

Impacts environnementaux de l'IA

Cours Intelligence Artificielle et Environnement

Master MVA

janvier 2024

Objectif du TD/TP : décrire qualitativement ou quantitativement les impacts directs et indirects d'une application numérique utilisant une solution d'IA.

On considérera une application de co-voiturage pour la mobilité quotidienne (courte distance), utilisant de l'IA pour améliorer l'appariement des voyageurs et la prédiction du trafic automobile.

1 Impacts de 2e et 3e ordre

Dans cette première partie, nous allons nous intéresser aux effets de 2e et 3e ordre de l'utilisation de l'application de co-voiturage.

1. Indépendamment de l'utilisation de cette application, quels sont les évolutions les plus probables du secteur de la mobilité quotidienne dans les 10 prochaines années ?
2. À quel(s) scénarios(s) de référence le scénario d'utilisation de cette application devrait-il être comparé ?
3. Quels sont les effets positifs de 2e ordre attendus c'est-à-dire les bénéfices environnementaux dus directement à l'utilisation de cette application ? Expliciter les conditions dans lesquelles ces bénéfices auront lieu.
4. Quel serait un ordre de grandeur de réduction intéressant par rapport aux stratégies nationales de réduction des émissions ?
5. Quels sont les effets (toujours de 2e ou 3e ordre, donc en plus des impacts liés au cycle de vie des équipements nécessaires à l'application) imaginables dans le secteur du numérique ?
6. Plus généralement, quels effets environnementaux positifs et négatifs pouvez-vous envisager du fait de l'utilisation de cette application ? En particulier, quels changements de comportement l'utilisation de cette application est-elle susceptible de provoquer chez ses utilisateurs ?
7. Quelles contre-mesures pouvez-vous imaginer afin de limiter les effets négatifs ?
8. L'utilisation de cette application risque-t-elle d'entrer en conflit avec d'autres leviers de décarbonation de la mobilité quotidienne et si oui lesquels ?
9. Quels problèmes de résilience face aux changements environnementaux l'utilisation de cette application risque-t-elle de poser, par exemple en cas de coupure de réseau ou de rupture d'approvisionnement matériel ou énergétique ?

2 Impacts du 1er ordre

L'outil [Green Algorithms](#) permet de calculer la consommation d'électricité et l'empreinte carbone de l'exécution d'un programme d'apprentissage automatique.

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.

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

En utilisant l'outil [Green Algorithms](#) :

1. Calculer la consommation d'énergie et l'empreinte carbone associée au scénario de référence avec l'outil Green Algorithms.
2. Le PUE utilisé est un PUE par défaut. Que devient votre empreinte carbone si vous utilisez un datacenter avec un PUE d'1,2?
3. Pour rappel le PUE se calcule de la façon suivante :

$$PUE = \frac{\text{consommation totale datacenter}}{\text{consommation informatique}}$$

Une partie importante de la consommation non informatique d'un datacenter vient de la climatisation. Qu'en déduisez-vous sur la pertinence d'utiliser un PUE moyen? (remettez le PUE à 1,67 après cette question)

4. Si l'on suppose que la consommation d'électricité du datacenter suit la répartition de la figure 1, quelle proportion de la consommation d'électricité a-t-on étudiée avec [Green Algorithms](#)? (attention cependant car dans cette figure, il n'y a que des CPU donc ce n'est pas réellement transposable à notre cas d'étude)
5. Quels avantages et inconvénients voyez-vous à utiliser un tel outil plutôt qu'un package python comme [CodeCarbon](#)?

2.1 Prise en compte du cycle de vie

1. Si l'on suppose que la répartition de l'empreinte carbone des serveurs du datacenter suit la répartition du tableau 2, quelle proportion de l'empreinte carbone a-t-on étudiée avec Green Algorithms?
2. Comment pensez-vous que l'empreinte carbone évolue si les serveurs sont sur un cloud plutôt qu'en local? Testez votre hypothèse avec un cloud Azure en "France Central" pour avoir une localisation similaire.
3. Est-ce que la comparaison cloud/serveurs locaux avec [Green Algorithms](#) de permet de conclure sur l'intérêt environnemental du cloud par rapport à des serveurs locaux?
4. Comparez l'empreinte obtenue avec un modèle plus ancien (NVIDIA P100 par exemple, de 2016) et un modèle plus récent et plus spécialisé (TPU v3 par exemple, de 2018). Qu'en concluez-vous?
5. De quelles informations auriez-vous besoin pour estimer l'empreinte carbone en incluant la fabrication des équipements?

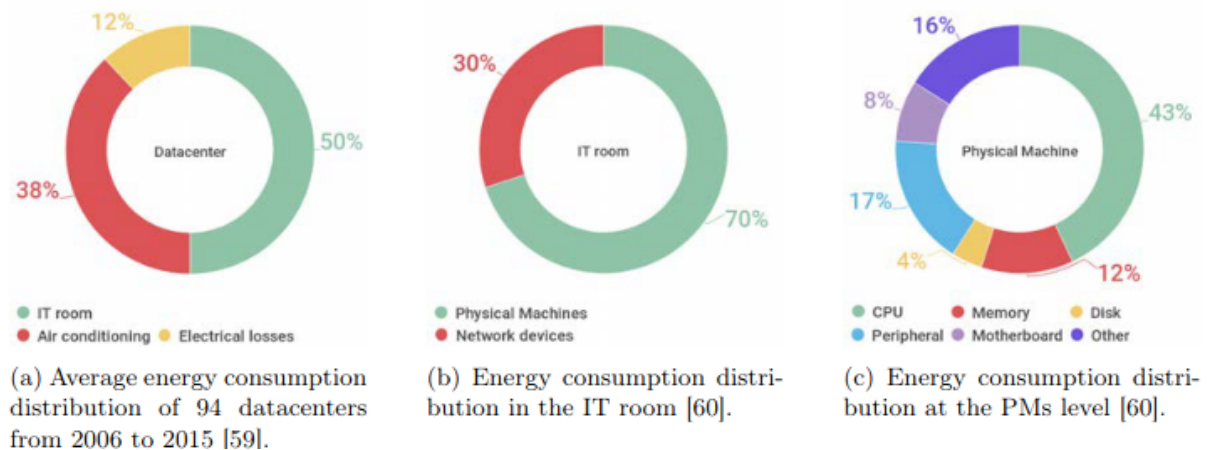


FIGURE 1 – Répartition de la consommation d’électricité dans des datacenters. Physical machines = serveurs. Source : David Guyon, 2018

Matériel	gCO2e	%
Serveurs de calcul (fabrication)	1,16	29
Serveurs environnement (fabrication)	0,38	10
Serveurs de calcul (usage : électricité)	1,79	45
Serveurs environnement (usage : électricité)	0,64	16

TABLE 2 – Empreinte carbone des serveurs du centre de calcul GRICAD pour une heure.coeur de calcul. Source : Berthoud et al, 2020

2.2 Compromis empreinte carbone / résultat

1. Si l’on avait arrêté l’entraînement au bout de 3 jours, quelle quantité d’émissions aurait-on économisé ?
2. Considérons la figure 2 qui présente les taux d’erreur pendant l’entraînement de systèmes de reconnaissance de la parole pour deux corpus (jeux de données) différents. Les auteurs de l’article correspondant ont calculé les taux d’erreur des systèmes lorsque 50% et 100% de l’empreinte carbone finale ont été émis. Quels sont les gains de précision pour chaque courbe ? Quand vous semble-t-il raisonnable d’arrêter l’entraînement d’un tel système ?

2.3 Entraînement vs inférence

Nous considérerons ici le [modèle de langue BLOOM créé récemment](#). Les informations sur l’entraînement et l’inférence sont disponibles dans le tableau 3. Nous supposons que le PSF de l’entraînement est de 2, et que la consommation des CPU est négligeable.

1. Calculez l’empreinte carbone de l’entraînement de ce modèle.
2. Au moment de l’écriture de l’article, le modèle avait été utilisé pendant environ 18 jours, la consommation de l’utilisation du modèle émettant environ 19kgCO2e par jour. Quelle est l’empreinte carbone associée ?

2.4 Électricité verte ?

1. Comment pensez-vous que l’empreinte carbone évolue si les serveurs sont localisés en Pologne ? Et en Suède ? Testez vos hypothèses.
2. D’après un article de recherche de Google, les achats d’énergie renouvelable de Google réduisent l’impact à zéro¹. Qu’en pensez-vous ?

1. «Google’s renewable energy purchases further reduce the impact to zero.», <https://arxiv.org/pdf/2204.05149.pdf>

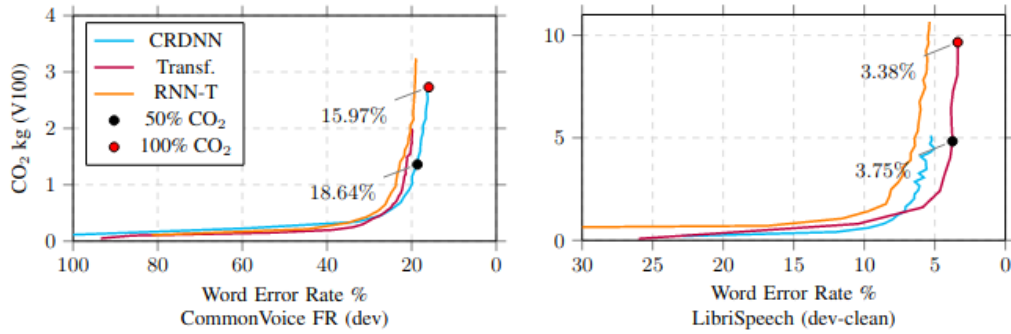


Figure 2: CO₂ emitted in kg (in France) by different E2E ASR models with respect to the word error rate (WER) on the dev sets of LibriSpeech and CommonVoice. The curves exhibit an exponential trend as most of the training time is devoted to slightly reduce the WER. The black and red dots indicates the WER obtained with 50% and 100% of the emitted CO₂. On LibriSpeech, 50% of the carbon emissions have been dedicated to reach SOTA results with an improvement of 0.37%.

FIGURE 2 – Taux d’erreur vs empreinte carbone. Source : T. Parcollet et al, 2021

Temps d’exécution	118 jours (2 832 heures), 5 heures, 41 min
Nombre de GPU	416
Modèle de GPU	Nvidia A100
Mémoire disponible par GPU	80GO (33 280 en tout)
Localisation du serveur	Orsay, France
PUE	1,2

TABLE 3 – Informations sur l’apprentissage du modèle

2.5 Comparaison CodeCarbon et Green Algorithms

1. Testez la librairie python [CodeCarbon](#) sur un de vos programmes d’apprentissage (rapide!) ou un exemple test.
2. Comparez le résultat avec celui fourni par [Green Algorithms](#). S’il y a des différences, à quoi les attribuez-vous ?