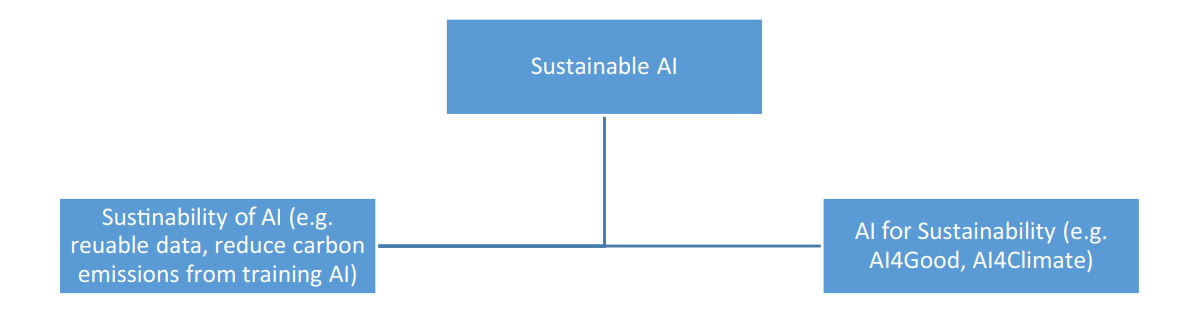
Travail effectué du 19 au 29 avril 2022 :

* 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.
* Apprentissage du python en autonomie

Sustainable AI: AI for sustainability and the sustainability of AI

La problématique :

Présenter le concept d'IA durable comme un terme générique pour couvrir deux branches ayant des objectifs et des méthodes différents : l'IA pour la durabilité et la durabilité de l'IA.



La méthode :

Dans un premier temps, l’article montre que l’IA durable est très prometteuse. L’IA durable serait la combinaison entre utiliser l’IA pour le développement durable et rendre l’IA la plus durable possible. Cependant, dans cet écrit, nous apprenons que des efforts ont déjà étais faire pour utiliser l’IA pour le développement durable mais que ces efforts ont un prix et qu’il est d’abord nécessaire de se concentrer sur la durabilité de l’IA, aspect très peu étudier et rehcercher.

* La durabilité de l’IA est une branche de l'IA durable qui s'intéresse à la manière de mesurer la durabilité du développement et de l'utilisation des modèles d'IA, par exemple en mesurant l'empreinte carbone, la puissance de calcul des algorithmes d'entraînement, etc. En d'autres termes, l'IA pour le développement durable ne peut être réalisée sans aborder simultanément la durabilité de l'IA.

L’article explique ensuite ce que l’IA durable devrait être en commencer par expliquer la notion de durabilité.

* L'IA durable doit donc prendre en compte les innovations en IA afin d'atteindre les objectifs du développement durable mais aussi cibler la durabilité de sa formation et de son utilisation. Le débat mondial sur l'IA doit explicitement aborder, non seulement les droits de l'homme et/ou les questions éthiques, mais aussi la tension entre les besoins de l'environnement, de l'économie et de la société.

Par la suite, l’article présente des outils qui existent déjà permettant de sensibiliser les gens comme AI for Climate or AI4Good. Ici encore, l’auteur reconnait l’importance de ces outils mais il souligne tout de même le fait que pour lui il faudrait se focaliser sur la durabilité de l’IA. Il continue dans cette lancé en présentant cette fois ci des articles partageant le même avis comme celui de Strubell qu’il utilise pour montrer que bien que l’IA devrait utiliser des énergies renouvelables, la priorité serait devrait peut-être pour les foyer qui n’ont pas accès à l’électricité moderne plutôt que prioriser l’entrainement d’algorithme qui ont pour but de battre des records de certains jeux.

* Les faits énoncés par Strubell sont essentiels pour permettre aux décideurs politiques de prendre des décisions concernant la proportionnalité de certaines méthodes d'IA par rapport à leur application prévue. En d'autres termes, il est temps que ces derniers régissent l'IA à un niveau plus détaillé et suggèrent que certaines méthodes. Par exemple la mise au point d'un modèle NLP d'apprentissage profond, ne devraient pas être autorisées pour des tâches éthiques telles que le recrutement de nouveaux employés ou la prédiction d'employés sur le point de démissionner. La raison en est que les coûts environnementaux sont tout simplement trop élevés pour justifier une application aussi subalterne (sans parler des problèmes éthiques).

L’article rejoint encore Strubell sur le fait que les chercheurs devraient tous rapporter le temps de formation et la sensibilité aux hyperparamètres dans leurs articles. L’auteur ajoute qu’il existe aujourd’hui des outils capables de calculer nos consommations mais que très peu de personnes les utilise. Il existe même un outil capable de prédire la consommation d’un entrainement et arrêter se dernier s’il dépasse les prédictions.

* Cet outil devrait être imposer par les décideurs de politiques

Enfin, l’auteur explique sa solution pour inscrire l’IA dans le développement durable :

Les contributions, les nouveautés : révision des articles, les recommandations et la définition de l’IA durable en deux parties durabilité de l’IA et L’IA pour le développement durable

Les résultats :

1. il faut tout d'abord conceptualiser l'IA comme une expérience sociale menée sur la société, il est alors impératif que des garanties éthiques soient mises en place pour protéger les personnes et la planète.
2. Intégrer la durabilité de l’IA dans la politique, avoir des groupes de travail sur l'IA au sein des gouvernements, qui s'engagent activement dans la recherche d'avis d'experts sur l'impact environnemental de l'IA avec pour but de le réduire. les entreprises et les institutions publiques devraient être tenues de fournir des rapports sur les émissions de carbone pour toutes les formations et tous les réglages des systèmes d'IA qu’elles effectuent
3. les institutions publiques et privées, par exemple la Commission européenne, dans le cadre de leurs options réglementaires pour l'IA, devraient créer un "cadre de proportionnalité" pour évaluer si la formation ou le réglage d'un modèle d'IA pour une tâche particulière est proportionnel à l'empreinte carbone et à l'impact environnemental général de cette formation et/ou de ce réglage. Obligation aux entreprises d’utiliser des outils tels que le "Carbon Tracker" pour suivre l'empreinte carbone de la formation d'un certain modèle mais aussi pour prédire l'empreinte afin d'arrêter la formation du modèle si le coût environnemental prédit est dépassé

L'impact des résultats :

Critique personnelle : Je pense que cet article est très intéressant car il donne de bonnes recommandations mais il n’apporte rien de nouveau et propose seulement des solutions plus ou moins politiques.

Perspectives d’amélioration : donner des résultats concrets sur les outils proposés comme le Carbon Tracker pour montrer comment et à quel point ils aident

Cet article pourrait mener à un appel pour des projets européens pour changer la politique

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

La problématique: How big is too big? What are the possible risks associated with this technology and what paths are available for mitigating those risks?

La méthodologie : Fait un résumé de plusieurs études pour donner des recommandations

L’article commence par parler de l’évolution des LMs, puis évoquer leurs couts financier et environnemental avec d’autres articles comme celui de Strubell en faisant un parallèle sur les impacts que les LMs ont sur les personnes marginalisées.

* Première recommandation : Les chercheurs doivent donner la priorité à l'efficacité et au coût de l'énergie afin de réduire l'impact négatif sur l'environnement et l'accès inéquitable aux ressources - deux facteurs qui affectent de manière disproportionnée les personnes déjà marginalisées.

Par la suite, l’article discute de la façon dont les grands ensembles de données sur Internet nuisent davantage aux minorités car elles ne sont pas tier ou hors de leur contexte. De plus les chercheurs et d'autres personnes confondent les gains de performance induits par les LM avec une réelle compréhension du langage nature.

Les contributions, les nouveautés : aucun

Les résultats :

L'impact des résultats :

Critique personnelle : Cet article montre bien le danger de vouloir entrainer des algorithmes NLP trop vite et avec beaucoup de data. C’est un bon moyen de sensibilisé les gens avec des exemples concrets. **Ne sert à rien pour notre sujet**

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

La problématique :

Réduire la consommation d'énergie et les émissions de carbone des charges de travail ML

La méthodologie :

Les auteurs ont identifié les meilleures pratiques et choix qui peuvent réduire l'énergie jusqu'à 100 fois et les émissions de carbone jusqu'à 1000 fois par rapport aux choix orthodoxes (" 4M ") :

1. Model: Selecting efficient ML model architectures while advancing ML quality, such as sparse models versus dense modes, can reduce computation by factors of ~5–10.

2. Machine: Using processors optimized for ML training such as TPUs or recent GPUs (e.g., V100or A100), versus general-purpose processors, can improve performance/Watt by factors of 2–5.

3. Mechanization: Computing in the Cloud rather than on premise improves datacenter energy efficiency, reducing energy costs by a factor of 1.4–2.

4. Map: Moreover, Cloud computing lets ML practitioners pick the location with the cleanest energy, further reducing the gross carbon footprint by factors of 5–10^5

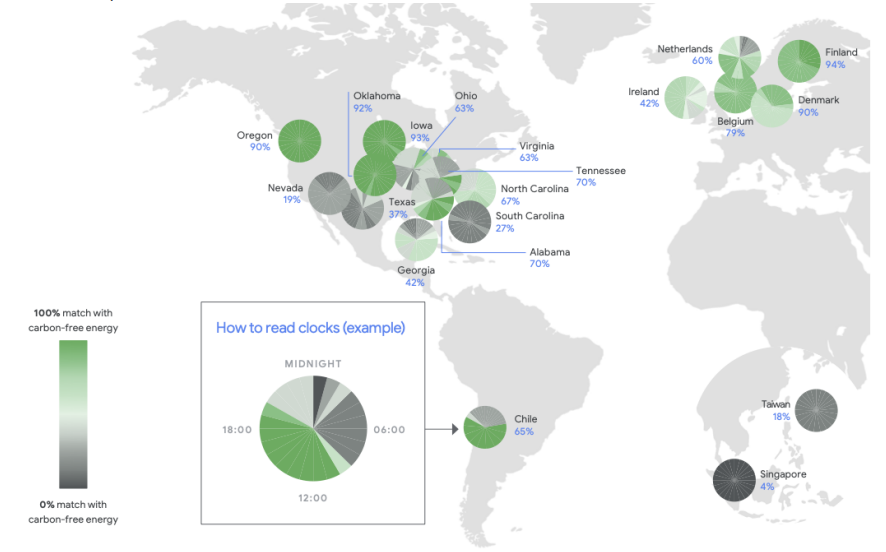
Vérifications :

Les auteurs ont mesuré la consommation d’énergie des TPU GPU et CPU utilisé pour le ML chez Google de manière rétroactive sur la base des données d'une semaine d'avril en 2019, 2020 et 2021. À chaque fois, la part de ML représentait 10 à 15 % de la consommation énergétique totale de Google pour cette semaine-là

Calculs:

𝑀𝑊ℎ = (𝐻𝑜𝑢𝑟𝑠 𝑡𝑜 𝑡𝑟𝑎𝑖𝑛 × 𝑁𝑢𝑚𝑏𝑒𝑟 𝑜𝑓 𝑃𝑟𝑜𝑐𝑒𝑠𝑠𝑜𝑟𝑠 × 𝐴𝑣𝑒𝑟𝑎𝑔𝑒 𝑃𝑜𝑤𝑒𝑟 𝑝𝑒𝑟 𝑃𝑟𝑜𝑐𝑒𝑠𝑠𝑜𝑟) × 𝑃𝑈𝐸

𝑡𝐶𝑂2𝑒 = 𝑀𝑊ℎ × 𝑡 𝐶𝑂2𝑒 𝑝𝑒𝑟 𝑀𝑊ℎ

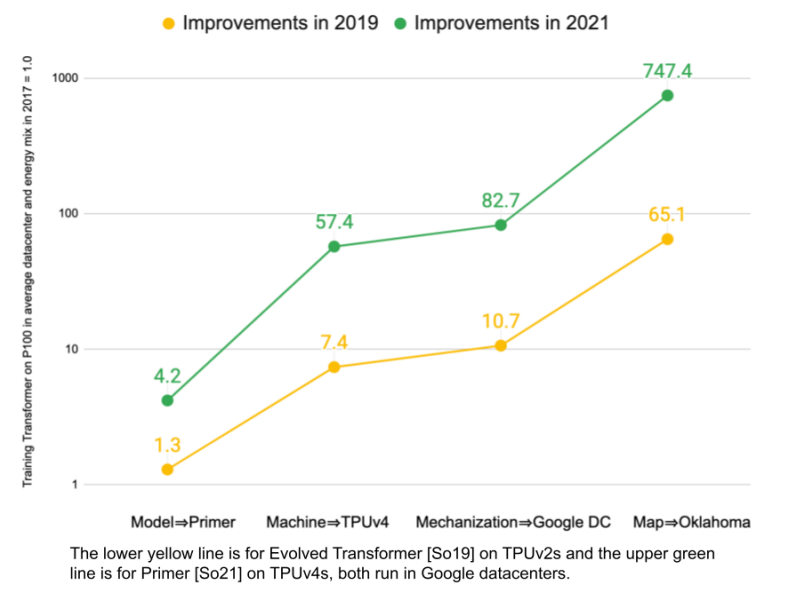


Les contributions, les nouveautés :

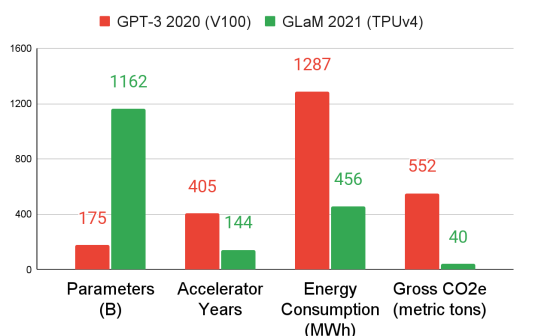
* La méthode de réduction l’application du modèle « 4Ms » sur différents algorithmes
* Une analyse qui montre que les études publiées ont surestimé le coût et l'empreinte carbone de la formation ML parce qu'elles n'avaient pas accès aux bonnes informations ou parce qu'elles extrapolaient des données ponctuelles sans tenir compte des améliorations algorithmiques ou matérielles.
* Les études de cas

Les résultats :

* Une réduction des émissions de 747x sans perte de précision du Transformer



* GLaM peut réduire l'empreinte carbone brute d'environ 14 fois avec une précision accrue par rapport au GPT-3,



* La stabilité de la consommation d’énergie pour ML de Google est une forte indication que malgré la popularité croissante de ML, si les meilleures pratiques sont suivies, sa consommation d'énergie ne monte pas en flèche, contrairement aux craintes communément exprimées.

Critique personnelle : donne de véritables résultats, présente une véritable méthode de réduction, apporte des critiques d’autres articles et études, analyse très intéressante de la consommation d’énergie pour le ML chez Google

Manque de détails sur les études de cas, reconnaitre chaque m pour les études de cas

A préciser -> deuxième études localisation

Perspectives d’amélioration : développer plus la partie sur l’étude de cas en énonçant clairement les situations et les paramètres de test en effet il manque quelques infos dessus

Implémentation possible : appliquer la règle des 4M sur la classification des images ou des méthodes similaires

KNAS: Green Neural Architecture Search

La problématique :

Peut-on évaluer les architectures sans entrainement ?

Il est admis que les gradients, induits par les réseaux neuronaux, sont le facteur le plus direct pour décider des résultats de convergence et de généralisation.

Hypothèse de recherche : les gradients peuvent être utilisés comme un proxy grossier de l'entraînement en aval pour évaluer les architectures initialisées de manière aléatoire

La méthodologie :

* Montre que les gradients dépendent de la conception des réseaux, notamment la profondeur du réseau, la fonction d'activation, l'initialisation, etc. Par conséquent, des architectures différentes donneront lieu à des gradients totalement différents.
* Démontrer théoriquement que la gram matrix de gradients GM est une mesure cruciale et complète pour évaluer la qualité des architectures.
* Montre que MGM a de bonnes corrélations avec la perte de formation et la précision de validation en échantillonnant 200 architectures provenant d'un benchmark NAS, NAS-Bench-201, et en montrant la relation entre les valeurs du kernel et les résultats d'optimisation sur CIFAR100
* Propose une solution NAS légère basée sur le kernel, appelée KNAS.
* Génération d’s architectures possibles en tant qu'espace de recherche S. Pas de politique de génération fixée afin de conserver la flexibilité sur diverses tâches et scénarios dans ce travail. Pour chaque architecture, il faut calculer le MGM en fonction de l'équation suivante :

Une image contenant texte

Description générée automatiquement

Les architectures ayant les scores les plus élevés sont sélectionnées comme candidates, et sont ensuite entraînées à partir de zéro pour obtenir leurs résultats sur un ensemble de validation. L'évaluation de MGM ne nécessite aucune étape d'entraînement.

Une image contenant texte

Description générée automatiquement

Contrairement aux autres approches NAS qui entraînent des centaines d'architectures, cette nouvelle approche permettrait de considérablement réduire les coûts de recherche

* Compare cette nouvelle approche avec différentes model déjà existants
* Vérification de la capacité de généralisation de KNAS avec plus d'expériences sur des tâches de classification de textes

Les contributions, les nouveautés :

Ont proposé et validé une hypothèse permettant d’obtenir une version verte du NAS pouvant trouver des architectures performantes avec des ordres de grandeur plus rapides que les solutions NAS "train-then-test"

Les résultats :

Une image contenant table

Description générée automatiquement

Nous pouvons voir que le KNAS une évolution dans la précision très importantes et propose des réulstats très compétiteurs à moindre couts

Une image contenant table

Description générée automatiquement

KNAS a de meilleures performances par rapport à RoBERTA-large avec des améliorations de précision de 1,24 et 0,24 sur les ensembles de données MRPC et RTE.

L'impact des résultats : montre clairement qu’il est possible de choisir le bon algo sans entrainement

Critique personnelle : revoir cet article encore après plus de lecture sur le sujet car pas assez de connaissances sur le sujet article assez difficile à comprendre

Travail à faire pour les 2 semaines suivantes :

* Continuer la bibliographie dans le but de faire un tableau comparatif avec 5 articles principaux
* Chercher tout ce qui impact de près ou de loin l’environnement durant l’intégralité du cycle de vie d’un maximum de source d’énergie
* Pour la première semaine faire le modèle de regression sur python donné lundi
* Commencer à rédiger la section d'introduction en anglais