Travail effectué du 2 au 13 mai 2022 :

1. Etude de différents articles citant l’article de référence : Strubell, E.; Ganesh, A.; McCallum, A. Energy and Policy Considerations for Deep Learning in NLP. arXiv 2019, arXiv:190602243.
2. Finir le modèle de régression
3. Recherche sur le cycle de vie des sources d’énergie (éolienne)
4. Etude de différents articles :

Sustainable AI: AI for sustainability and the sustainability of AI

La problématique :

Comment le concept d'IA durable peut-il être défini et opérationnalisé pour guider les décisions du secteur public ?

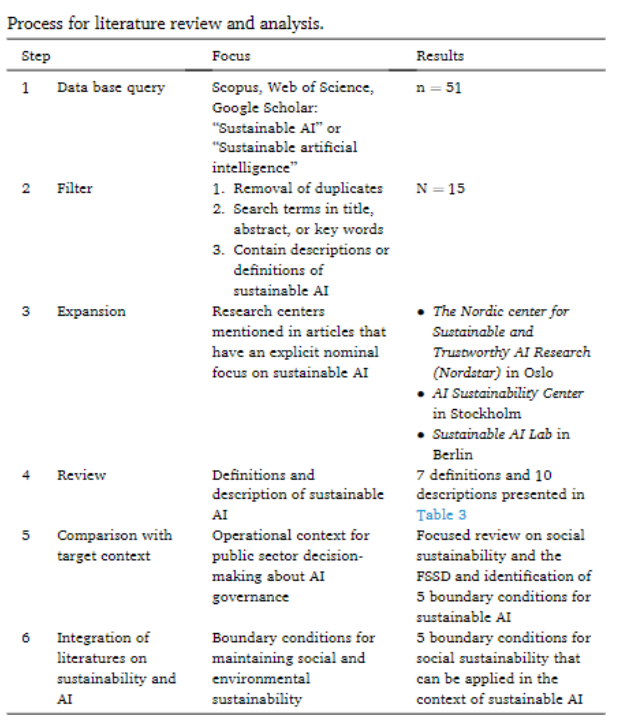
La méthode :

Literature review :

Les auteurs de l’articles commencent par présenter le contexte de la prise de décision du secteur public en matière de gouvernance de l'IA ainsi que la façon dont le concept de durabilité a été élaboré dans le contexte du développement durable et des objectifs du développement durable.

* FSSD - Framework for Strategic Sustainable Development a été mis en place dans une différents contextes et avec une variété d'acteurs du secteur public et à but non lucratif, testant son utilité de donner « des orientations sur la manière dont toute région, organisation ou projet peut développer une vision encadrée par des principes de durabilité sociale et écologique, analyser et évaluer la situation actuelle par rapport à cette vision et ainsi clarifier l'écart, générer des idées d'actions possibles qui pourraient aider à combler le fossé, et classer ces actions par ordre de priorité dans un plan progressif et économiquement attractif. »

L’article présente ensuite l’approche de recherche.



La section 4 de l’article présente les résultats de la revue de la littérature des auteurs et compare les différentes façons dont l’IA durable est conceptualisée et abordée dans ces différents articles.

Une image contenant texte

Description générée automatiquementUne image contenant texte

Description générée automatiquement

Par la suite, les auteurs identifient et analyse les conditions limites nécessaire de l’IA en intégrant la littérature sur la durabilité et sur l'IA du secteur public afin d'aboutir à un modèle conceptuel préliminaire d'IA durable pour le secteur public.

Les 5 conditions limites :

1. La diversité : éviter la dégradation de la durabilité sociale par la monopolisation des systèmes d'IA par les élites ou l'exclusion des parties prenantes concernées de la gouvernance de l'IA.
2. La capacité d’apprendre : percevoir les changements et d'y répondre efficacement et comprend la capacité d'apprendre de l'expérience, en tant que mécanisme
3. La capacité d’auto-organisation : relire cette partie car pas très bien comprise
4. Un sens commun : s'assurer que les systèmes d'IA n'incarnent pas ou ne manifestent pas de valeurs contraires aux valeurs sociétales ou aux valeurs des groupes de parties prenantes concernés.
5. La confiance : veiller à ce que l'IA ne soit pas conçue et utilisée d'une manière qui, si elle était connue, trahirai la confiance dans cette IA ou dans les systèmes et services dans lesquels elle est intégrée.

Contribution personnelle :

La section 5 discute de l'analyse et présente le modèle intégré des conditions limites de l'IA durable, ainsi que des considérations opérationnelles sur la façon dont il peut être appliqué dans le processus décisionnel du secteur public. La dernière partie de cette section présente quelques remarques finales ainsi qu'un aperçu des limites de cette étude.

Une image contenant texte

Description générée automatiquement

Critique personnelle :

Cet article parle surtout de l’impact sociétal de l’IA dans les différents secteurs et ne prends donc pas du tout en compte la partie environnemental ou économique de L’IA durable. Aucune application de ce model donc aucun résultat

Perspectives d’amélioration : Intégrer les questions environnementales à son modèle

RESUMER LE TABLEAU ASPECT SOCIAL POUR LE DEV DURABLE POUR METTRE EN PLACE DES LIMITES

A Holistic Approach for Designing Carbon Aware Datacenters

La problématique :

Avoir un carbon footprint nul en utilisant seulement des énergies renouvelables est très dur car les énergies renouvelables ne sont pas constantes et il arrive donc que parfois elles ne sont pas suffisantes poussant les datacenters à utiliser des énergies à forte émission de carbone. Ces datacenters deviennent donc low ou 0 carbone sur la base annuelle mais pas sur la base horaire.

* Cet article souhaite proposer un cadre et une solution 100% 0 carbone 24/7

La méthode :

1. Collecte et évaluation des données à une granularité fine, horaire, sur les réseaux électriques à différents endroits géographiques : les auteurs collectent d'abord la consommation électrique horaire des centres de données Meta pour toutes les régions des États-Unis. Ils préparent également les données de production d'énergie horaire pour chaque région géographique au niveau de l'autorité d'équilibrage (BA), où se trouvent les parcs de centres de données Meta. Ils ont aussi mené une analyse de demande et d'offre d'énergie des centres de données à l'échelle industrielle et les intensités de carbone qui en résultent

Une image contenant table

Description générée automatiquement

Cette étude a permis de dégager 3 scénarios permettant aux datacenters qui suivent ces scénarios de progresser considérablement vers des opérations sans carbone sur une base horaire.

* + Mix grid : fournit une baseline et est particulièrement pertinent pour les petits centres de données qui n'ont pas encore personnalisé leurs achats d'énergie.
  + Net Zero : décrit comment les investissements dans les énergies renouvelables réduisent considérablement l'intensité carbonique des opérations des centres de données
  + 24/7 Carbon Free : décrit comment des investissements supplémentaires et des programmateurs de réponse à la demande pourraient réduire davantage l'intensité de carbone.

1. L’évaluation équitable et comparative des différentes solutions alternatives :
   * Déploiement des énergies renouvelables : suite à une analyse de la couvrance de la consommation 24/7 en fonction des investissements en énergie solaire et éolienne, on voit qu’on atteint toujours un plateau et que l’investissement nécessaire pour atteindre 100% de la demande est très important, ne fait qu’augmenter plus on s’approche de 100% et dépend de la localisation. Il faut donc utiliser une méthode de stockage
   * Déploiement du stockage de l'énergie : Utiliser des batteries lithium-ion habituellement utilisé pour les périodes de maintenance plus souvent pour atteindre l’objectif 24/7 en les rechargeant lors de surproduction d’énergie et en les utilisant lors des sous productions. Cette solution permettrait aux régions hybrides (solaire éolienne) d’atteindre 24/7 carbon free énergie opérationnelle dans le datacenter de l’Utah de Meta.

Une image contenant table

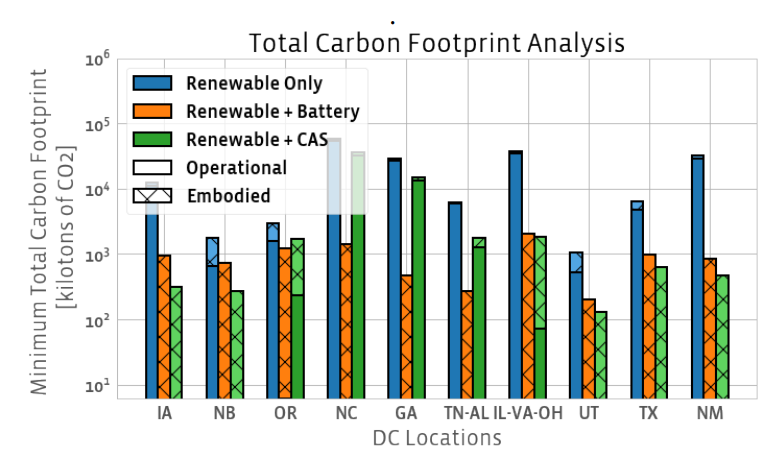
Description générée automatiquement

* + Programmation respectueuse du carbone : Cela consiste à faire tourner les machines pendant les périodes ou l’intensité carbone est la plus basse ou encore lorsque qu’on est en surproduction d’énergie et non lorsqu’on est en sous production.

1. La prise en compte de l'empreinte carbone opérationnelle et intrinsèque dans les analyses : limiter l’empreinte carbone opérationnelle s'accompagne souvent d'un coût supplémentaire non négligeable en termes de carbone incorporé. Ainsi, pour obtenir des datacenters nul en émission carbone, il faut prendre en compte l'empreinte carbone opérationnelle et l'empreinte carbone intrinsèque de manière holistique afin de minimiser l'empreinte carbone globale des centres de données. Énergie renouvelable seule non suffisante ; zone avec seulement énergie solaire ont souvent l’empreinte carbone la plus élevée ; RENOUVELABLE + programmation respectueuse du carbone est très efficaces pour les régions d’énergie hybride ou majoritairement éolienne

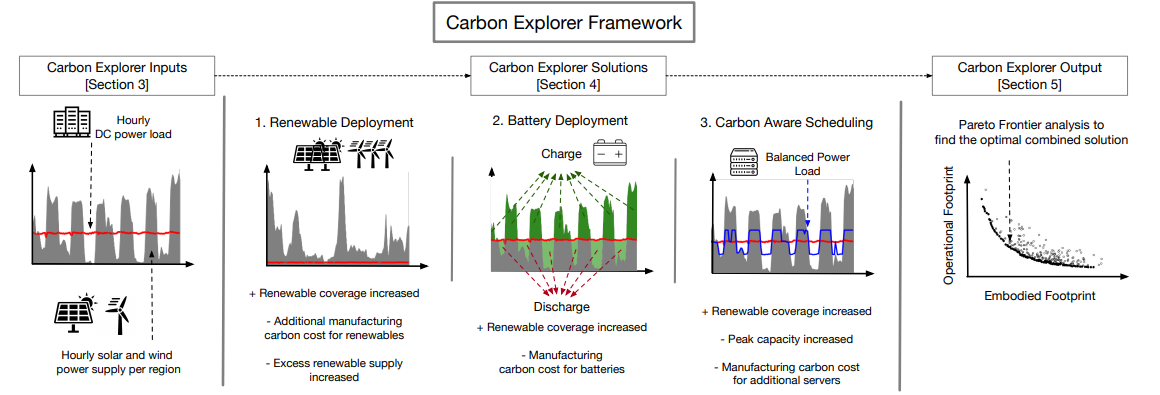
* Les solutions dépendent donc des caractéristiques de production d'énergie renouvelable des régions où se trouvent les centres de données.

Résultats :



* Les solutions Renouvelable + Batterie et renouvelable + programmation respectueuse du carbone peuvent offrir une réduction d'un ordre de grandeur de l'empreinte carbone totale des centres de données. Comme les batteries ne sont pas limitées par la flexibilité de la charge de travail comme renouvelable + CAS, renouvelable + batterie permet d'atteindre une couverture opérationnelle de 100 % pour tous les centres de données. En outre, le déploiement de batteries est la seule solution viable pour atteindre une couverture 24/7 pour le centre de données dans les régions où l'énergie solaire est absente. Enfin, RENOUVELABLE + CAS offre une alternative compétitive à Renouvelable + Batterie dans la plupart des régions uniquement éoliennes et hybrides.

Contribution personnelle : Carbon explorer



L'impact des résultats :

Critique personnelle : énergies renouvelables limité à l’éolienne et au solaire, pas de recherches poussées sur la façon de stocker l’énergie, seulement une sorte de batterie proposée, pas d’infos sur les installations utilisées dans les différents datacenters (?)

Perspectives d’amélioration : élargir les données sur lesquelles l’étude est faite plus de régions, plus de méthode de stockage

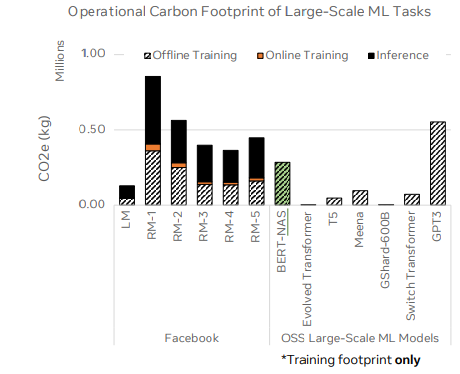
Sustainable AI: Environmental Implications, Challenges and Opportunities

La problématique :

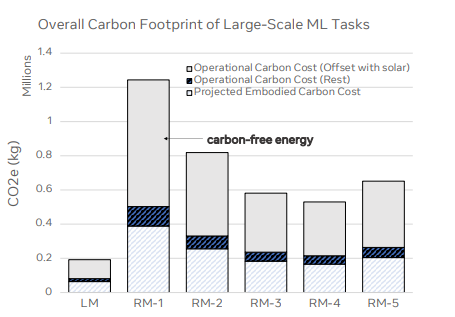
Présenter une analyse de bout en bout de l'empreinte carbone opérationnelle et intrinsèque pour la formation et l'inférence de l'IA.

La méthode :

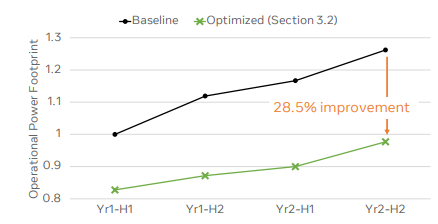
1. Caractériser l'empreinte carbone de l'IA en étudiant le cycle de développement des modèles dans les cas d'utilisation de l'apprentissage automatique à l'échelle industrielle chez Facebook.
2. Analyse de l’empreinte carbone de différents modèles d'apprentissage automatique représentatifs utilisés chez Facebook et de 7 open-source modèles de grande échelle (Leurs tâches sont basées sur les architectures de modèle « vanille » et peuvent ne pas refléter les cas d'utilisation en production.)



* L'empreinte carbone du modèle LM est dominée par l'inférence, tandis que, pour RM1 à RM5, l'empreinte carbone de la formation et de l'inférence est à peu près égale. L'empreinte carbone moyenne des tâches de formation du modèle LM chez Facebook est 1,8 fois plus importante que celle de Meena et 0,3 fois plus importante que l'empreinte carbone de GPT-3

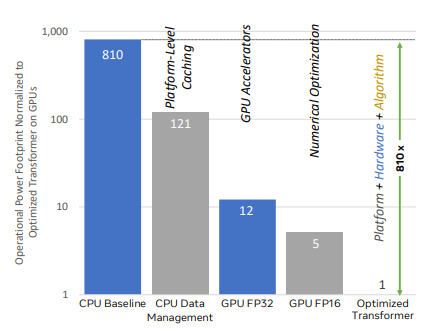


* Si on prend en compte l’offset c’est le coût du carbone de fabrication qui est la source dominante de l'empreinte carbone de l'IA.

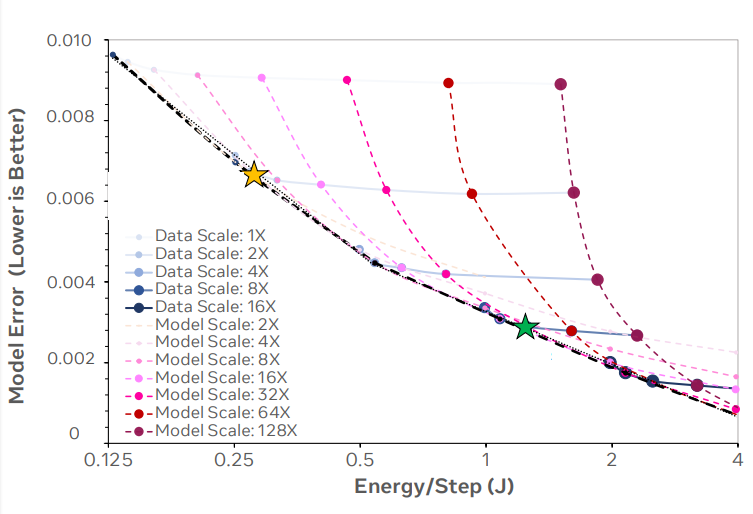
1. Optimisation de l’empreinte carbone à partir d’un co-design hardware-software. Réduction de l'empreinte énergétique du stack hardware-software de l'apprentissage automatique de 20 % tous les six mois.
2. 

* L'effet net, avec le paradoxe de Jevon, est une réduction de 28,5 % de l'empreinte énergétique opérationnelle sur deux ans. Mais l'infrastructure d'IA continue de se développer et la demande augmente.

1. Réduction de plus de 800 fois de l'empreinte carbone opérationnelle obtenue grâce à 3 choix de co-design hardware-software judicieux pour un modèle de langage universel basé sur Transformer.



* + Platform-level Caching : le pré-calcul et la mise en cache d'intégrations fréquemment utilisées. En utilisant des dispositifs de stockage RAM et Flash comme caches permet un partage entre les applications et les cas d'utilisation.
  + GPU acceleration : utilisé du hardware IA spécialisé basé sur les GPU
  + Algorithmic optimization : Réduire la précision de moitié, utiliser des opérateurs personnalisés pour planifier les étapes de codage dans un seul noyau du module Transformer

1. « L'efficacité devrait être un critère d'évaluation pour la publication de la recherche ML sur les modèles à forte intensité de calcul, au-delà des mesures liées à la précision. »
   * Data Utilization Efficiency : dégrader un petit peu la qualité des modèles peu permettre une très bonne économie d’énergie, les datas perdent leur valeur prédictive et cela augmente fortement les émissions carbones, prédire le temps de demi-vie des données, permet de réduire les besoins en ressources pour le pipeline de stockage et d'ingestion des données, ce qui diminue le temps de formation, ainsi que les besoins de stockage
   * Experimentation and Training Efficiency : Utilisation de méthodes NAS et HPO beaucoup plus efficaces en termes d'échantillons. Réduction du nombre d'expériences de formation, détection et arrêt précoce des flux de formation peu performants
   * Efficient, Environmentally-Sustainable AI Infrastructure and System Hardware : les GPU ne sont pas utilisé entièrement 30-50% chez facebook il faut donc utilisé la virtualisation et la multi-location pour les accélérateurs d’IA car il permet une meilleure utilisation des ressources , repenser la conception des systèmes hardware pour minimiser l'empreinte carbone en prenant en compte le cycle de vie complet du développement du matériel et du modèle ML, stocker l’énergie, éviter le on-device learning même si il permet une meilleure confidentialité

Résultats : Même recommandation

Contribution personnelle :

L'impact des résultats :

Critique personnelle :

1. Recherche sur le cycle de vie des sources d’énergie (éolienne)

L’éolienne :

Taux d’émission durant la vie entière d’une éolienne calculé par l’ADEME :

* Eolienne terrestre : taux d'émission de 12,7 g CO2 eq / kWh
* Eolien en mer : taux d'émission de 14,8 g CO2 eq / kWh

1. Mise en place des éoliennes
   1. Matériaux nécessaires pour construire des éoliennes :

En moyenne, pour chaque mégawatt de puissance installée, un aérogénérateur nécessite 169 tonnes de métaux divers et 455 tonnes de béton et de ferrailles pour les fondations.

Une éolienne peut se séparer en plusieurs parités qui sont les suivantes :

* Le rotor

Une image contenant texte, table

Description générée automatiquement

* La nacelle

Une image contenant table

Description générée automatiquement

Les aimants actuellement utilisés sont de type néodyme-fer-bore (NdFeB) qui utilise deux terres rares1 : le néodyme et le dysprosium.

* Le mât : soit en acier (88% des cas) soit en béton avec le bout du mât en acier
* Les fondations : Pour une éolienne d'une puissance de 3MW, 800 tonnes de béton armé, 40 tonnes d'acier et 400 tonnes de terre recouvrant le radier (correspond à la consommation en béton d'un immeuble de 12 appartements). Les fondations dépendent énormément de la nature du sol, du vent et de la présence de nappe d'eau
  1. Transport :

L’incertitude sur le transport est très importante. L’étude menée par l’ADEME se base sur les critères suivants :

* Transport associé à l’importation des composants sur le site des assembliers : on suppose une distance de transport de 600 km par camion des composants vers les assembliers.
* - Transport sur site de l’installation : on suppose différentes distances de transport selon les composants de l’éolienne et un type de transport camion pour chaque cas (nacelle : 1025 km ; rotor : 1025 km ; tour : 600 km ; fondation : 50 km ; autres : 600 km).
  1. Assemblage :

Ici aucune donnée disponible sur l’utilisation d’une grue de levage, l’hypothèse d’Ecoinvent (Bastien Burger, 2007) suivante a été utilisée : 0,5 kWh par kilogramme assemblé d’éolienne

* 1. Construction des routes d’accès :

L’hypothèse d’Ecoinvent (Bastien Burger, 2007) de 10 kilomètres de route par parc éolien est utilisée. -> 5810 kilomètres de routes construites pour l’accès aux différents parcs en France

* 1. Raccordement réseau :

2 câbles

La distance entre chaque raccordement varie selon les parcs ; après lecture des rapports ACV des industriels, nous prenons l’hypothèse d’une distance d’un kilomètre entre l’éolienne et le transformateur. Le câblage après le transformateur est d’après le périmètre de l’étude non pris en compte.

D’après les données Vestas (Vestas, 2006), les câbles 32 kV internes au parc sont constitué à 40 % d’aluminium, 36 % de thermoplastiques isolant (PE) et de 24 % de cuivre et représente une masse d’une tonne de matériaux par kilomètre.

La totalité du raccordement réseau représente 3658 kilomètres de câbles pour une masse de 3658 tonnes.

1. Production d’énergie électrique

Pour le calcul des impacts environnementaux de la production d’électricité par des éoliennes de différentes puissances, il est nécessaire de modéliser la production d’électricité de chaque installation de l’atlas français.

La méthode de calcul pour la production d’électricité durant une année se traduit par la formule suivante : Production électrique (kWh) = 8760 × L ×P

* D’après (RTE, 2015) en France, le facteur de charge annuel moyen des éoliennes terrestre sur la période 2010 – 2014 est de 22,9%
* Les puissances nominales des éoliennes sont issues de la base de données des éoliennes installées en France

1. Maintenance et réparation

La maintenance représente en moyenne un changement de 15% de la nacelle et d’une pale tout au long de la durée de vie de l’éolienne (Padey, 2013).

Le transport des pièces de remplacement jusqu’au parc éolien reprend les mêmes hypothèses que le transport des pièces au premier assemblage de l’éolienne.

Pour le transport associé à la maintenance préventive et corrective ne nécessitant que le déplacement d’un technicien une distance annuelle de 2160 km en camionnette et considéré par l’ADEME.

1. Démontage et fin de vie

On reprend le 0,5 kWh par kilogramme d’éolienne pour démonter l’éolienne.

|  |  |  |
| --- | --- | --- |
| Matériaux | Scenario | Référence |
| Acier | 90 % recyclé, 10 % enfouis | (Elsam Engineering, 2004) |
| Fonte | 90 % recyclé, 10 % enfouis | (Martinez et al. 2009) |
| Cuivre | 90 % recyclé, 10 % enfouis | (Vestas. 2006) |
| Aluminium | 90 % recyclé, 10 % enfouis | (Vestas. 2006) |
| Plastiques | 100 % incinéré | (Schleisner, 2000) |
| Béton | 100 % recyclé | (Martinez et al. 2009) |
| Fibre de verre, époxy | 100 % incinéré | (Milanese, 2009) |
| Aimant permanet | 100 % enfouis | Cycleco |

Travail à faire pour la semaine prochaine :

* Continuer l’analyse du cycle de vie pour les autres énergies surtout solaire
* Voir et comprendre l’économie transactionnelle et régulée, les libertariens, le concept de volatilité et de la valeur réelle
* Faire une synthèse des articles et son tableau un améliorant l’analyse les articles en particulier la partie critiques personnelles