empreinte carbone est calculée selon la formule suivante :

$$CarbonFootprint = C_{total} \times CI$$

avec :

- CI le facteur d'émission (*carbon intensity*) de l'électricité de la région considérée

Nous considérerons les paramètres d'entraînement du tableau 1.

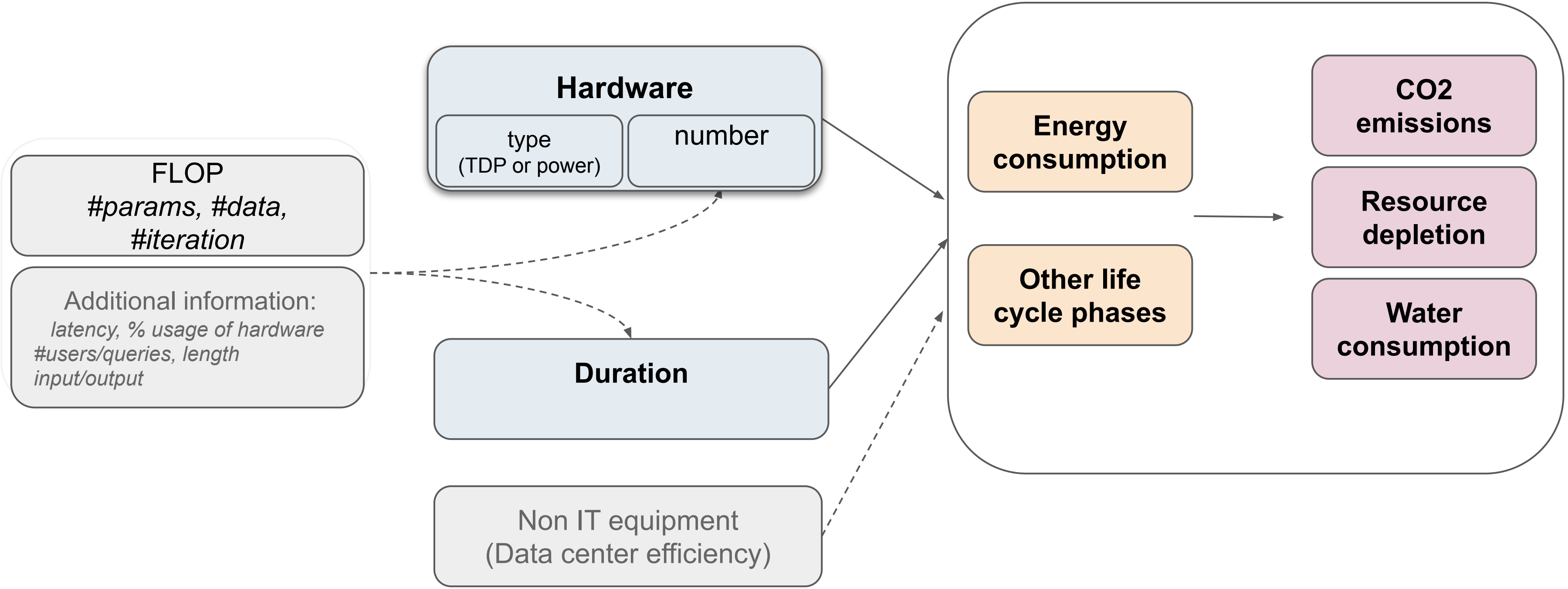
1. Consommation d’énergie et empreinte carbone pour la table 1 ?
2. Expliquez les différences observées en changeant :
   - le PUE
   - cloud/local
   - le type de GPU
   - la localisation

Temps d’exécution	190 h
Nombre de CPU	4
Modèle de CPU	Xeon E5-2683 v4
Nombre de GPU	4
Modèle de GPU	Tesla V100
Mémoire disponible par GPU	32GO
Localisation du serveur	Orsay, France

TABLE 1 – Informations sur l’apprentissage du modèle

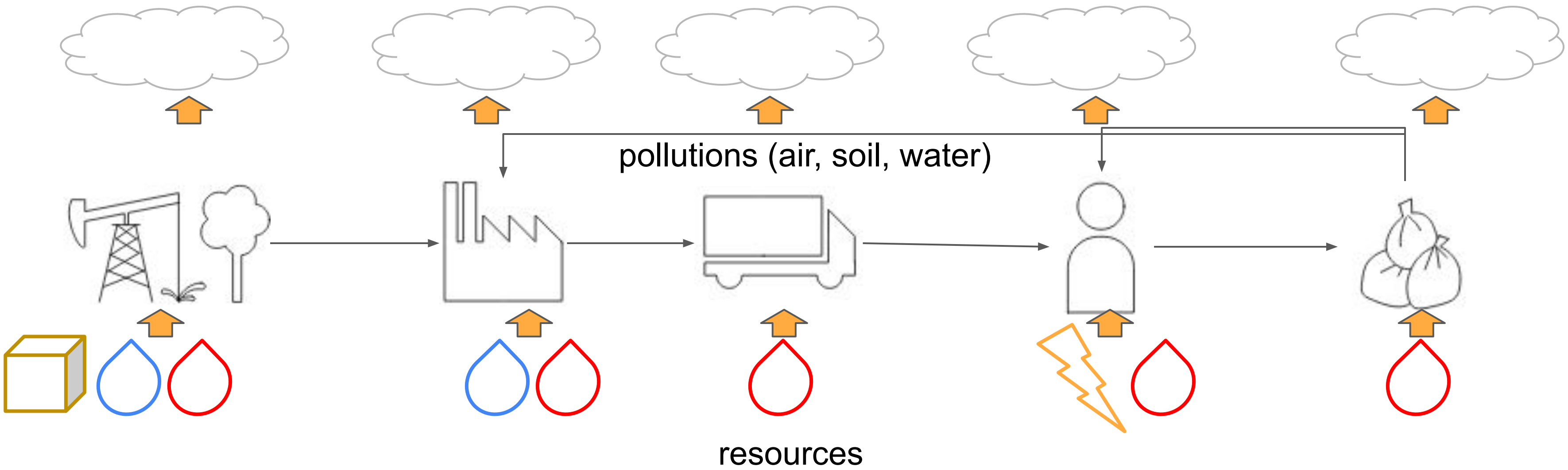
3. Qu’est-ce qui est et n’est pas pris en compte ?
4. Comment évalueriez-vous la part de l’inférence par rapport à l’entraînement ?
5. D’après un article de recherche de Google, les achats d’énergie renouvelable de Google réduisent l’impact à zéro. Qu’en pensez-vous ?

Estimating environmental impacts



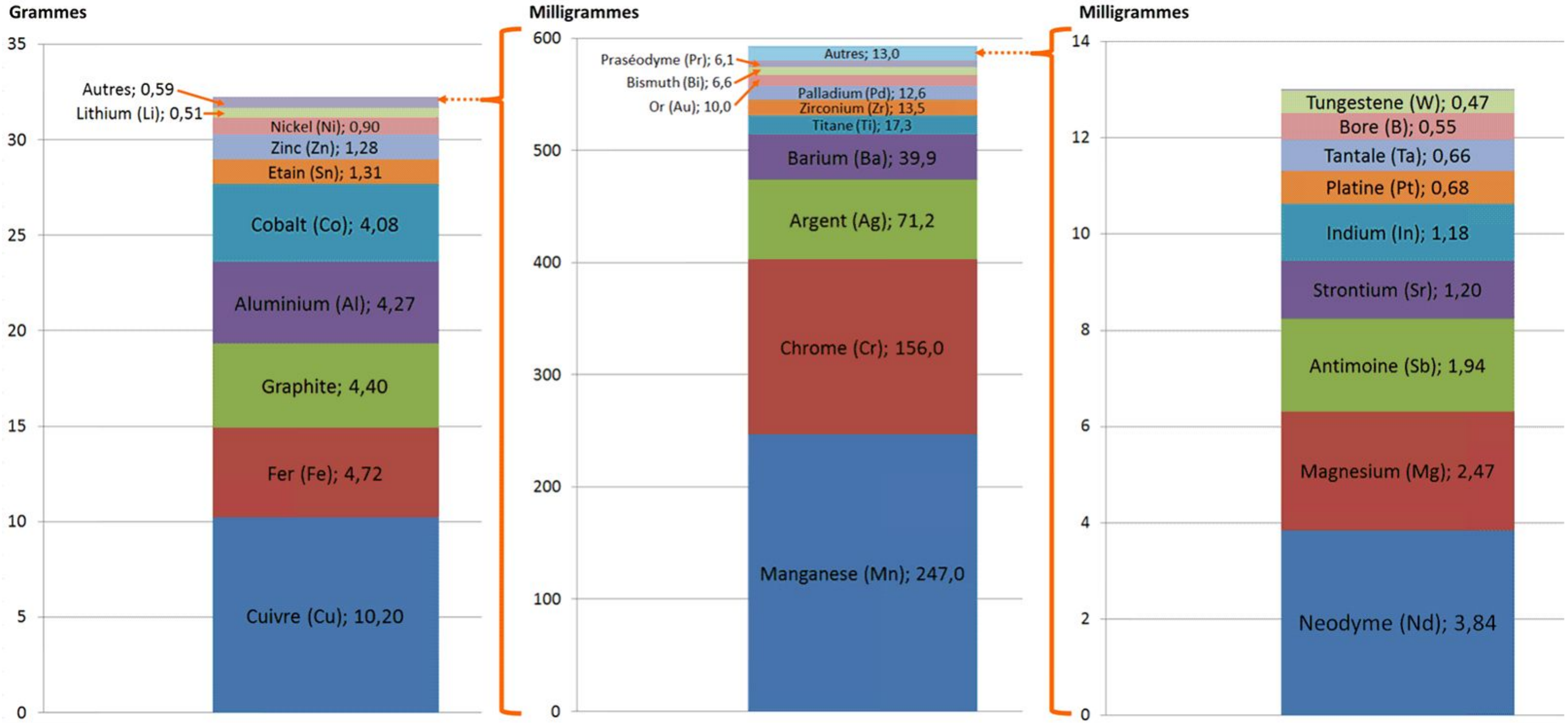
source: A. Bugeau & A.-L. Ligozat

Life Cycle Assessment



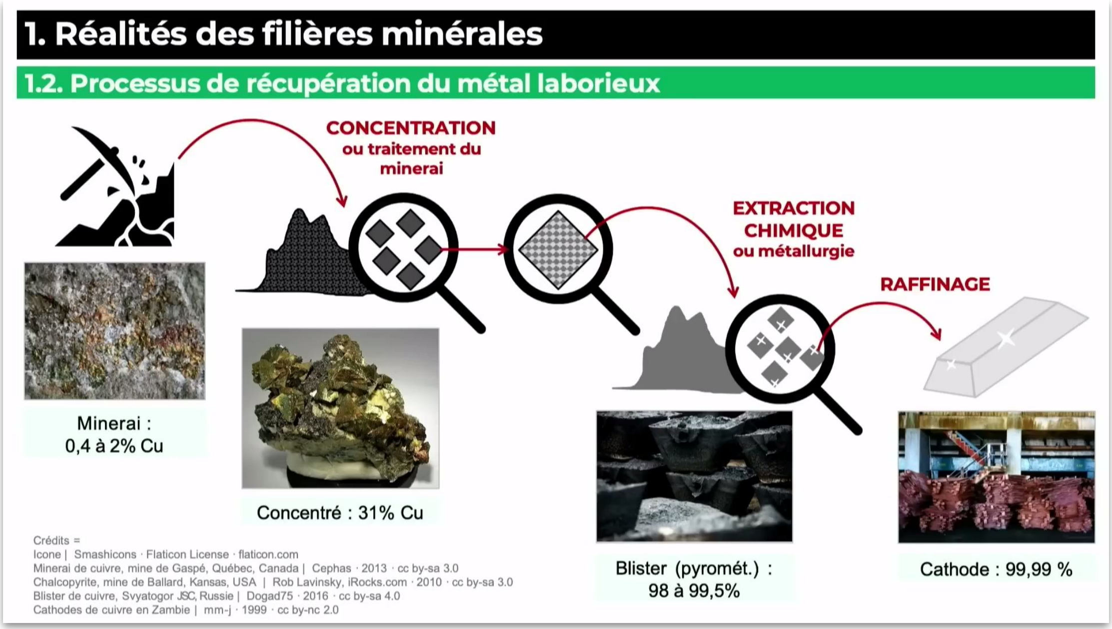
schema from Jacques Combaz

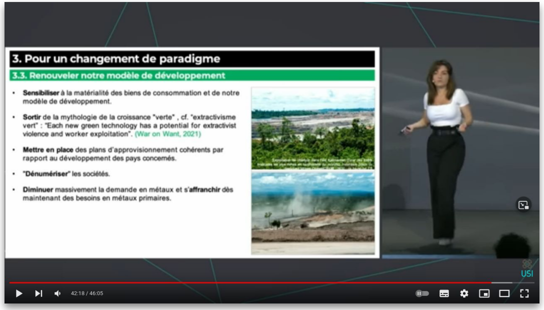
Composition of a smartphone



Source: [report from French Sénat on smartphones]

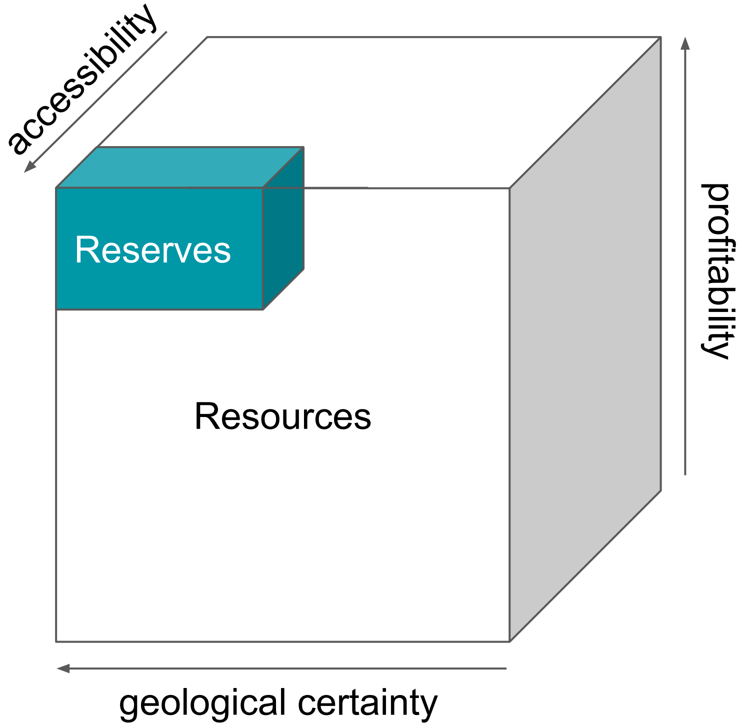
Metal recovery

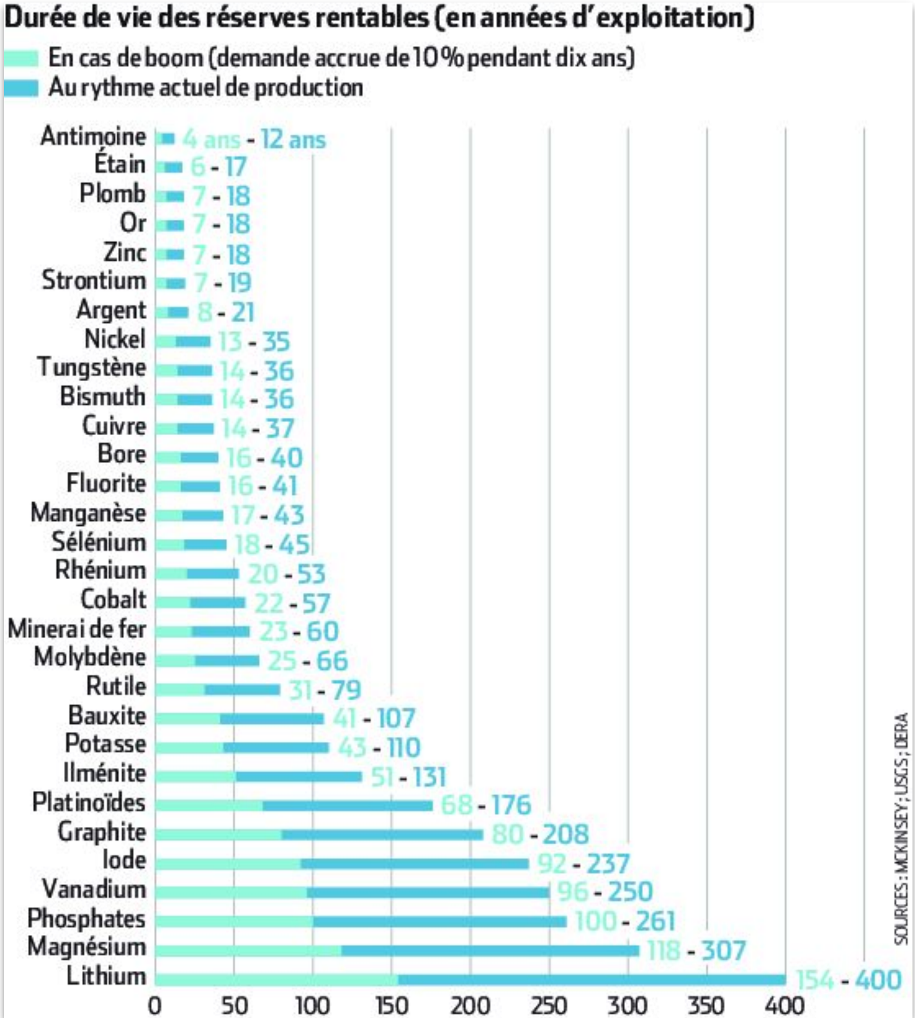




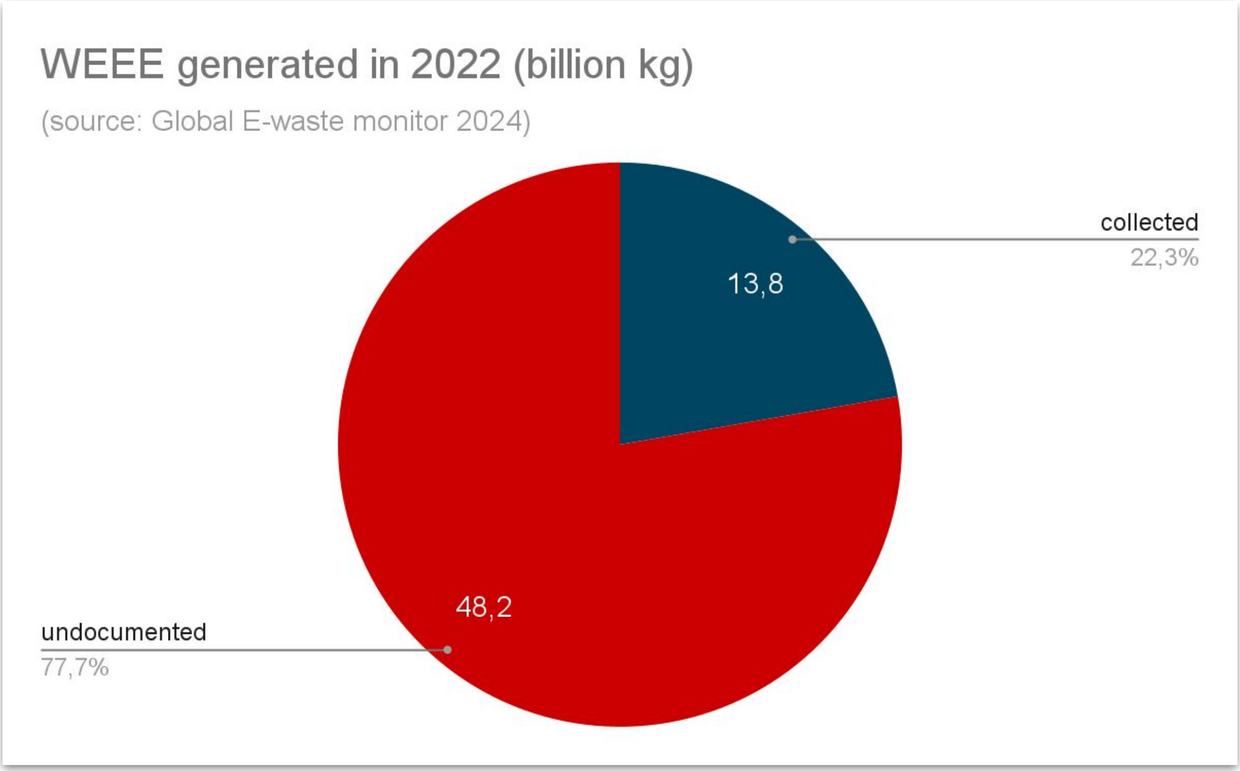
[Ruée minière au XXIè siècle : jusqu'où les limites seront-elles repoussées ?]() - Aurore Stephant at USI

Raw material availability





E-waste



Source: Global E-wasteMonitor 2024

Informal recycling

Dumping and processing of electronic waste in Agbogbloshie, Accra, Ghana

Estimating the environmental footprint for an AI model

For each piece of equipment used:

Use phase impact = Energy consumption x Impact per kWh

(possible allocations too, for example if use of a shared infrastructure)

Embodied impact = Production impact x Allocation factor

(given by manufacturers) (generally time-based)

MLCA: towards comprehensive environmental evaluation

	ADP	GWP	PE	Human toxicity	Water Consumption	...
Extraction	✓	✓	✓	✗	✗	✗
Manufacturing	✓	✓	✓	✗	✗	✗
Transport	✗	✗	✗	✗	✗	✗
Usage	✓	✓	✓	✗	✗	✗
End of Life	✗	✗	✗	✗	✗	✗

Modeling graphics card

Manufacturing impacts attribution

Infrastructure consumption

PARIS 2015

Putting impacts in perspective

(Morand, 2023)

a BigScience initiative

Results for BLOOM training

176B params · 59 languages · Open-access

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

training BLOOM

- GWP: $59tCO_2$ eq
  - annual emissions of 59 person (PB_{GWP})
  - annual emissions of 29 person (SNBC)
- ADP: 1.2 kgSb eq
  - annual resource extraction of 38 person (PB_{ADP})
- PE: 9800000 MJ

LCA of an AI as a service

Stable Diffusion = open-source text-to-image generative deep-learning model

FU1 = client vision i.e., average impact of a person visiting the website

FU2 = host vision i.e., impact of the service for one year

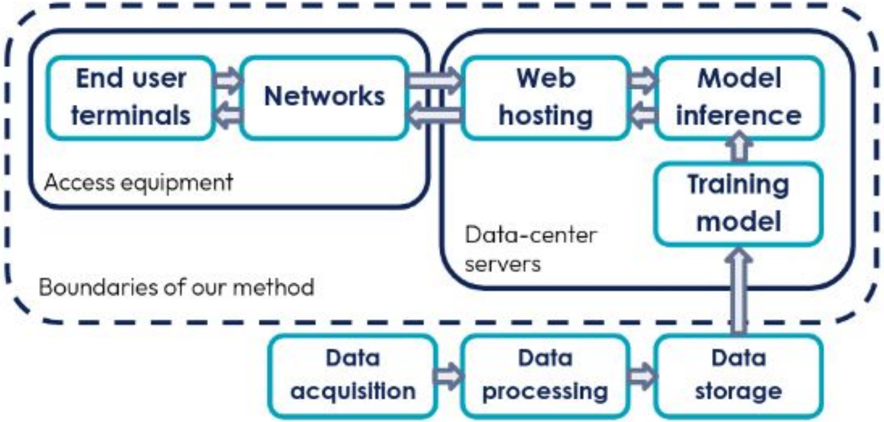


Figure 1: Structure of Gen-AI service considered by the methodology

s

source: (Berthelot et al, 2024)

LCA of an AI as a service

Table 2: Environmental impact of Stable Diffusion for FU1 and FU2

FU	Abiotic Depletion Potential (kgSb eq)	Warming Potential ($kgCO^2$ eq)	Primary energy (MJ)
FU1 - Single use of service	$6.72e^{-08}$	$7.84e^{-03}$	$2.02e^{-01}$
FU2 - A year of service	$4.64e^{+00}$	$3.60e^{+05}$	$8.93e^{+06}$

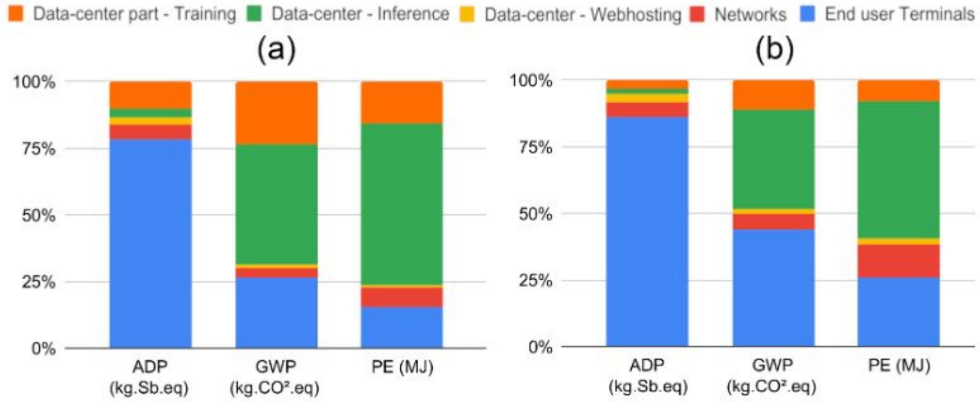
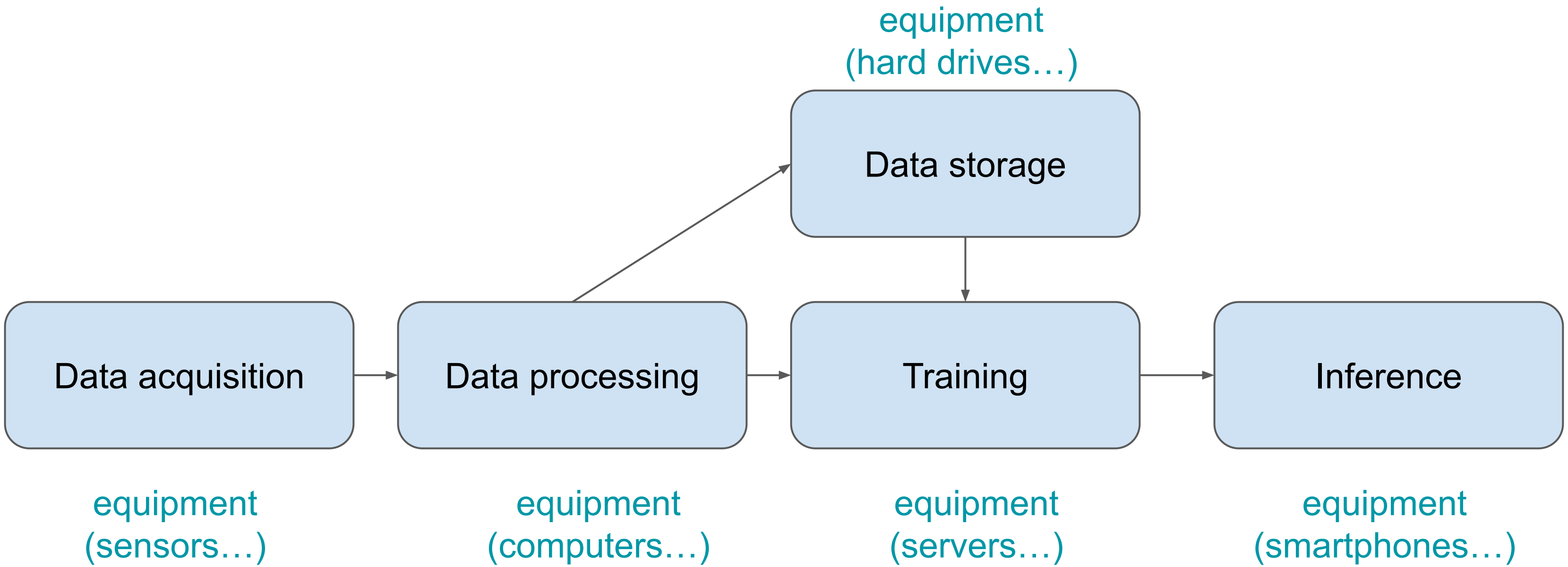


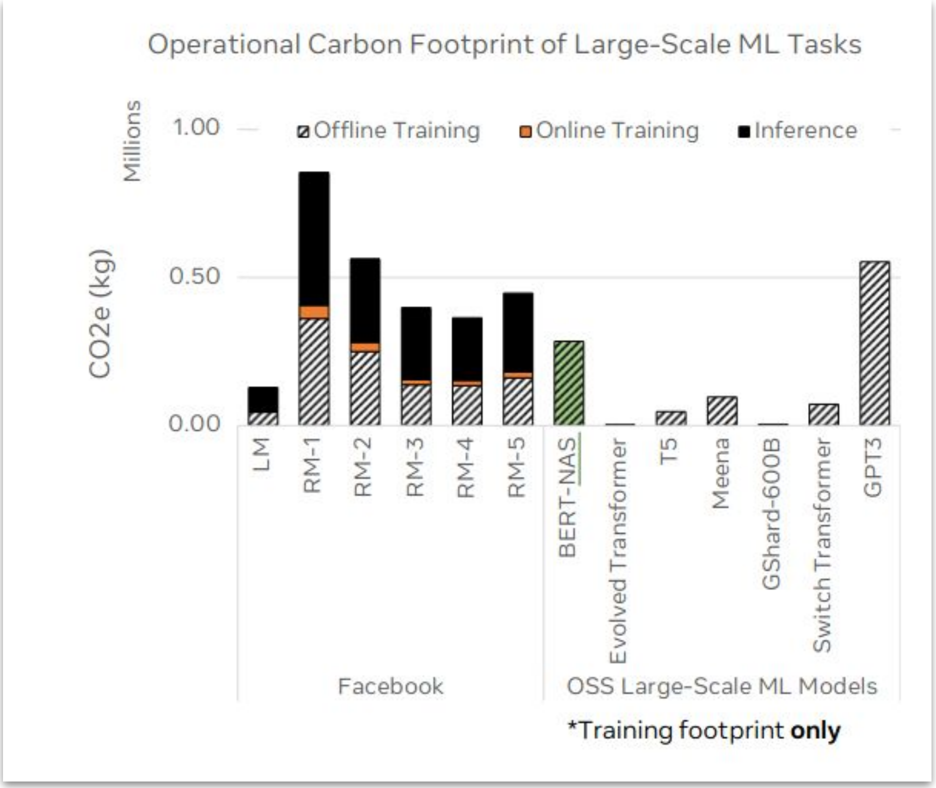
Figure 2: Impact distributions for (a) FU1 and (b) FU2

source: (Berthelot et al, 2024)

AI: which tasks?



Training vs inference (Wu et al., 2021)



Energy consumption of the inference phase (Luccioni et al. 2024)

- 88 models, 10 ML tasks; 1,000 inferences for 3 datasets, experiments repeated 10 times
- Hardware: 1 node of 8 NVIDIA A100-SXM4-80GB GPUs on AWS
- Estimation with Code Carbon:
    - uses software integrated tools to measure power of RAM, CPU, GPU for total duration

Image modality increases impacts

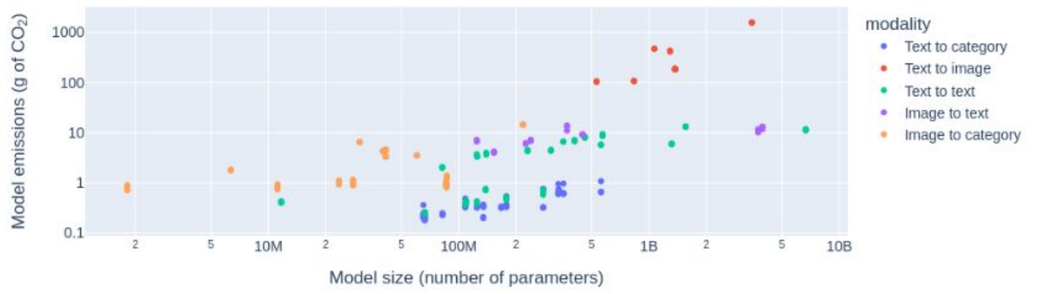
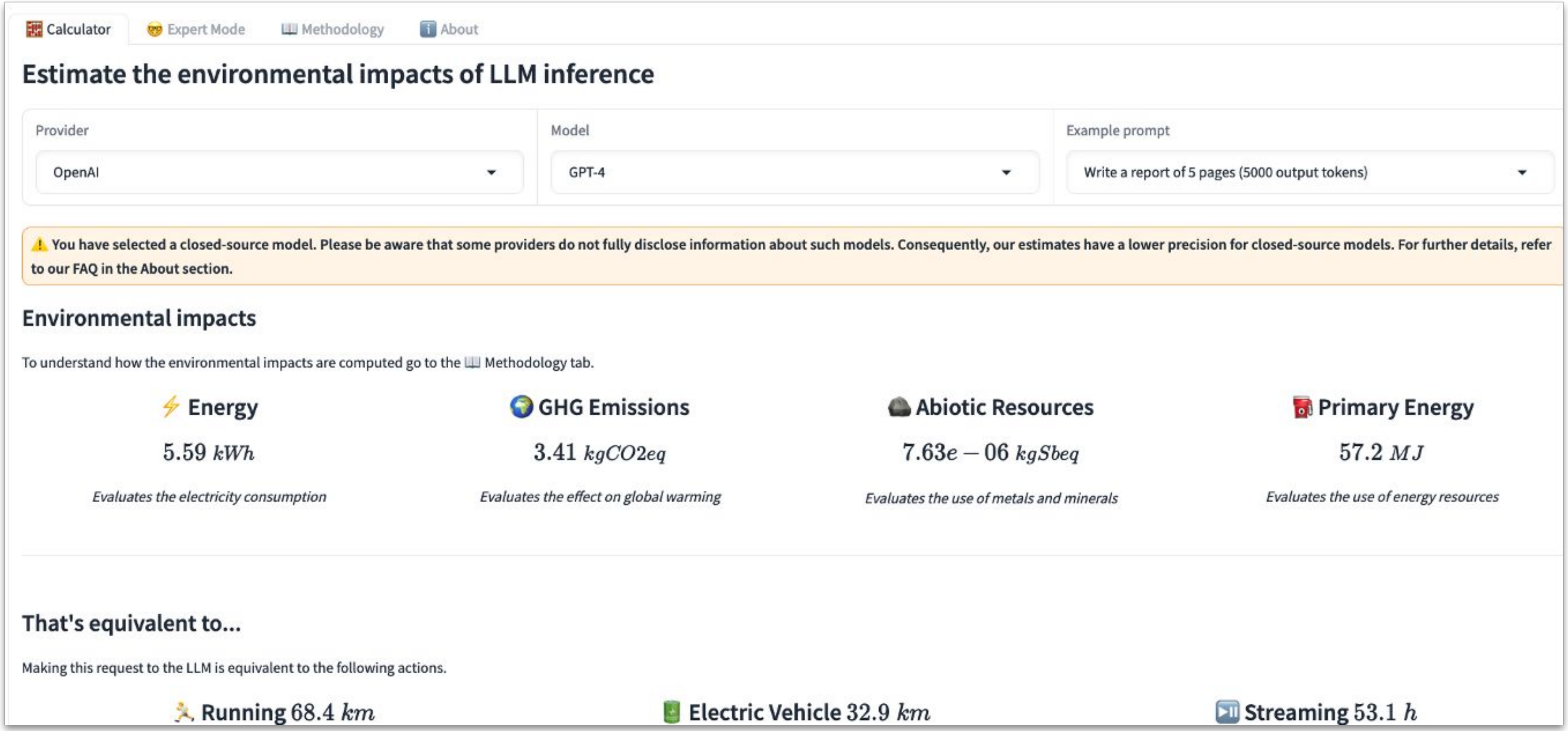


Figure 2: The 5 modalities examined in our study, with the number of parameters of each model on the x axis and the average amount of carbon emitted for 1000 inferences on the y axis. NB: Both axes are in logarithmic scale.

Results for the inference phase of large ML models



Gen AI Impact, ecologits, https://huggingface.co/spaces/genai-impact/ecologits-calculator

Let’s test it!

https://huggingface.co/spaces/genai-impact/ecologits-calculator

No instructions here, just choose a
case study that's relevant to you!

Water use of AI

3 types of water use:

- on-site open-loop cooling systems
- off-site electricity production
- AI chip and server manufacturing

ChatGPT3's training (US average) ~ 5 million L

source: (Li et al, 2024)

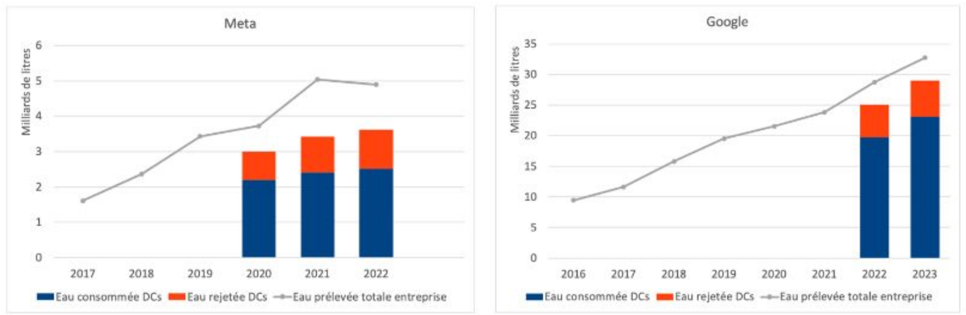
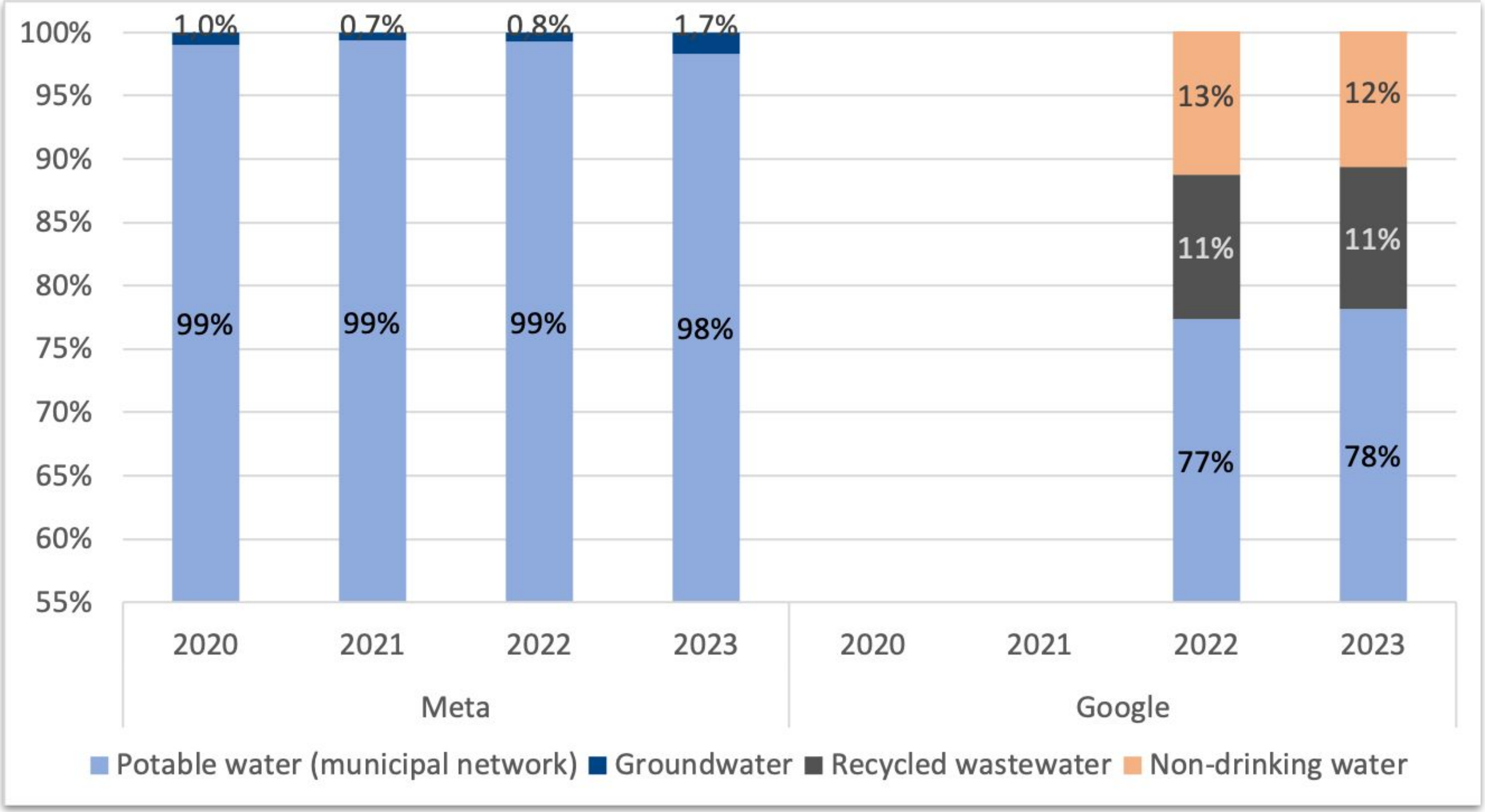


Figure 3 : Consommations et rejets d'eau en milliards de litres dans les DCs en comparaison avec les prélèvements totaux d'eau des entreprises (Source : Meta, 2023 ; Google, 2023 ; Google, 2024).

Source: (Bouveret et al., 2024)

Water use

source: Meta and Google environmental reports, data compiled by Aurélie Bugeau



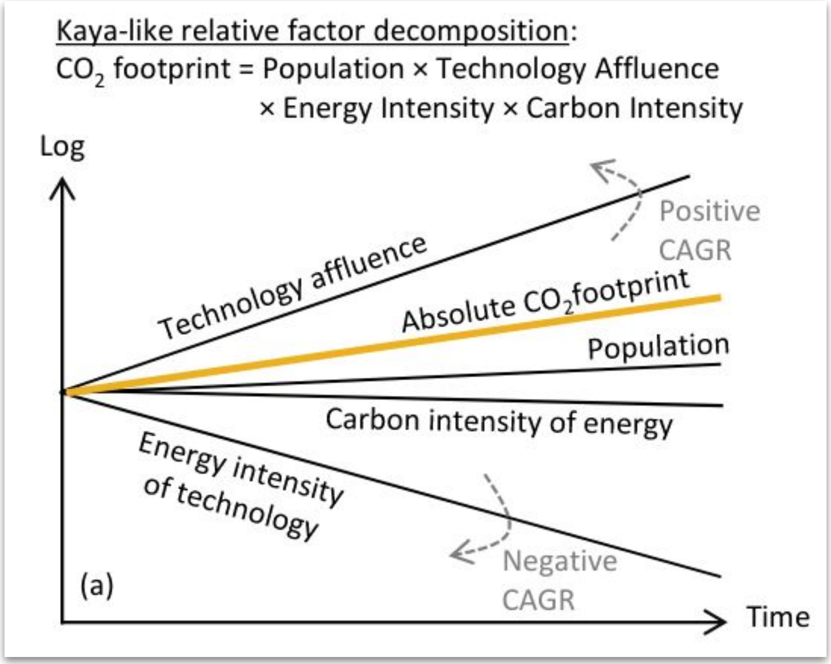
Do AI optimizations lead to decreasing impacts?

Optimization in terms of:

- Hardware
- Algorithm
- Carbon

Does it lead to decreasing impacts? *(Patterson et al, 2022)*

Growth of ICT often absorbs optimization *(Bol et al, 2021)*



Available data on AI models

Companies and academics share some data on AI models

ex: epoch AI

- heterogeneous and potentially inconsistent data
- but enables to analyze training and inference of AI models

Results for the training phase of large AI models

results of *How Green Can AI Be? (Morand et al., 2024)*

Methodology:

- Gathering information on Graphics cards for Machine Learning

TECHPOWERUP

TechPowerUp
GPU database

Wikipedia list of
NVIDIA graphics cards

Other sources
(e.g. Google documentation)

NVIDIA Workstation graphics cards
between 2013 & 2023

other graphics cards
(e.g. TPUs)

Results for the training phase of large AI models

results of *How Green Can AI Be? (Morand et al., 2024)*

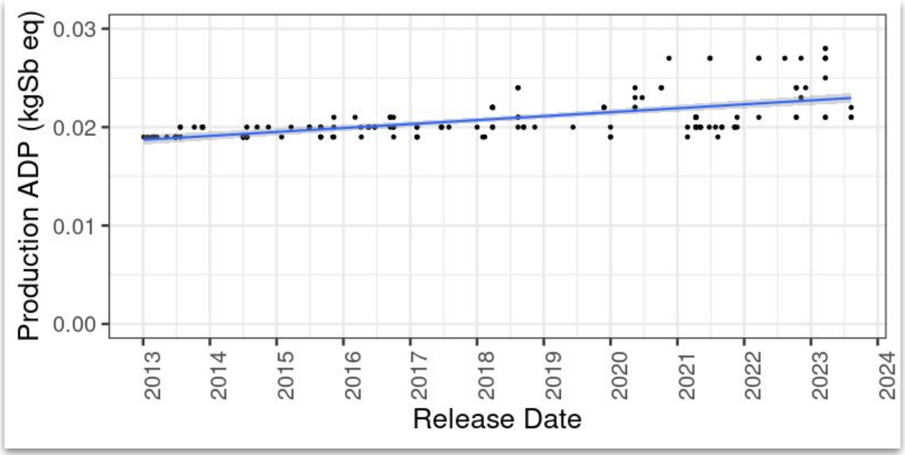
Methodology:

- Estimating training duration
  - If training duration and #cards available (13% of entries), used this info
  - Else if training hardware and #FLOP (25% of entries), estimated duration with average performance ratio

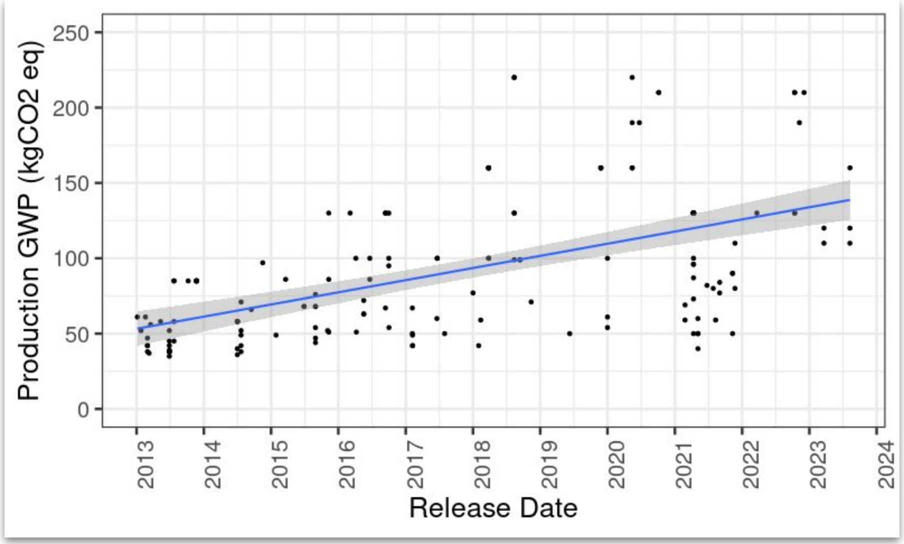
  - No information on memory
  - Hypotheses: 1.1 PUE, 50% usage rate of infrastructure, 3 years server life time, 100% hardware usage during training, carbon intensity of the country of the model producer…

Graphics card evolution

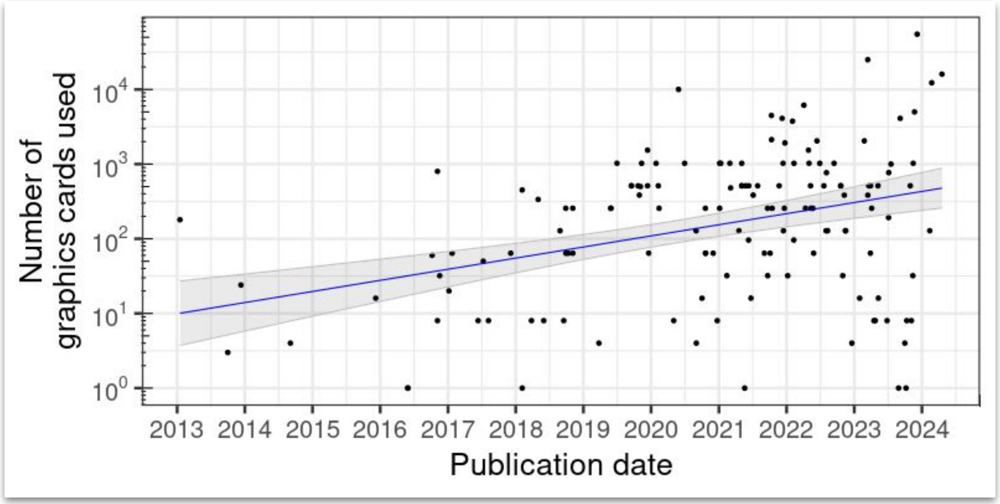
Resource consumption of production



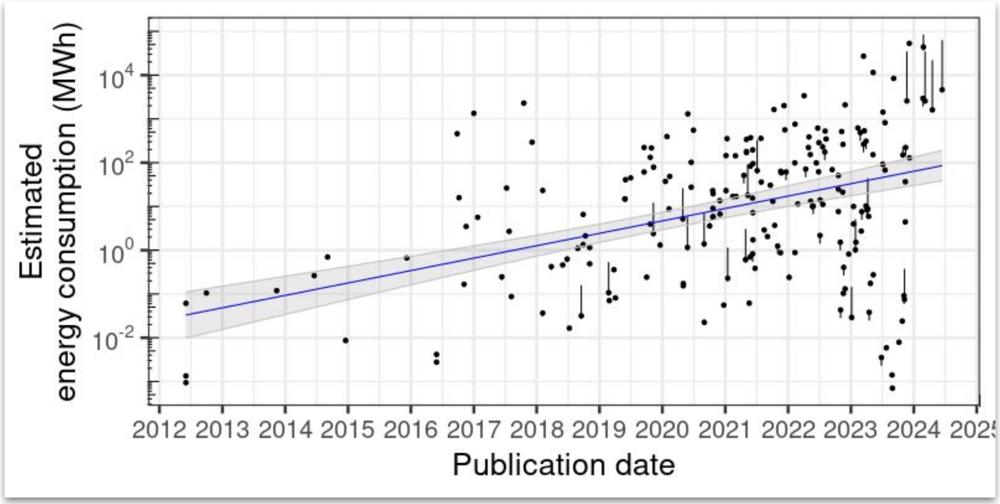
Carbon footprint of production



Large increase in the number of cards to train models

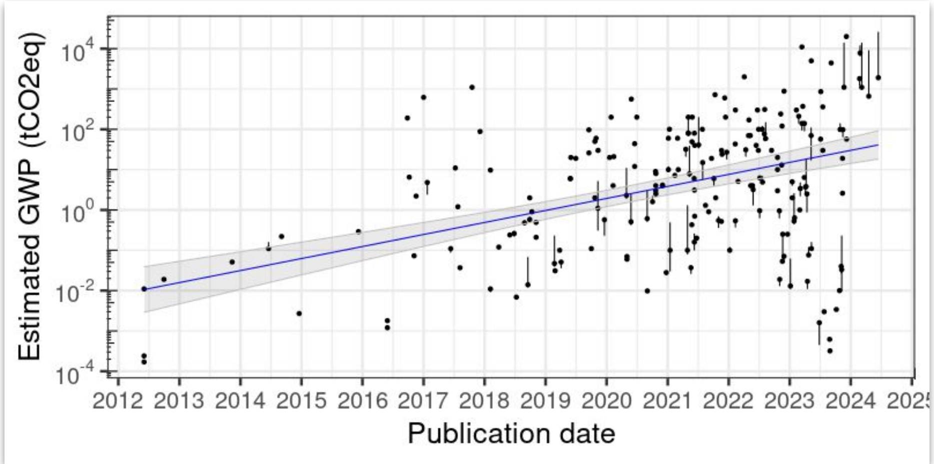


Exponential increase in the energy consumption of models training

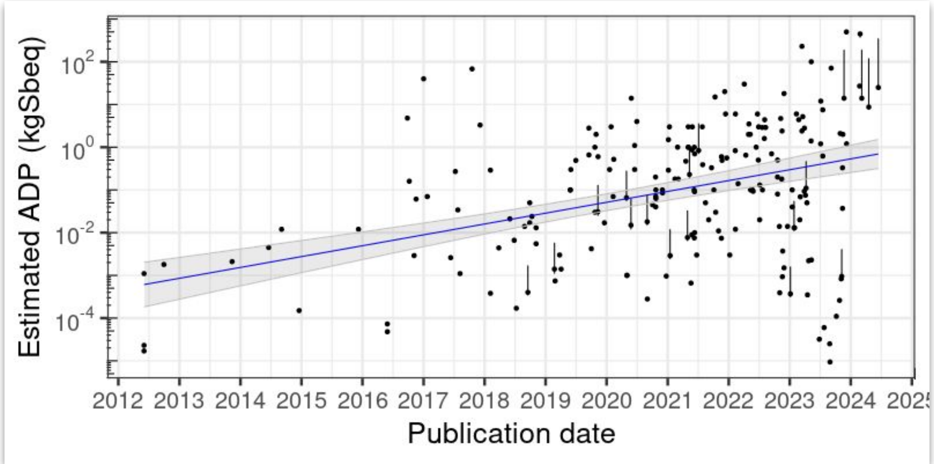


Exponential increase in the environmental damages of models training

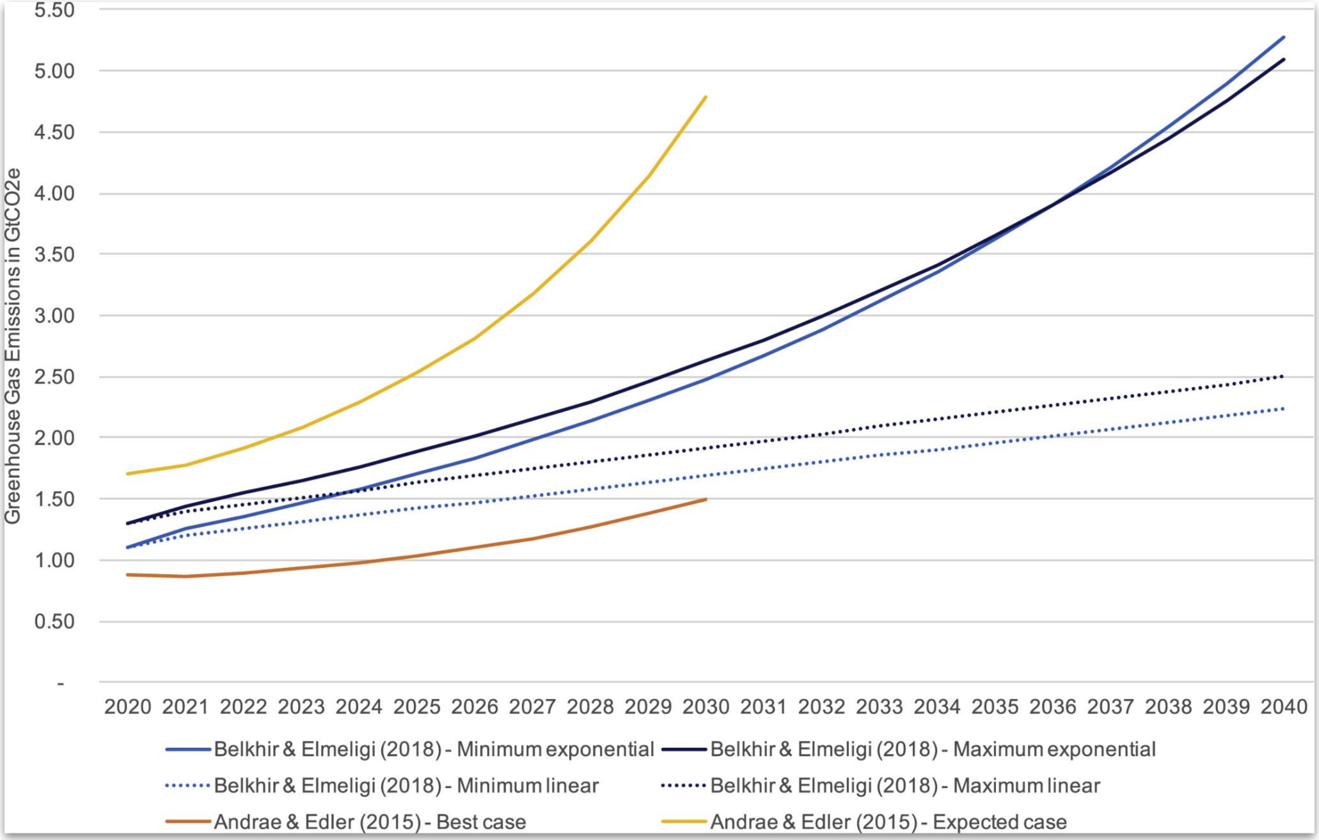
Carbon footprint of training



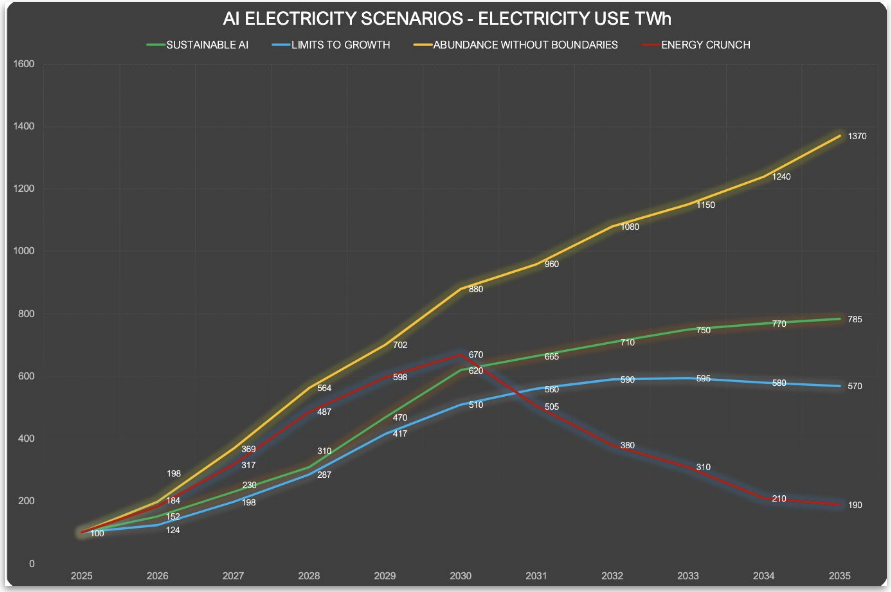
Resource consumption of training



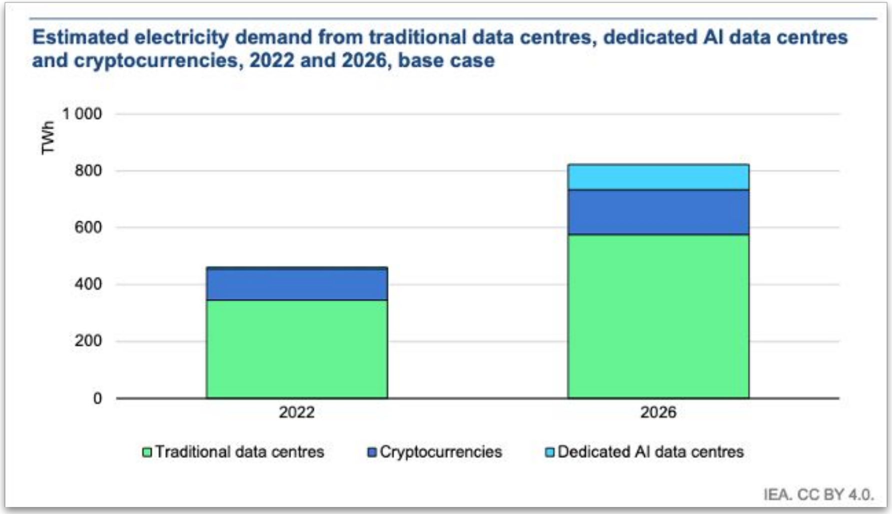
Projections of ICT's GHG emissions from 2020 (Freitag et al, 2021)

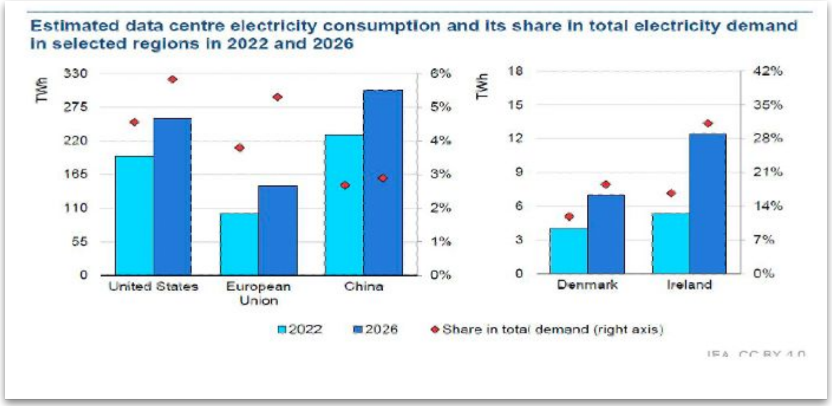


Some general trends



(Wijnhoven et al., Schneider Electric, 2024)





(IEA, 2024)

Google's answer to (Strubell et al., 2019)

The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink

David Patterson[1,2], Joseph Gonzalez[2], Urs Hölzle[1], Quoc Le[1], Chen Liang[1], Lluis-Miquel Munguia[1], Daniel Rothchild[2], David So[1], Maud Texier[1], and Jeff Dean[1]

Best practices proposed:

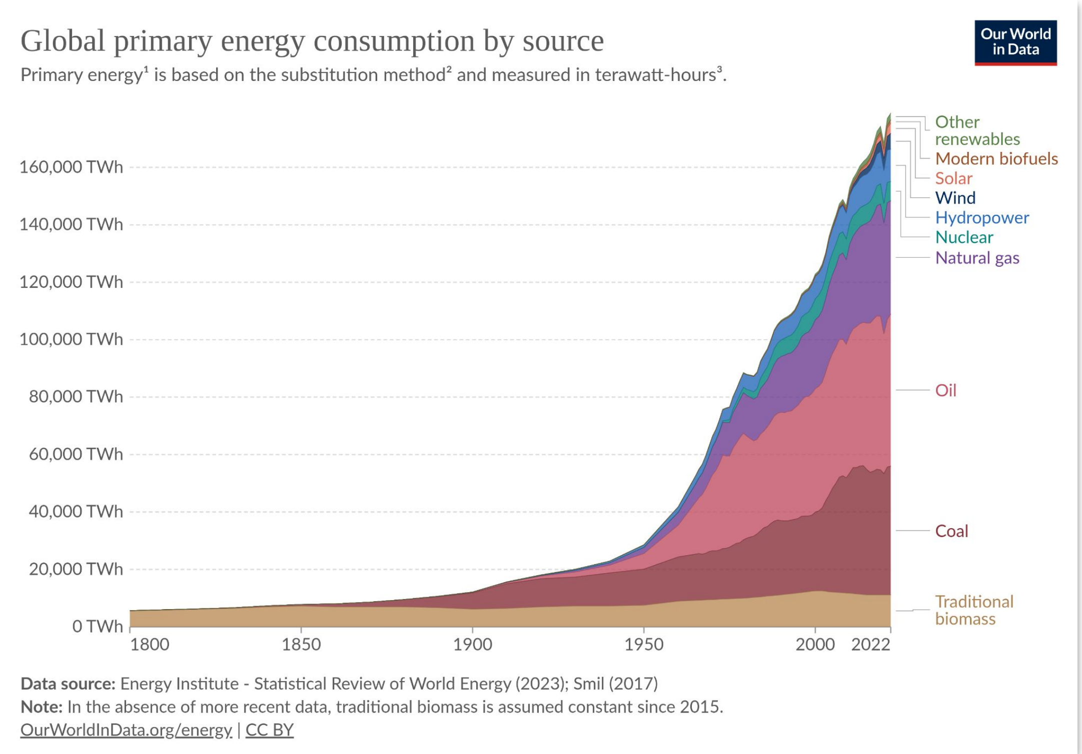
- Efficient ML model
- Processors optimized for ML training
- Cloud for better energy efficiency
- Location with the "cleanest" energy

and «Google's renewable energy purchases further reduce the impact to zero»

but:

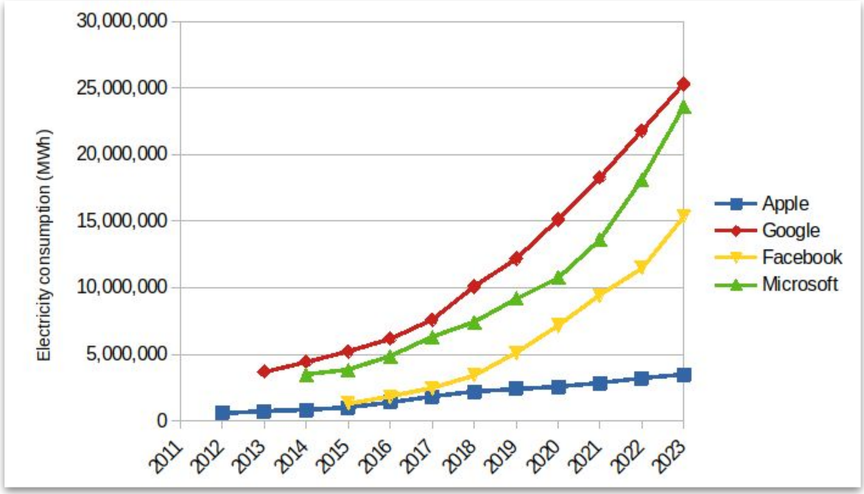
- what about the life cycle?
  - recent processors ⇒ carbon footprint ↗
- what about inference?
- «carbon free» energy and «net zero impact»?
- potential carbon footprint if everything optimized, but not actual one
- focus on carbon footprint

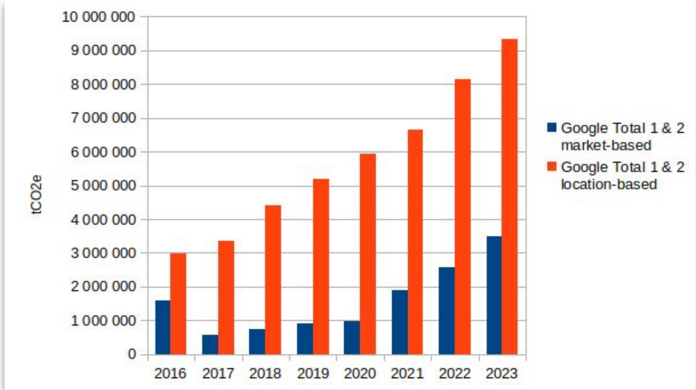
Decarbonization of energy?



Influence of AI in carbon footprints of companies

Google: «As we further integrate AI into our products, **reducing emissions may be challenging** due to increasing energy demands from the **greater intensity of AI compute**, and the emissions associated with the expected increases in our technical infrastructure investment.»





source: environmental reports, data compiled by Anne-Laure Ligozat

And still, Google in 2025

«Google's efforts to reduce the emissions intensity (gCO2 e/kWh) and limit water use in high-stress watersheds have significantly reduced the impact of AI serving on the environment»

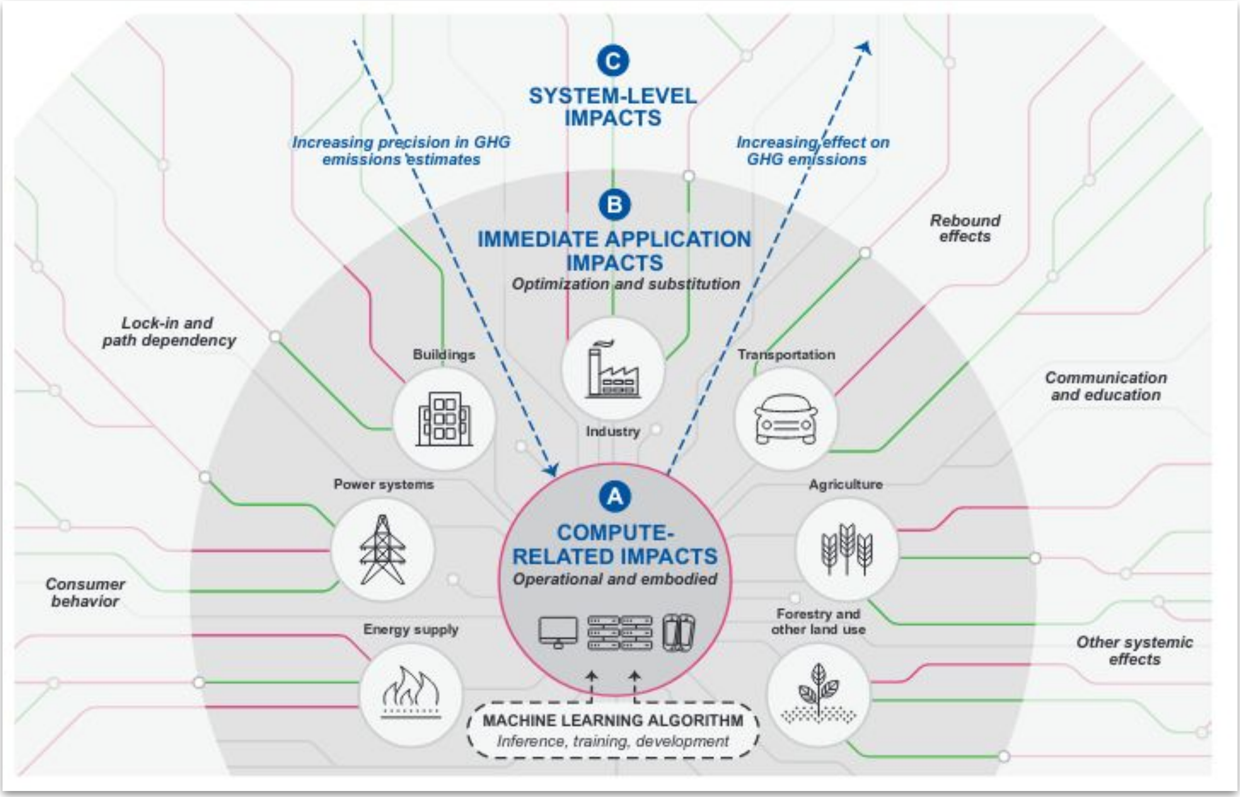
«Google continues to procure clean energy generation in pursuit of our 24/7 carbon-free ambition. In the 2025 annual Environmental Report, we have shown that despite continued growth of electricity consumption from 2023 to 2024, Scope 2 MB emissions have decreased over the same timeframe. This demonstrates an important decoupling between electricity consumption and emissions impact for Google's data centers»

Measuring the environmental impact of delivering AI at Google Scale, 2025
https://services.google.com/fh/files/misc/measuring_the_environmental_impact_of_delivering_ai_at_google_scale.pdf

AI’s own footprint is increasing

Is AI leading to a sustainable world?

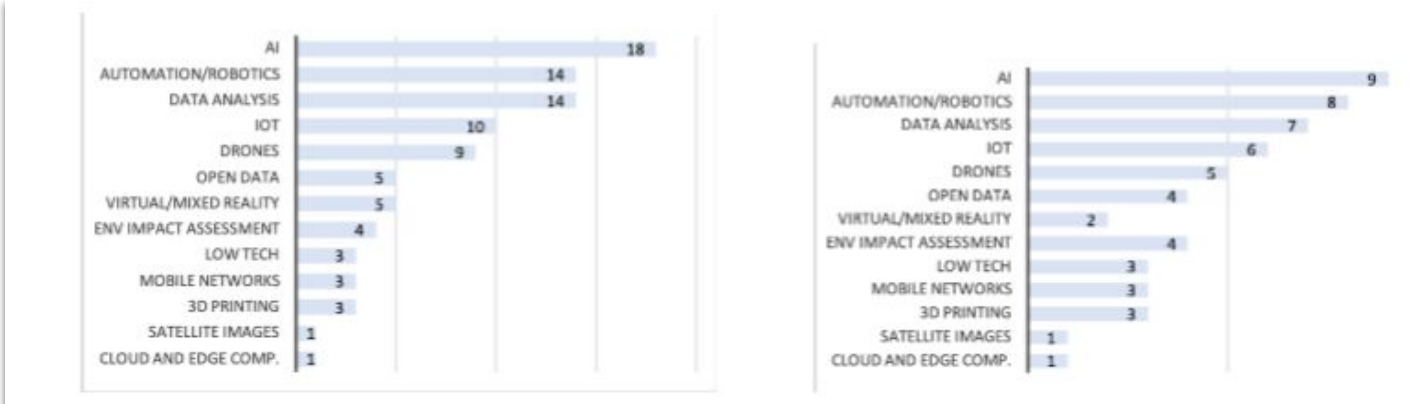
First, second and third order impacts of AI



(Kaack et al., 2021)

AI as a solution?

Work on prospective studies (Bugeau & Ligozat, 2023)

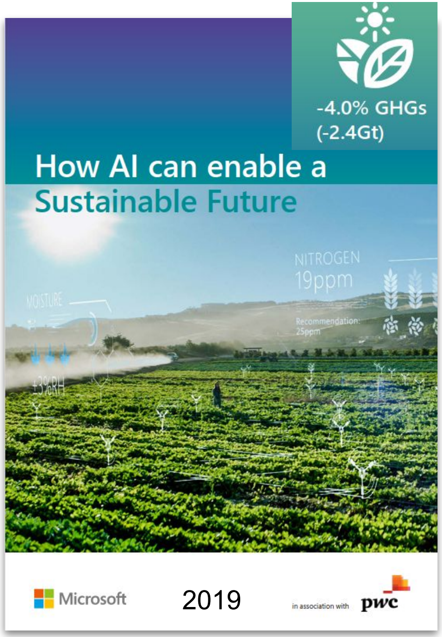


a) Digital technologies by scenario

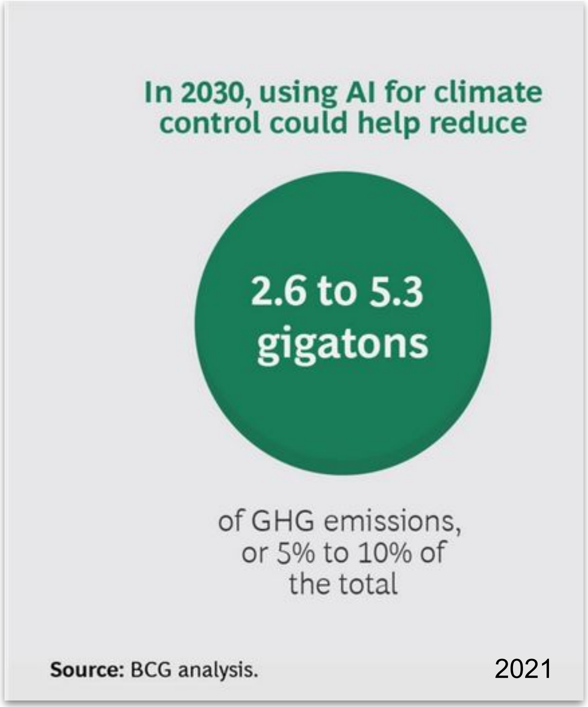
b) Digital technologies by studies

IPCC 2022	*DDC* 2020
Ademe 2022	*SNBC* 2020
negaWatt 2021	*RTE* 2022
EU green deal 2019	*Shift* 2020
Eionet 2022	*France 2072* 2018
Arup 2019	*D&A* 2022
	CNIL 2021
	Digit. Challenge 2022

AI as a solution?

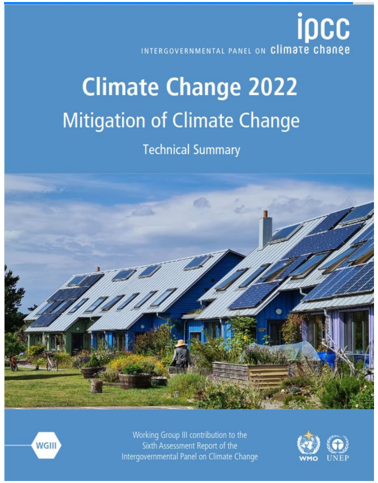


«AI can enable our future systems to be more productive for the economy and for nature. This supports the proposition that we can use AI to help 'decouple' economic growth from GHG emissions.»



AI as a solution?

«Artificial Intelligence (AI) can be deployed for a wide range of applications to promote the goals of the European Green Deal. However, adverse environmental impacts of AI could jeopardise the attainment of these goals.»



(....) artificial intelligence can improve energy management in all sectors, increase energy efficiency, and promote the adoption of many low-emission technologies, including decentralised renewable energy, while creating economic opportunities. However, some of these climate change mitigation gains can be reduced or counterbalanced by growth in demand for goods and services due to the use of digital devices.

«Tackling Climate Change with Machine Learning» (Rolnick et al., 2019)

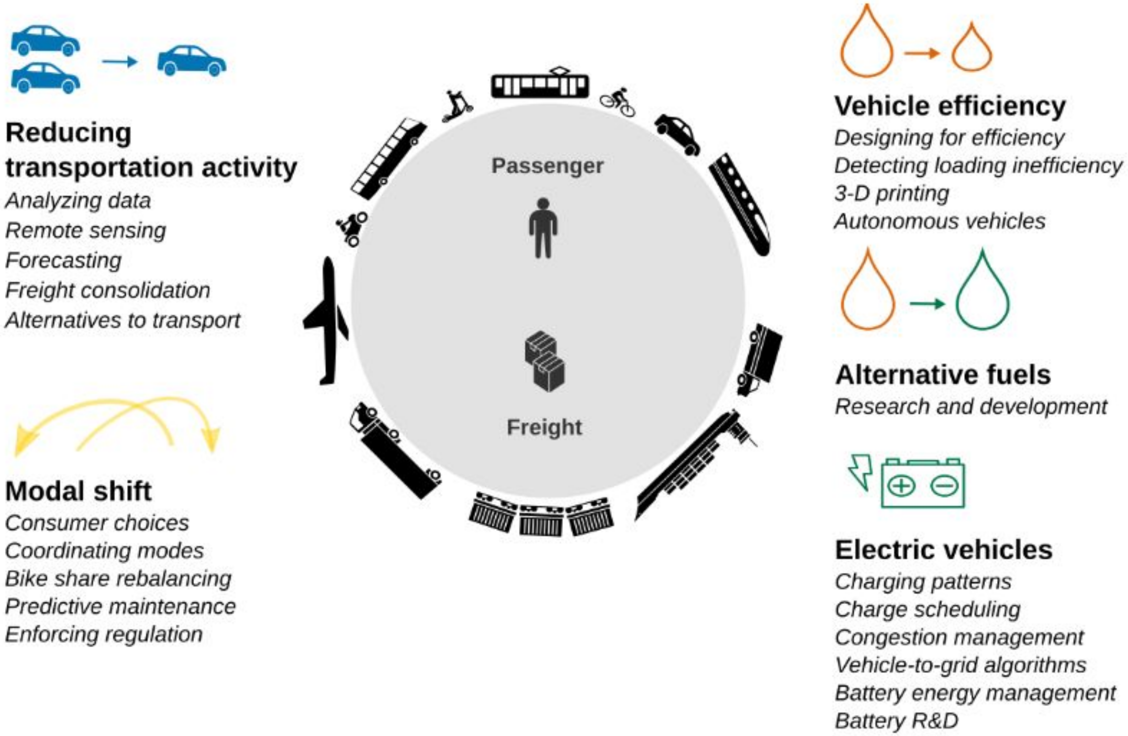
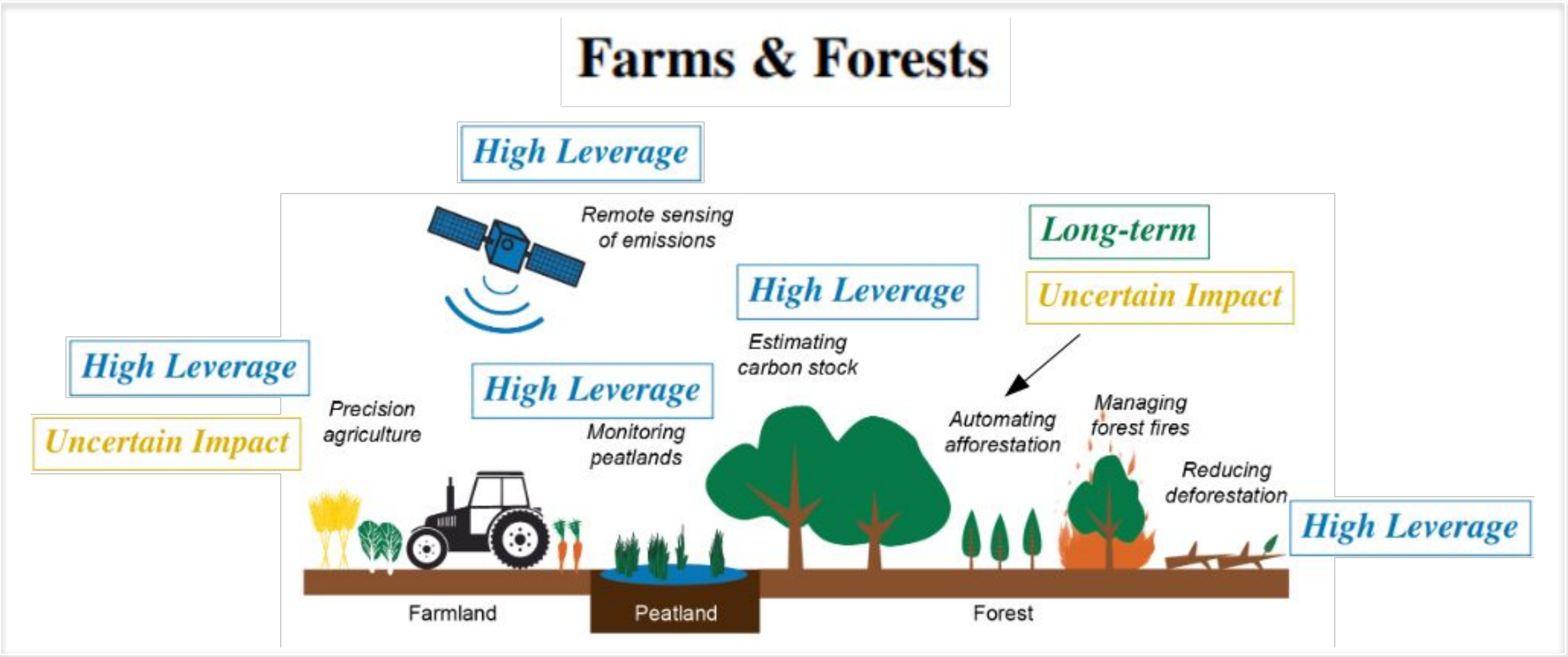


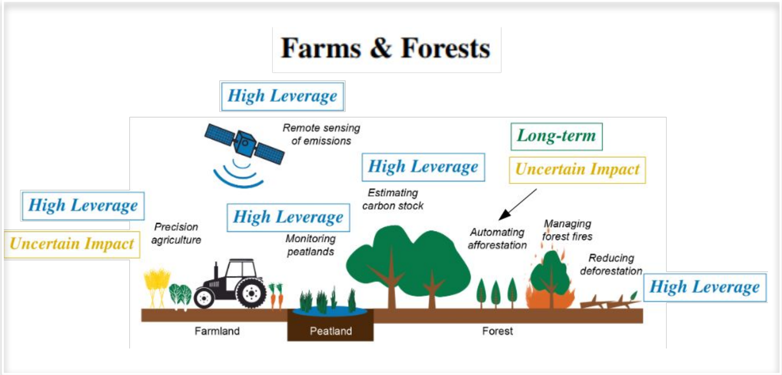
Figure 2: Selected strategies to mitigate GHG emissions from transportation using machine learning.

«Tackling Climate Change with Machine Learning» (Rolnick et al., 2019)



(Brooks,Thorslund &Biørn-Hansen, 2023)

AI4Green and AI4Good vs AI4Bad?



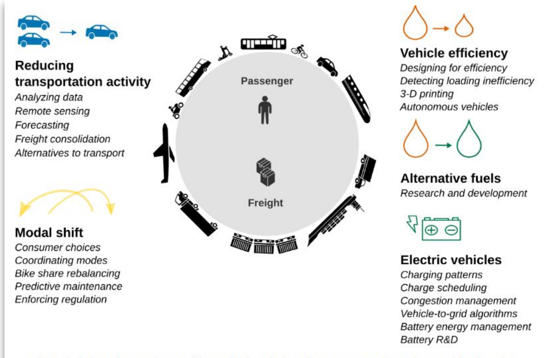
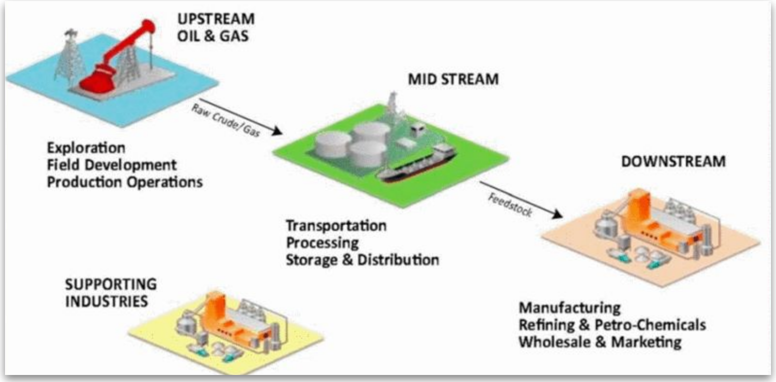
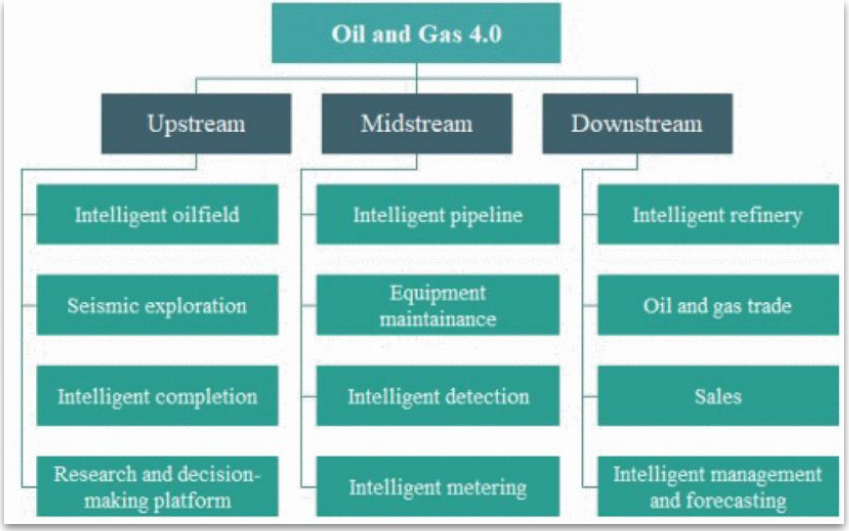


Figure 2: Selected strategies to mitigate GHG emissions from transportation using machine learning.





Is this something we should be concerned about?

Yes: science is inscribed in society

Pilier II

Budget pour les pôles et le JRC

aux prix courants

Pôle 1	**Santé**	**8,246 milliards d'EUR** (dont 1,35 milliard provenant de Next Generation EU)
Pôle 2	**Culture, créativité et sociétés inclusives**	**2,28 milliards d'EUR**
Pôle 3	**Sécurité civile pour la société**	**1,596 milliard d'EUR**
Pôle 4	**Numérique, industrie et espace**	**15,349 milliards d'EUR** (dont 1,35 milliard d'EUR provenant de Next Generation EU)
Pôle 5	**Climat, énergie et mobilité**	**15,123 milliards d'EUR** (dont 1,35 milliard d'EUR provenant de Next Generation EU)
Pôle 6	**Alimentation, bioéconomie, ressources naturelles, agriculture et environnement**	**8,952 milliards d'EUR**
	JRC (actions directes non nucléaires)	**1,97 milliard d'EUR**

Les pôles comprennent un budget consacré aux partenariats et aux missions.
Next Generation EU est l'instrument de l'Union européenne pour la relance

Commission européenne

A. Petit, PDG du CNRS

«Investir aujourd'hui dans la recherche, c'est (...) aider la France et l'Europe à **gagner les guerres** et combats actuels, la lutte contre le changement climatique, une transition énergétique harmonieuse, la **création d'emplois et de valeurs**, le combat contre la radicalisation et les obscurantismes, une **numérisation du monde** au bénéfice du plus grand nombre.» Tribune Les échos 2019

B. Sportisse, PDG de l'Inria
« L'Inria doit accorder la **priorité à son impact économique** sur le tissu industriel français. »
Assemblée nationale 2021

A responsibility of research

Comité d'éthique du CNRS (COMETS)
« COMETS understands [research's] responsibility in a broad sense: it requires thinking about how to limit the footprint of 'everyday' research practices (buying better and less, optimising the use of digital technology, limiting travel and work-related trips (hereinafter referred to as 'scientific missions'), improving the energy performance of buildings); but it must also lead us to consider the environmental footprint of **research topics** and the ways in which they can be addressed»
« [Its specific purpose - to produce knowledge in the service of society -] confers on it the particular **responsibility of also questioning the uses that may be made of this knowledge** (in particular its transformation into innovations) and **how such uses can meet the problems encountered by society or, on the contrary, perpetuate and even aggravate them**.»

Labos 1point5,
https://labos1point5.org/les-textes-positionnement/RechercheEthiqueEnvironnementale#textecollectif

« Il relève de la responsabilité individuelle et collective des scientifiques d'**évaluer les impacts potentiels sur l'environnement de leurs propres travaux et de placer la limite entre des sujets à traiter et d'autres à laisser de côté en raison de leur impact négatif**.»

Indirect impacts

direct impacts

indirect impacts

optimize traffic flow?

use of new connected objects, sensors…

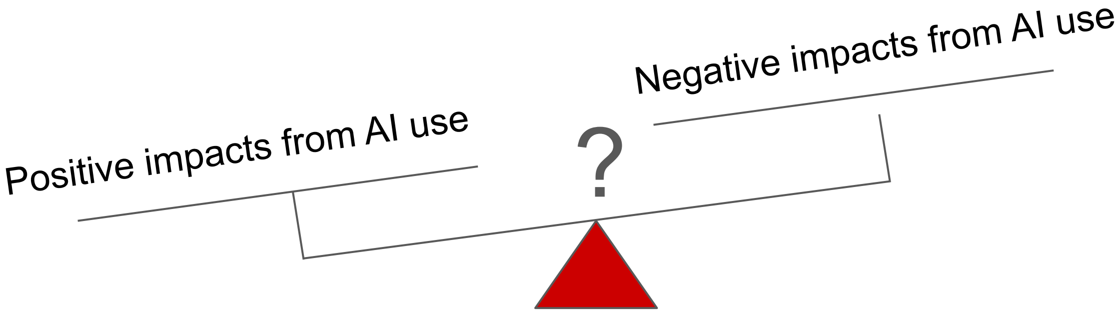
lower fuel consumption

rebound effect
smoother traffic flow => time savings => greater distance from home => urban sprawl

path dependency
prolongs current system, vs. public transport, active mobility...

priority to systems with significant impacts?

AI for environmental applications



at least with Life Cycle Assessment

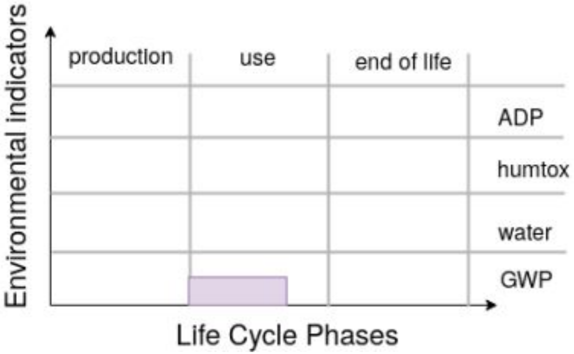
taking into account as many indirect effects as possible

Life cycle assessment of AI systems

(Ligozat et al, 2021)

Assessing the environmental impacts of an AI system should at least include a Life Cycle Assessment

How are AI for Green systems benefits assessed?

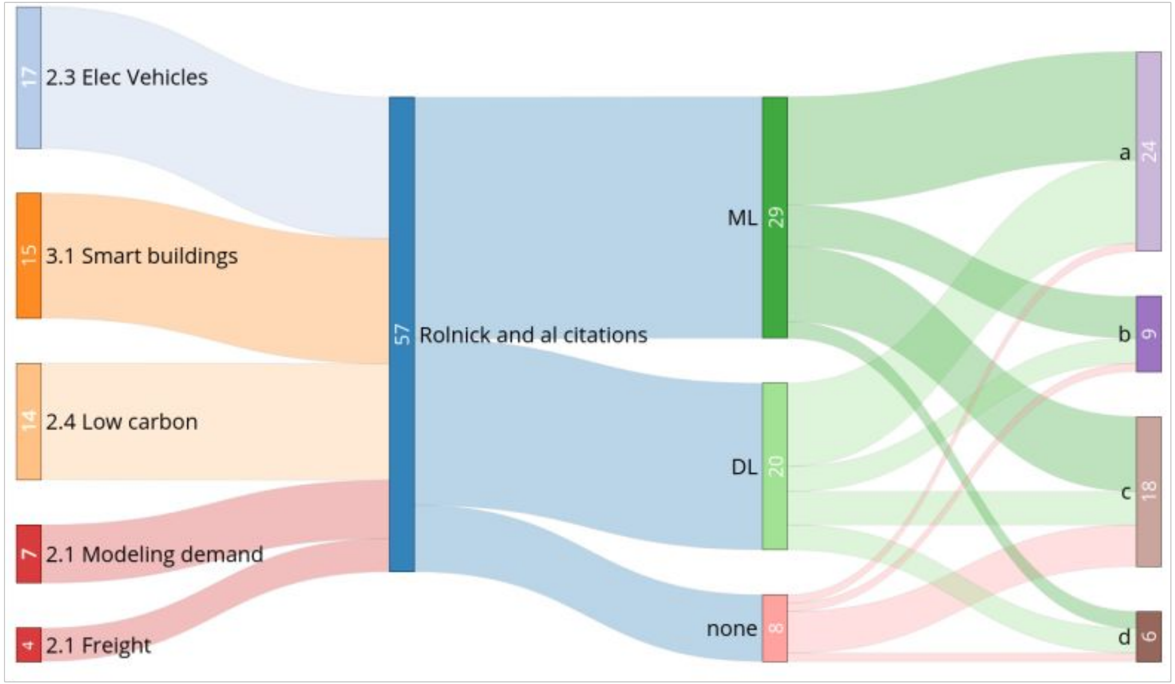


$$\Delta(M_2|M_1) = LCA(M_2) - LCA(M_1) \in \mathbb{R}^d \quad (1)$$

with:

- M_1 the reference application without using the AI service,
- M_2 the application enhanced by AI,
- $LCA(x)$ a quantification of d types of environmental impacts (e.g., GHG emissions, water footprint, etc.). The LCA methodology is described in Section 3.2. Note that $LCA(M_2)$ includes the impacts of the AI service itself, i.e., $LCA_{AI}(M_2)$.

Evaluations in (Rolnick et al., 2019)



a. No mention of the environmental gain

b. General mention of the environmental gain

c. A few words about the environmental gain but no quantitative evaluation or only indirect estimation

d. Evaluation of the energy gain without taking the AI program into account

Biases of impact studies (Rasoldier et al., 2022)

Perimeter

- life cycle not taken into account: (Ligozat et al., 2021) for AI
- indirect (2nd and 3rd order) not taken into account: 5G

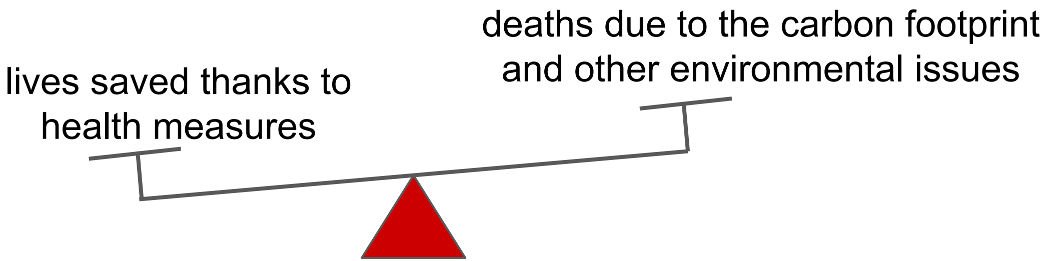
Hypotheses

- comparison to what reference scenario?

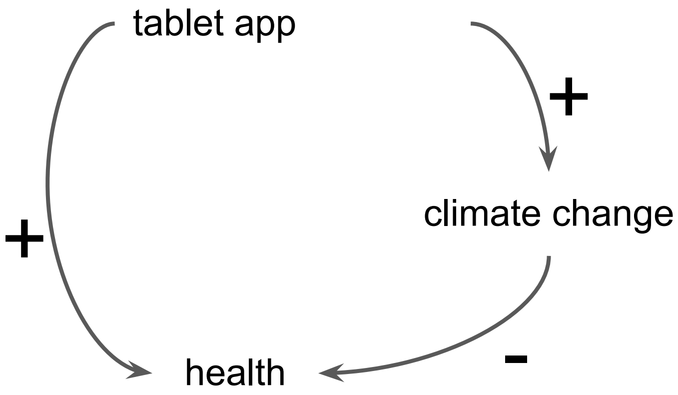
Disconnection from global scenarios

- minimal benefits + poorly managed uncertainties
- incompatibility between measures

Example in the health sector



[Conférence Comprendre et Agir -- Valérie d'Acremont]



Case study

Système considéré : optimisation de la consommation de chauffage et climatisation d’un bâtiment par analyse de la présence

1. Quels sont les impacts environnementaux directs du système ?
2. Quels sont les impacts environnementaux indirects du système ?
    - impact principal : quels impacts positifs sont visés ?
    - autres changements possibles
3. Critères IA Frugale
    - la nécessité de recourir à un système d’IA plutôt qu’à une autre solution moins consommatrice pour répondre au même objectif a été démontrée ;
    - de bonnes pratiques (..) sont adoptées par le producteur, le fournisseur et le client d’IA pour diminuer les impacts environnementaux du service utilisant un algorithme d’IA ; (c’est-à-dire que le service est optimisé)
    - les usages et les besoins visent à rester dans les limites planétaires et ont été préalablement questionnés.
4. Inscription dans les scénarios de l’Ademe

LA SOCIÉTÉ EN 2050

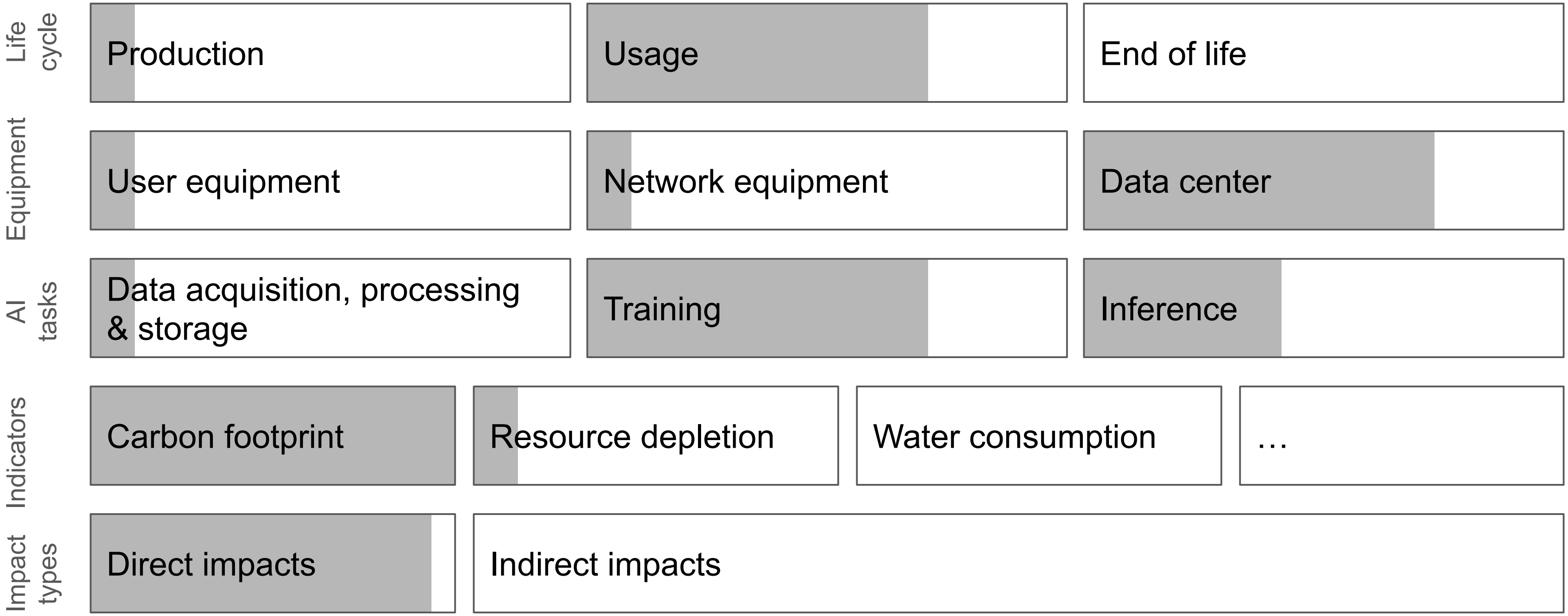



		S1 GÉNÉRATION FRUGALE	S2 COOPÉRATIONS TERRITORIALES	S3 TECHNOLOGIES VERTES	S4 PARI RÉPARATEUR
MODES DE VIE	Société	• Recherche de sens • **Frugalité choisie mais aussi contrainte** • Préférence pour le local • Nature sanctuarisée	• Évolution soutenable des modes de vie • **Économie du partage** • Équité • Préservation de la nature inscrite dans le droit	• **Plus de nouvelles technologies que de sobriété** • Consumérisme « vert » au profit des populations solvables, société connectée • Les services rendus par la nature sont optimisés	• Sauvegarde des modes de vie de **consommation de masse** • La nature est une ressource à exploiter • Confiance dans la capacité à réparer les dégâts causés aux écosystèmes
	Alimentation	• Division par 3 de la consommation de viande • **Part du bio : 70 %**	• **Division par 2 de la consommation de viande** • Part du bio : 50 %	• Baisse de 30 % de la consommation de viande • **Part du bio : 30 %**	• Consommation de viande quasi-stable (baisse de 10 %), complétée par des **protéines de synthèse ou végétales**
	Habitat	• Rénovation massive et rapide • **Limitation forte de la construction neuve** (transformation de logements vacants et résidences secondaires en résidences principales)	• Rénovation massive, **évolutions graduelles mais profondes des modes de vie** (cohabitation plus développée et adaptation de la taille des logements à celle des ménages)	• **Déconstruction-reconstruction** à grande échelle de logements • Ensemble des logements rénovés mais de façon peu performante : la moitié seulement au niveau Bâtiment Basse Consommation (BBC)	• Maintien de la **construction neuve** • La moitié des logements seulement est rénovée au niveau BBC • **Les équipements se multiplient**, alliant innovations technologiques et efficacité énergétique
	Mobilité des personnes	• **Réduction forte de la mobilité** • Réduction d'un tiers des km parcourus par personne • La moitié des trajets à pied ou à vélo	• **Mobilité maîtrisée** • - 17 % de km parcourus par personne • Près de la moitié des trajets à pied ou à vélo	• Mobilités accompagnées par l'État pour les maîtriser : infrastructures, télétravail massif, covoiturage • **+ 13 % de km parcourus par personne** • 30 % des trajets à pied ou à vélo	• Augmentation forte des mobilités • + 28 % de km parcourus par personne • Recherche de **vitesse** • 20 % des trajets à pied ou à vélo
	Technique Rapport au progrès, numérique, R&D	• Innovation autant organisationnelle que technique • **Règne des *low-tech***, réutilisation et réparation • Numérique collaboratif • **Consommation des *data centers* stable** grâce à la stabilisation des flux	• Investissement massif (efficacité énergétique, EnR et infrastructures) • Numérique au service du développement territorial • **Consommation des *data centers* stable** grâce à la stabilisation des flux	• Ciblage sur les **technologies les plus compétitives pour décarboner** • Numérique au service de l'optimisation • Les *data centers* consomment 10 fois plus d'énergie qu'en 2020	• Innovations tout azimut • Captage, stockage ou usage du carbone capté indispensable • Internet des objets et intelligence artificielle omniprésents : les ***data centers* consomment 15 fois plus d'énergie** qu'en 2020
	Gouvernance Échelles de décision, coopération internationale	• **Décision locale**, faible coopération internationale • Réglementation, interdiction et rationnement *via* des quotas	• Gouvernance partagée • **Fiscalité environnementale** et redistribution • Décisions nationales et coopération européenne	• Cadre de **régulation minimale** pour les acteurs privés • État planificateur • Fiscalité carbone ciblée	• Soutien de l'offre • Coopération internationale forte et ciblée sur quelques filières clés • **Planification centralisée du système énergétique**
	Territoire Rapport espaces ruraux – urbains, artificialisation	• Rôle important du territoire pour les ressources et l'action • **« Démétropolisation »** en faveur des villes moyennes et des zones rurales	• **Reconquête démographique des villes moyennes** • Coopération entre territoires • Planification énergétique territoriale et politiques foncières	• **Métropolisation**, mise en concurrence des territoires, villes fonctionnelles	• Faible dimension territoriale, **étalement urbain**, agriculture intensive
ÉCONOMIE	Macro-économie	• **Nouveaux indicateurs de prospérité** (écarts de revenus, qualité de la vie…) • Commerce international contracté	• Croissance qualitative, **« réindustrialisation »** de secteurs clés en lien avec territoires • Commerce international régulé	• **Croissance verte**, innovation poussée par la technologie • Spécialisation régionale • Concurrence internationale et échanges mondialisés	• **Croissance économique carbonée** • Fiscalité carbone minimaliste et ciblée • Économie mondialisée
	Industrie	• **Production au plus près des besoins** • 70 % de l'acier, mais aussi de l'aluminium, du verre, du papier-carton et des plastiques viennent du recyclage	• Production en valeur plutôt qu'en volume • **Dynamisme des marchés locaux** • 80 % de l'acier, mais aussi de l'aluminium, du verre, du papier-carton et des plastiques viennent du recyclage	• **Décarbonation de l'énergie** • 60 % de l'acier, mais aussi de l'aluminium, du verre, du papier-carton et des plastiques viennent du recyclage	• Décarbonation de l'industrie pariant sur le **captage et stockage géologique de CO2** • 45 % de l'acier, mais aussi de l'aluminium, du verre, du papier-carton et des plastiques viennent du recyclage

AI’s benefits on other sectors are unclear

What can we do?

What is presently assessed



Red vs Green AI (Schwartz et al., 2020)

Red AI

- improve accuracy rather than efficiency, through the use of massive computational power while disregarding the cost
    - even though relationship between model performance and model complexity is at best logarithmic
- yet valuable: contributes to what we know about pushing the boundaries of AI

but

⇒ allow for more equitable comparisons, eg reporting training curves

⇒ recognize Green AI work

Green AI

novel results encouraging a reduction in resources spent

Responsible AI?

- < >
  Déclaration de Montréal
  IA responsable_
  </ >
  (Dilhac et al., 2018)
    - AI systems and associated equipment must aim for maximum energy efficiency and minimize the carbon footprint over their entire lifecycle, as well as impacts on ecosystems and biodiversity...
- Villani report (2018)
    - (...) AI can lead to numerous rebound effects. For example AI can prevent us from rethinking our modes of growth, consumption, and measurement of wealth produced, and instead to consume just as much as before, if not more.

AFNOR General framework for frugal AI (2024)

Discussions about:

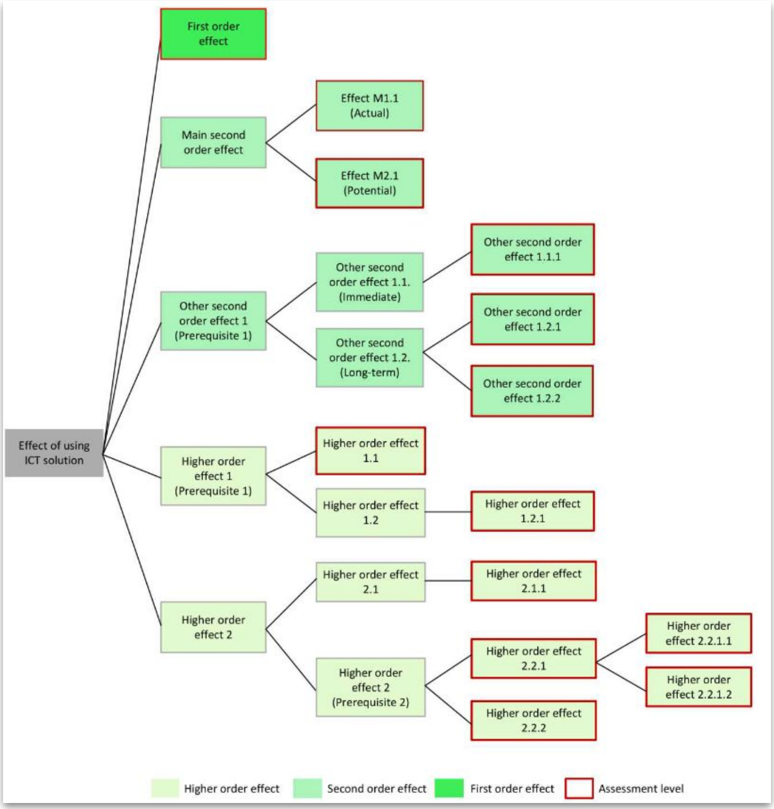
- perimeter of the specification (who is it addressed to?)
- definition of frugality (vs efficiency): redefinition of needs and usages
  - frugal service: includes AI if necessary + usages and needs questioned and aim at staying in planetary boundaries

	Definition	Related concepts	Reasoning	Approach	Comments
Efficiency	Ability to optimize resources to achieve a defined outcome	Effectiveness, optimization	Relative/per unit of use Need takes precedence: optimizing a solution deemed to best meet the need	Seeking a local optimum or compromise on a highly constrained outcome level	Consideration of first order effects in order to minimize them Consideration of AI stakeholders
Frugality	Ability to accept an outcome level deemed sufficient by redefining uses and needs	Sufficiency [10)]	Holistic Constraints on resources take precedence: seeking the solution using the least resources possible and meeting the need in a satisfactory way	Seeking an overall optimum or a broad compromise on an outcome level, implying a broader or more flexible approach to the need	Consideration of first and second order effects in order to minimize negative environmental impacts Broad consideration of all players beyond AI stakeholders alone

AFNOR General framework for frugal AI

Evaluation

- 1st order effects (multi-indicators and multi-phases of the life cycle)
  - follows ACV ISO standards, ITU-T L.1410 recommendation and PCR (Product Category Rules) Digital Services
  - cf. MLCA tools (Morand et al., 2024)
- but also 2nd and 3rd order
  - according to ITU-T L .1480 recommendation



Rethinking AI and computer science

What place for AI and more generally ICT (Information and Communication Technologies) in prospective sustainable scenarios?

- How does ICT fit into existing mitigation scenarios
- How can focusing on ICT build new scenarios from the degrowth perspective

How to build digital alternatives for sustainable worlds?

- What are the present and desirable values in ICT design?
- What digital infrastructures would be sustainable?

ADN
INRIA
team

Conclusion

- Comprehensive evaluation of the environmental impacts remains a work in progress
- But tools for partial evaluation of 1st order impacts exist and can easily be used
- As well as guidelines for a discussion of 2nd and 3rd order impacts
- But be careful with partial indicators

- Need for discussion on the role of AI in a green transition
- And of the role of computer science research

References (1/2)

- Anthony, L. F. W., Kanding, B., & Selvan, R. (2020). Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models. *ArXiv:2007.03051 [Cs, Eess, Stat]*. http://arxiv.org/abs/2007.03051
- Berthoud, F.; Bzeznik, B.; Gibelin, N.; Laurens, M.; Bonamy, C.; Morel, M.; Schwindenhammer, X. Estimation de l'empreinte carbone d'une heure.coeur de calcul. Research report, UGA - Université Grenoble Alpes ; CNRS ; INP Grenoble ; INRIA, 2020. https://hal.archives-ouvertes.fr/hal-02549565v4/
- Bol, D., Pirson, T., Dekimpe, R. (). Moore's Law and ICT Innovation in the Anthropocene. IEEE Design, Automation and Test in Europe Conference 2021. http://hdl.handle.net/2078.1/243578
- Brooks, I. , Thorslund, M. L. and Biørn-Hansen, A. (2024) "Tech4Bad in the Oil and Gas Industry: Exploring Choices for ICT Professionals," 2023 International Conference on ICT for Sustainability (ICT4S), Rennes, France, 2023, pp. 142-153, doi: 10.1109/ICT4S58814.2023.00023
- Bugeau, A., Ligozat, A.-L. (2023). How digital will the future be? Analysis of prospective scenarios. https://hal.science/hal-04362220v2
- Dilhac, M.-A., Abrassart, C., Voarino, N., et al. (2018). Rapport de la déclaration de montréal pour un développement responsable de l'intelligence artificielle. https://www.declarationmontreal-iaresponsable.com/la-declaration
- Freitag, C., Berners-Lee, M., Widdicks, K., Knowles, B., Blair, G. S., Friday, A. (2022). The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations. Patterns, Volume 3, Issue 8, 12 August 2022, Pages 100576. https://doi.org/10.1016/j.patter.2021.100340
- Guyon, D. (2018). Supporting energy-awareness for cloud users. PhD thesis, Université Rennes 1.
- Heinrich, F. C., Cornebize, T., Degomme, A., Legrand, A., Carpen-Amarie, A., Hunold, S., Orgerie, A.-C., and Quinson, M. (2017). Predicting the energy-consumption of mpi applications at scale using only a single node. In 2017 IEEE International Conference on Cluster Computing (CLUSTER).
- Hershcovich, D., Webersinke, N., Kraus, M., Bingler, J. A., & Leippold, M. (2022). Towards Climate Awareness in NLP Research. arXiv:2205.05071 [cs]. http://arxiv.org/abs/2205.05071
- Mathilde Jay, Vladimir Ostapenco, Laurent Lefèvre, Denis Trystram, Anne-Cécile Orgerie, et al.. An experimental comparison of software-based power meters: focus on CPU and GPU. CCGrid 2023 - 23rd IEEE/ACM international symposium on cluster, cloud and internet computing, May 2023, Bangalore, India. pp.1-13. https://cnrs.hal.science/hal-04030223v2
- Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2021). *Aligning artificial intelligence with climate change mitigation*. https://hal.archives-ouvertes.fr/hal-03368037

References (2/2)

- Ligozat, A.-L., Lefèvre, J., Bugeau, A., & Combaz, J. (2021). Unraveling the hidden environmental impacts of AI solutions for environment. *arXiv:2110.11822 [cs]*. http://arxiv.org/abs/2110.11822 and Sustainability https://www.mdpi.com/2071-1050/14/9/5172/htm
- Ligozat, A.L.; Luccioni, A. A Practical Guide to Quantifying Carbon Emissions for Machine Learning researchers and practitioners. 462 Technical report, Bigscience project, LISN and MILA, 2021. https://hal.archives-ouvertes.fr/hal-03376391
- Luccioni, A. S., Viguier, S., Ligozat, A.-L. Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model. Journal of Machine Learning Research, 2023, 24 (253). https://www.jmlr.org/papers/volume24/23-0069/23-0069.pdf
- Parcollet, T., Ravanelli, M.. The Energy and Carbon Footprint of Training End-to-End Speech Recognizers. 2021. https://hal.science/hal-03190119
- Rasoldier, A., Combaz, J., Girault, A., Marquet, K., Quinton, S. (2022). How realistic are claims about the benefits of using digital technologies for GHG emissions mitigation?. LIMITS 2022 - Eighth Workshop on Computing within Limits, Jun 2022, Virtual, France. https://hal.science/hal-03949261/
- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., … (2019). Tackling Climate Change with Machine Learning. *ArXiv:1906.05433 [Cs, Stat]*. http://arxiv.org/abs/1906.05433
- Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, *63*(12), 54-63. https://doi.org/10.1145/3381831
- Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbhahn, M., & Villalobos, P. (2022). Compute Trends Across Three Eras of Machine Learning. arXiv:2202.05924 [cs]. http://arxiv.org/abs/2202.05924
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *ArXiv:1906.02243 [Cs]*. http://arxiv.org/abs/1906.02243
- Villani, Cédric. (2018). *For a meaningful artificial intelligence*. https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf
- Wu, C.-J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Behram, F. A., Huang, J., Bai, C., Gschwind, M., Gupta, A., Ott, M., Melnikov, A., Candido, S., Brooks, D., Chauhan, G., Lee, B., Lee, H.-H. S., … Hazelwood, K. (2021). Sustainable AI: Environmental Implications, Challenges and Opportunities. *arXiv:2111.00364 [cs]*. http://arxiv.org/abs/2111.00364