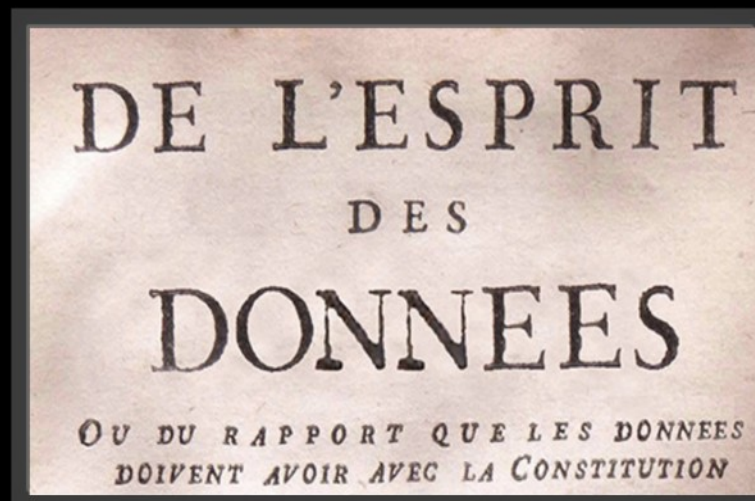


Intelligence artificielle

Vous devez répondre à un appel à projet, France 2030 ou autre, interroger l'impact environnemental de votre projet d'intelligence artificielle, mieux comprendre la taxonomie financière de l'Union européenne et comment elle s'applique à vos projets numériques?

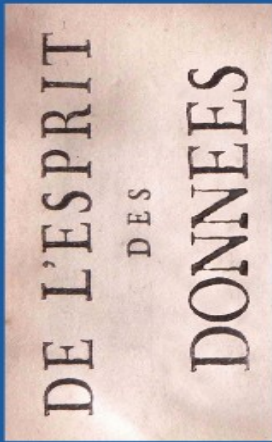


DE L'ESPRIT
DES
DONNEES

IA frugale et numérique écoresponsable, une réflexion française

De quoi va-t-on parler ?

- D'où vient l'IA frugale dans la stratégie nationale française ?
- Le contexte UE, FR et ministériel (transition écologique)
- IA frugale, numérique écoresponsable, de quoi parle-t-on ?
- Définitions Made in France
- La notion d'ordre de grandeur
- Avoir un regard au-delà
- Quelles solutions ?



Energie

Agriculture-
forêt

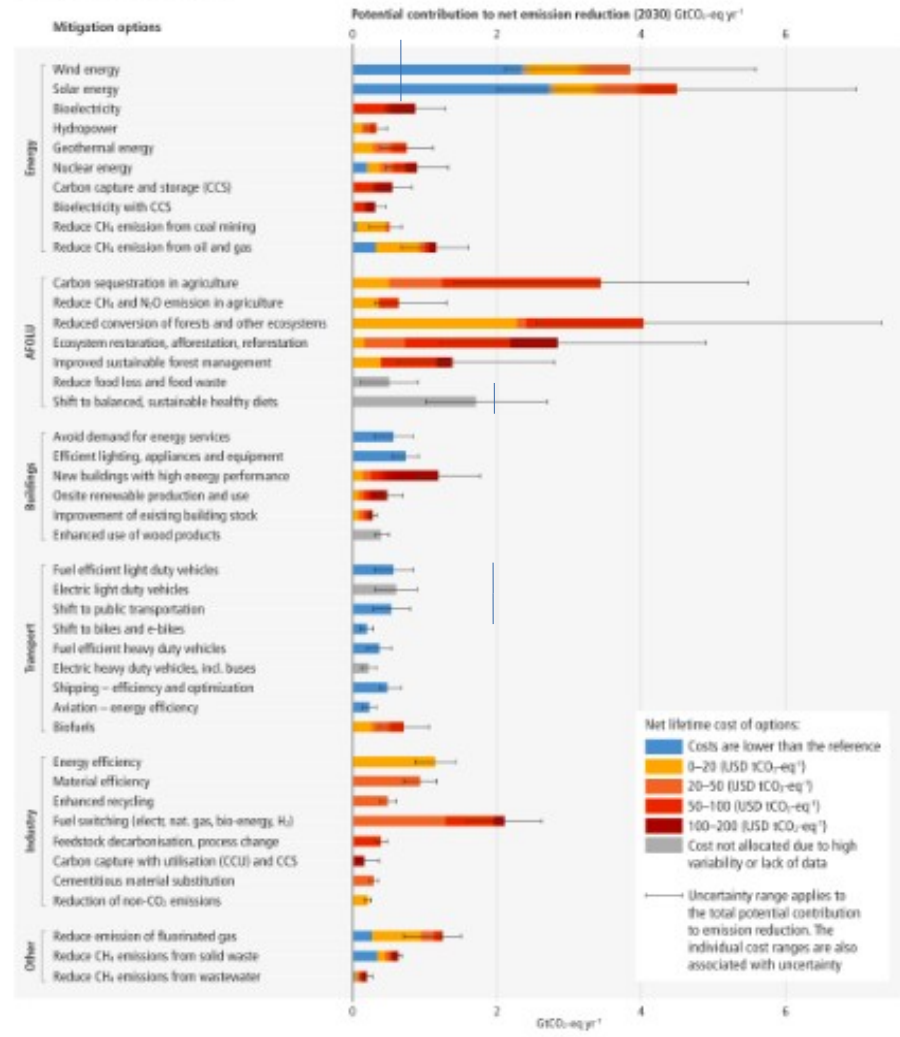
Bâtiments

Transport

Industrie

Autres
(réduction CH4...)

Many options available now in all sectors are estimated to offer substantial potential to reduce net emissions by 2030. Relative potentials and costs will vary across countries and in the longer term compared to 2030.

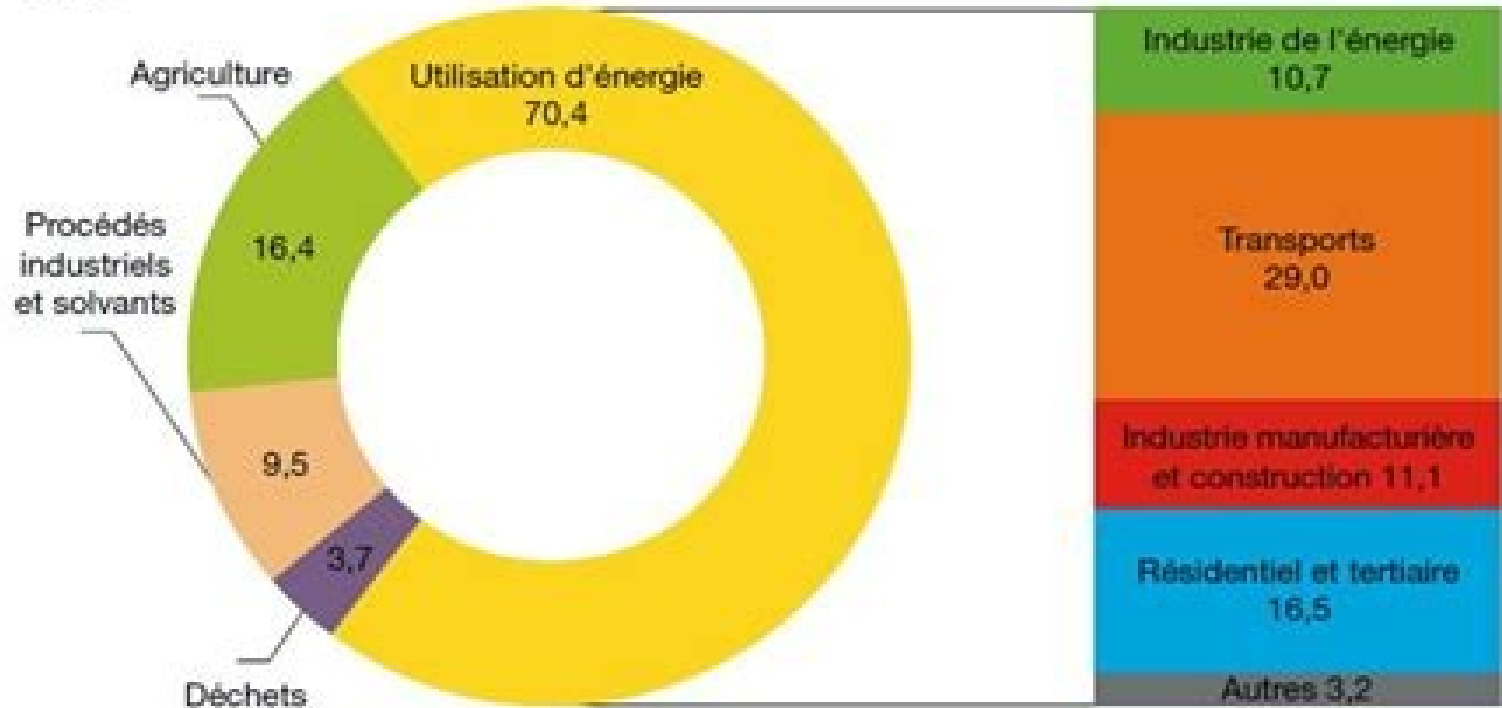


Vol.3, GIEC,
2022

La situation

RÉPARTITION DES ÉMISSIONS DE GAZ À EFFET DE SERRE EN FRANCE EN 2017²

En %



Le cadre européen

- Règlement (UE) 2020/852 du Parlement européen et du Conseil du 18 juin 2020 sur l'établissement d'un cadre visant à favoriser les investissements durables
- une activité économique est considérée comme durable sur le plan environnemental si cette activité économique:
 - a) contribue substantiellement à un ou plusieurs des objectifs environnementaux (...);
 - b) ne cause de préjudice important à aucun des objectifs environnementaux;
- Plusieurs règlements d'application dont un sur le numérique

DE L'ESPRIT
DES
DONNEES

Ne pas causer de préjudice important

1) Atténuation du changement climatique	Il a été démontré pendant l'activité que tout a été mis en œuvre pour appliquer les pratiques pertinentes énumérées en tant que «pratiques attendues» dans la version la plus récente du code de conduite européen relatif au rendement énergétique des centres de données ⁽⁶²⁴⁾ ou dans le document CLC TR50600-99-1 du CEN-CENELEC intitulé «Installations et infrastructures des centres de traitement de données— Partie 99-1: Pratiques recommandées relatives à la gestion énergétique» ⁽⁶²⁵⁾ et que l'activité a mis en œuvre toutes les pratiques attendues auxquelles a été attribuée la valeur maximale de 5 conformément à la version la plus récente du code de conduite européen relatif au rendement énergétique des centres de données.
3) Utilisation durable et protection des ressources hydrologiques et marines	Cette activité respecte les critères établis à l'appendice B de la présente annexe.
4) Transition vers une économie circulaire	L'équipement utilisé satisfait aux exigences établies par la directive 2009/125/CE pour les serveurs et les produits de stockage de données. L'équipement utilisé ne contient aucune des substances soumises à limitations visées à l'annexe II de la directive 2011/65/EU, sauf lorsque les valeurs de concentration en poids dans les matériaux homogènes n'excèdent pas les valeurs maximales énoncées dans cette annexe. Un plan de gestion des déchets est en place et garantit un recyclage maximum en fin de vie des équipements électriques et électroniques, y compris par l'intermédiaire d'accords contractuels avec des partenaires dans le recyclage, d'une prise en compte dans les projections financières ou dans les documents officiels du projet. À la fin de sa vie, l'équipement fait l'objet d'une préparation en vue du réemploi, d'opérations de valorisation ou de recyclage, ou d'un traitement approprié, y compris l'extraction de tous les fluides et un traitement sélectif conformément à l'annexe VII à la directive 2012/19/UE.
5) Prévention et contrôle de la pollution	Néant
6) Protection et rétablissement de la biodiversité et des écosystèmes	Néant

Déclinaison dans France 2030

- Réduction à « ne cause de préjudice important »

=> ne concerne que les centres serveurs

- Meilleure gestion énergétique
- Plan de gestion des déchets et de fin de vie
- Logiciels : néant

DE L'ESPRIT
DES
DONNEES

L'impact environnemental du numérique

L'empreinte carbone du numérique en France

17,2 Mt CO2 eq. soit 2,5 % de l'empreinte nationale

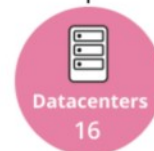
Consommation énergétique : empreinte carbone, radiations ionisantes et épuisement des ressources abiotiques fossiles

10 % de la consommation électrique française soit 48,7 TWh par an

L'épuisement des ressources abiotiques naturelles (minéraux & métaux) : représente de l'ordre de **27 % de l'impact** environnemental du numérique et est équivalent à l'extraction de 21 tonnes d'or

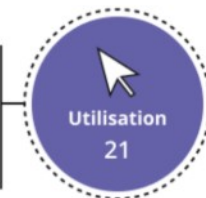
L'empreinte carbone du numérique dépend essentiellement des équipements et de leur fabrication

Répartition de l'empreinte carbone du numérique en 2020 par composantes du numérique (%)



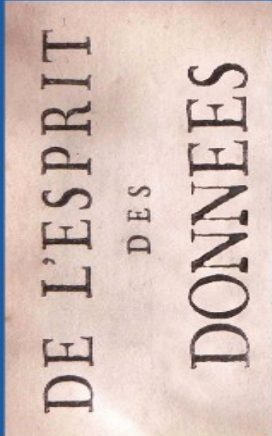
L'utilisation des équipements est responsable de 21 % des émissions du numérique et comprend l'utilisation des réseaux et datacenters

Répartition de l'empreinte carbone du numérique en 2020 par phase du cycle de vie (%)



La feuille de route française « Numérique et Environnement »

- **Connaître pour agir** : apporter des données précises et objectives sur les impacts positifs et négatifs de l'ensemble du cycle de vie des services numériques sur l'environnement.
- **Soutenir un numérique plus sobre** : maîtriser, voire réduire, l'empreinte environnementale du numérique
- **Faire du numérique un levier d'innovation pour la transition écologique** : s'appuyer sur le potentiel du numérique pour accélérer la transition écologique.



DE L'ESPRIT
DES
DONNEES

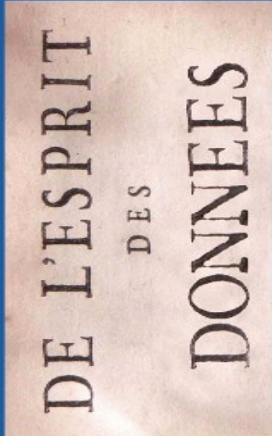
La notion de coût/bénéfice

- Toute action à un impact environnementale
- Comment mesurer le coût ?
- Comment estimer le bénéfice ?
- Comment détecter le verdissement abusif ?

DE L'ESPRIT
DES
DONNEES

L'IA frugale pour France 2030

- En donnée = « pas de big data » = < 1 To/jour
 - En pratique, c'est rarissime
- En énergie : c'est la cible principale
 - Liée au départ à l'IA embarquée (véhicules autonomes, satellites...)
 - capturée par la problématique du climat
 - Traduit en eqCO2

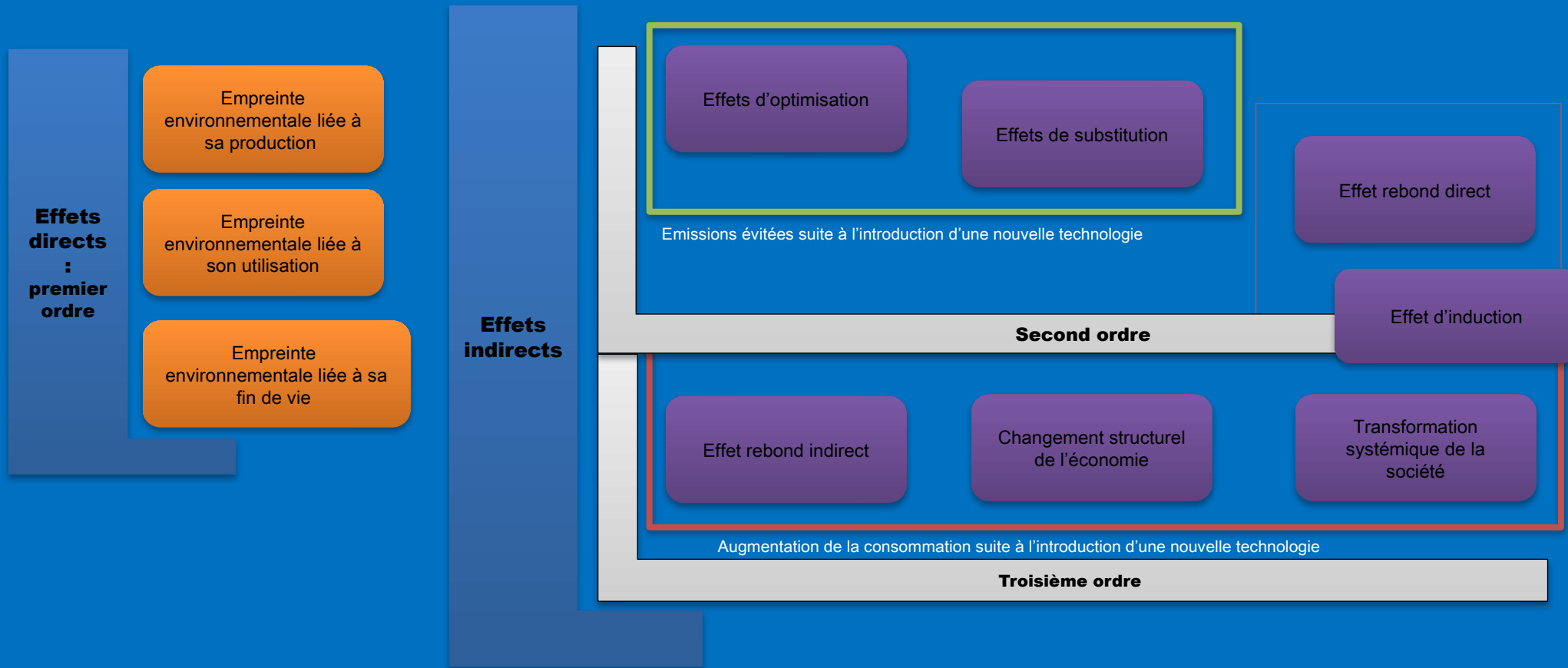


DE L'ESPRIT
DES
DONNEES

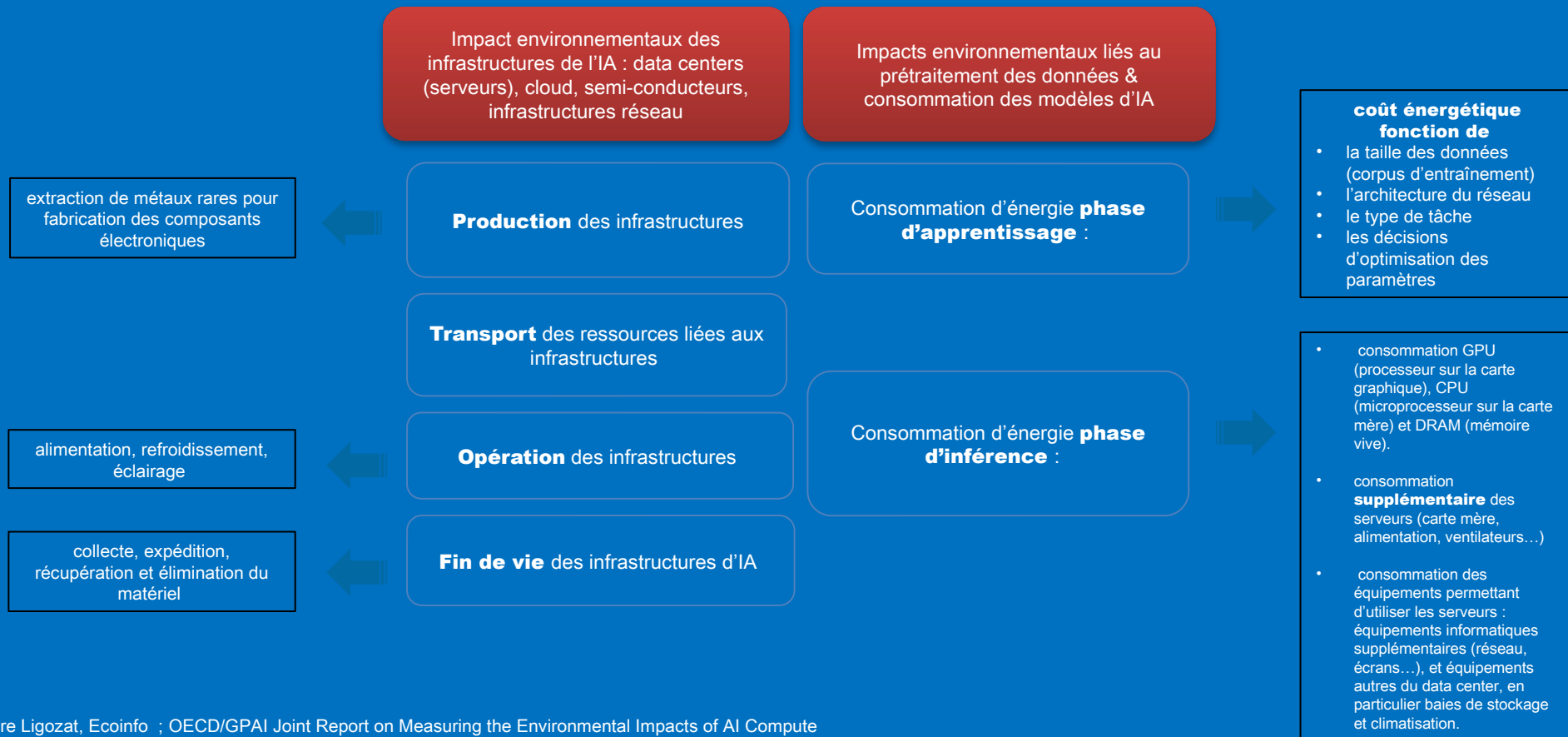
Promouvoir l'IA frugale ?

- 60 M CAD – 40 M €
- 2022 : appel à projet « IA frugale pour la transition écologique » à destination des collectivités françaises
- 4 projets retenus
- Echec de la recommandation d'estimer la consommation des systèmes d'IA
- Décembre 2022 : analyse de l'écosystème des outils en fonction des critères d'EcolInfo
- => Obligation de recourir à Green Algorithms

Cadre théorique : taxonomie d'analyse d'impact environnemental



Cadre théorique : impact environnemental d'un algorithme d'IA



$$E_{\text{total}} = \text{PUE} \times \sum E_{\text{ressource}}$$



$$\text{CO2eq} = E_{\text{total}} \times \text{FE}$$

$E_{\text{ressource}}$ = consommation ressources = CPU; GPU; DRAM

PUE : efficacité énergétique du data center ; la quantité d'énergie supplémentaire nécessaire pour faire fonctionner l'installation (par exemple, le refroidissement et l'éclairage).

Consommation électrique de l'ensemble du data center
Consommation électrique des équipements informatiques

FE = facteur d'émission de l'électricité ; permet de convertir une consommation d'électricité en émissions de gaz à effet de serre correspondantes.

Dépend :

- du mix énergétique du lieu où se situe le datacenter
- de la période à laquelle le programme d'IA est lancé

Source : Anne-Laure Ligozat,
Ecoinfo

Le calcul peut aussi être détaillé de la manière suivante

$$\text{MWh} = (\text{Hours to train} \times \text{Number of Processor} \times \text{Average Power per Processor}) \times \text{PUE}$$

MWh : consommation d'énergie
avec « Processors » : incluant tous les composants du serveur (CPU, GPU et DRAM)

$$t\text{CO2e} = \text{MWh} \times t\text{CO2e per MWh}$$

tCO2e per MWh = intensité carbone (dépend du lieu de localisation du datacenter)

Source : Patterson et al., Computer

Mesurer Estimer

- Choix de ***Green Algorithms (GT)***, Lannelongue et al
<https://www.green-algorithms.org/>
- Site en ligne gratuit,
- prise en compte du mix énergétique,
- estimation du nombre d'inférence
- Github réactif

DE L'ESPRIT
DES
DONNEES

Quelques évaluations

- 2023 : suivi de deux projets pour la transition écologique, 5 SIA entre 0,4 gr eqCO₂, 1 SIA à 2,7 tonnes (3D).
- Pour un gain estimé à terme de 1000 t/an.

DE L'ESPRIT
DES
DONNEES

Coût des IA ?

DE L'ESPRIT
DES
DONNEES

- Estimation du coût énergétique software-based : entre 18kg eqCO2 et 284 000 kg eqCO2 [environ un vol A-R Paris-NY]"
- Les plus gros modèles émettent l'équivalent d'un vol Paris-Hong-Kong
https://ecoinfo.cnrs.fr/wp-content/uploads/2021/04/Cours_CentraleSupelec_IA_responsable-1.pdf
- En 2019, 41 268 dessertes FR-USA, et N-Y est au plus proche
- Et 2200 dessertes de Hong-Kong.

Bénéfices des IA ?

- Mieux comprendre pour mieux agir :
- simulation des écosystèmes complexes
- > l'eau, la ville
- Traitement du langage : extraction des données dans les rapports en pdf, analyse des textes juridiques sur un même lieu,
- Imagerie : traitement des données satellitaires pour la forêt, la sécheresse des sols, la fréquentation des zones naturelles...

DE L'ESPRIT
DES
DONNÉES

La sobriété dans la conception

- Un modèle géant pour économiser lors des applications ?
- Des modèles hybrides ?

A vertical rectangular graphic with a light beige background and a dark blue border. The text "DE L'ESPRIT" is written vertically on the left, "DES" is in the center, and "DONNEES" is written vertically on the right.

DE L'ESPRIT
DES
DONNEES

La sobriété dans la conception

- Journées de Recherche en Apprentissage Frugal 2023
 - <https://jraf-2023.sciencesconf.org/>



DE L'ESPRIT
DES
DONNEES

Quelques autres pistes

■ Frugalité via les *architectures* d'apprentissage

- inférence intrinsèquement frugale ...
 - notamment via parcimonie & butterflies
- ... mais aussi entraînement intrinsèquement frugal
 - distribué, binarisation, sketching (of gradients, vectors, datasets) ...

$$A \approx \begin{matrix} \square \\ \diagdown \\ \square \end{matrix} \times \begin{matrix} \square \\ \diagdown \\ \square \end{matrix} \times \begin{matrix} \square \\ \diagdown \\ \square \end{matrix} \times \begin{matrix} \square \\ \diagdown \\ \square \end{matrix}$$

■ Frugalité via des *principes*

- limiter le gâchis du "end-to-end learning" !
- **incorporer nos connaissances durement acquises**
 - modèles physiques, connus ou paramétrés, e.g. via des EDP
 - symétries et invariances, au-delà de l'augmentation des données
 - graphes, bases de connaissance
 - apprentissage "quality-aware"; sélection des données pertinentes
- *effets bénéfiques sur* "interprétabilité", confidentialité, robustesse



Projet SHARP
PEPR IA 2023-2027

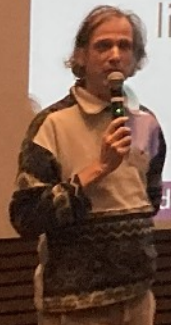
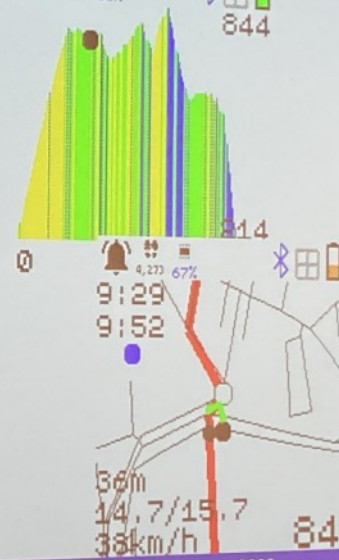
Proposition ALLyS
IA cluster Lyon



R. GRIBONVAL
IRAF - 13-14 décembre 2023 - Grenoble

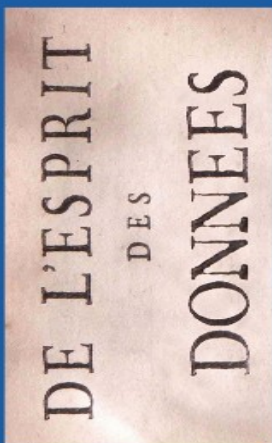
DE L'ESPRIT DES DONNEES

- 1500 lignes de javascript
 - espruino : interpréteur javascript pour l'embarqué (gordon williams)
 - le code js est plus petit que le code binaire correspondant
- entrée fichier gpx
 - suite de coordonnées en xml
- enrichissement avec openstreetmap
- pré-processing en rust (3500 lignes) compilé vers wasm



Environmental impacts of computing in health & life sciences research

7/11/2023



- <https://www.eicworkshop.info/#resources>.
- 9 présentations dont :
- Reducing the carbon footprint of digital pipelines: A case study/Nick Souter, University of Sussex
- Addressing environmental impacts in the lab/ Martin Farley, GreenLab Associates University College London
- The carbon footprint of high-performance computing/Loïc Lannelongue, University of Cambridge

La sobriété dans l'entraînement : entraîner « juste ce qu'il faut »

Via Nick Souter in https://www.eicworkshop.info/slides/PDF/7_eic_souter.pdf

Li & Chao (2021)

When classifying crop pests,
peak training accuracy could
be achieved with:

- 75% of total data (shallow model)
- 60% of total data (middle model)
- 40% of total data (deep model)

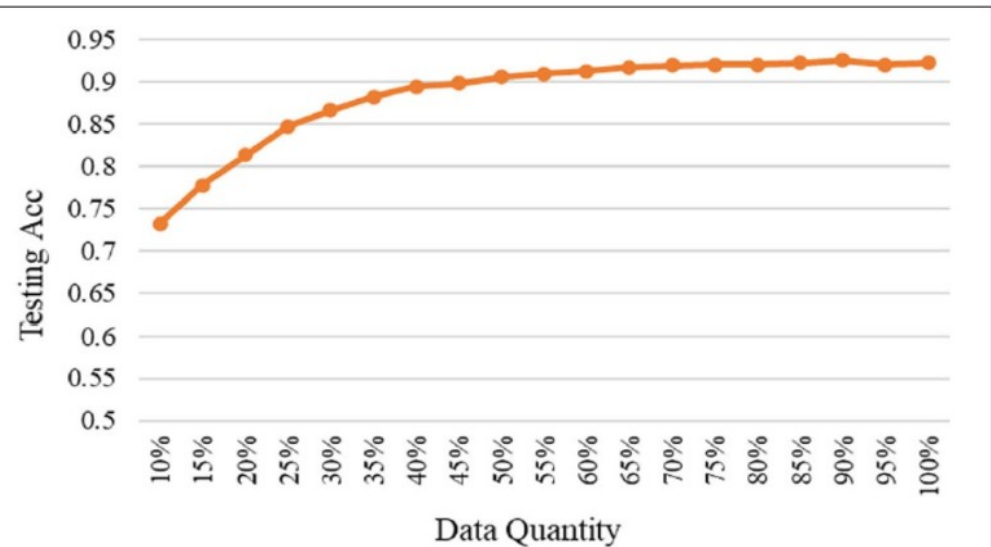
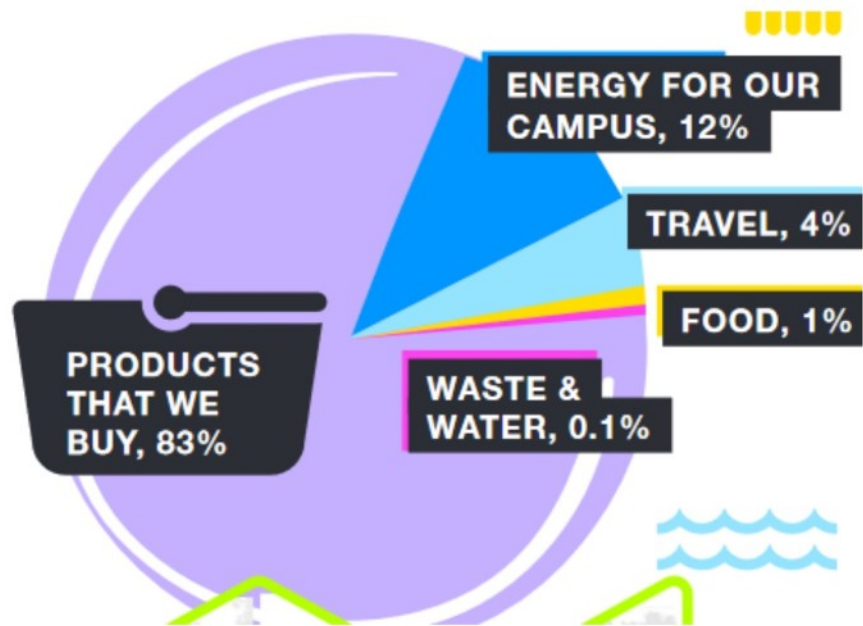


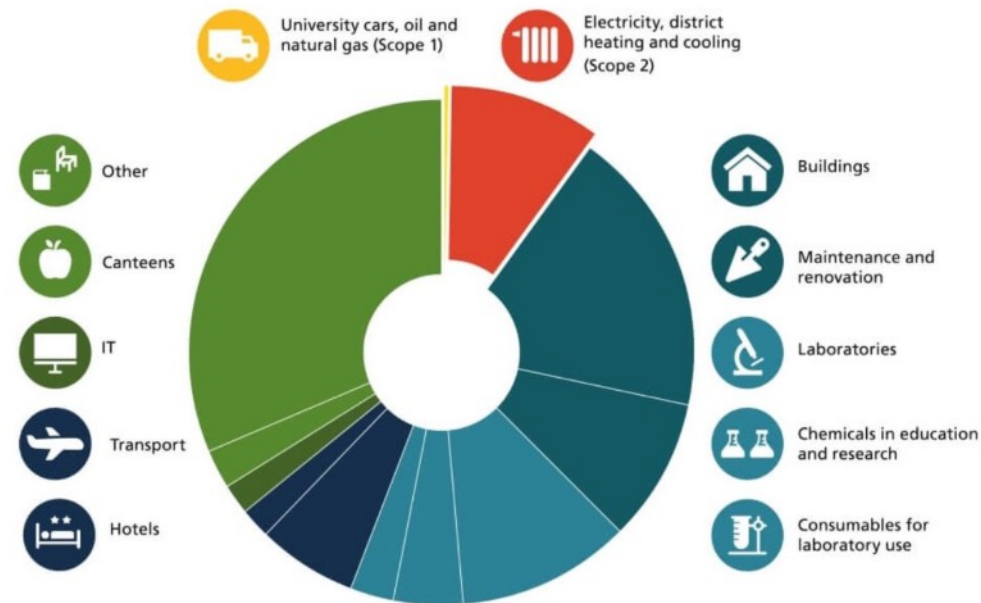
FIGURE 4 | The relation between accuracy and data quantity under the middle model.

La sobriété au-delà du numérique - Analyse du labo, où sont les gains faciles ?

Total CO2 Emissions from 2 European Institutions



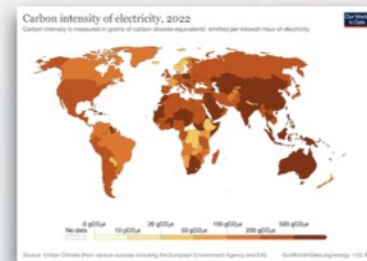
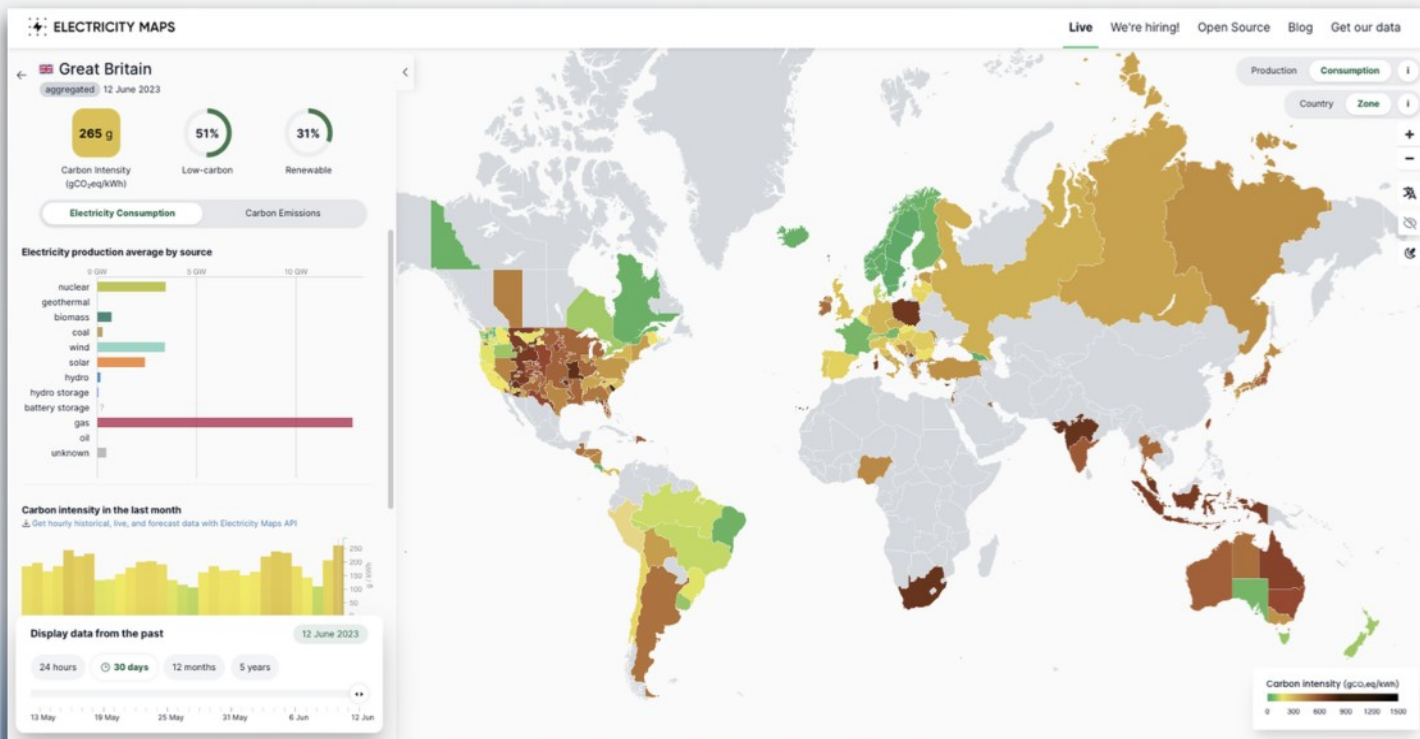
UCL, UK



University of Copenhagen, Denmark

THE CARBON FOOTPRINT OF COMPUTATION: CARBON INTENSITY

www.electricitymap.org



DE L'ESPRIT

L'IA générative, un nouveau paradigme ?

- ChatGPT : 1,3 GWh
- Ses inférences : x 10 (in Alex de Vries, The growing energy footprint of artificial intelligence)
- Production mondiale d'électricité 2021: 28 520 TWh
(agence internationale de l'énergie)
- Le coût complet d'un LLM pourrait être x 5 (Sasha Luccioni, Sylvain Viguié, Anne-Laure Ligozat in Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model)

Power Hungry Processing: ⚡ Watts ⚡ Driving the Cost of AI Deployment?

ALEXANDRA SASHA LUCCIONI and YACINE JERNITE, Hugging Face, Canada/USA

EMMA STRUBELL, Carnegie Mellon University, Allen Institute for AI, USA

« multi-purpose, generative architectures are orders of magnitude more expensive than task-specific systems for a variety of tasks, even when controlling for the number of model parameters. »

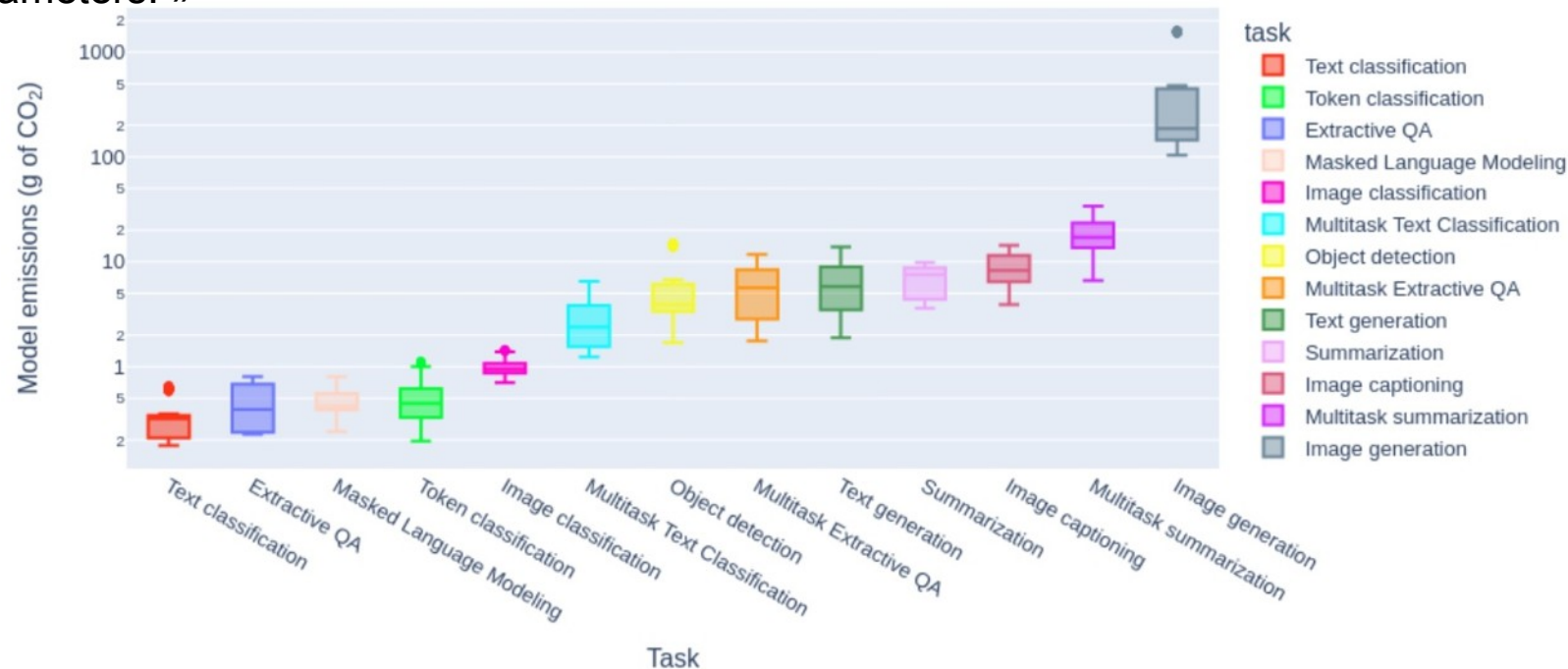


Fig. 1. The tasks examined in our study and the average quantity of carbon emissions they produced (in g of CO₂) for 1,000 queries. N.B. The y axis is in logarithmic scale.

DE L'ESPRIT
DES
DONNEES

Pour aller plus loin

- Améliorer l'analyse du cycle de vie des centres serveurs
 - <https://ecoinfo.cnrs.fr/>
 - <https://boavizta.org/>
- Dépasser eqC02 pour kWh, consommation des ressources et impact biodiversité

Conclusion ?

We believe this resolves all remaining questions on this topic. No further research is needed.

References

1. [Illegible]
2. [Illegible]
3. [Illegible]
4. [Illegible]

JUST ONCE, I WANT TO SEE A RESEARCH PAPER WITH THE GUTS TO END THIS WAY.

Sadly not yet

*So dedicated research efforts
are needed*

DE L'ESPRIT
DES
DONNEES