

L'intelligence artificielle est-elle la prochaine grande menace pour le changement climatique ?

De sérieux doutes ont été émis à propos de la consommation énergétique de l'intelligence artificielle mais les nouvelles technologies de puces et l'IA elle-même pourraient représenter des solutions d'auto-régulation viables.

Temps de lecture : minute

31 juillet 2019

[Retrouvez le billet initial en anglais sur la page Medium du MIT](#)

Lors d'une récente conférence à San Francisco, Gary Dickerson a formulé une prédiction audacieuse. Le directeur général d'Applied Materials, un des plus importants fournisseurs de l'industrie des semi-conducteurs (matériau qui a les caractéristiques d'un isolant, très utilisé dans les *data centers*), a prévenu qu'en l'absence d'innovation significative dans la conception et la fabrication des *data centers*, ces derniers pourraient représenter un dixième de l'utilisation de l'électricité mondiale d'ici 2025.

Aujourd'hui, les millions de *data centers* du monde entier absorbent un peu moins de 2% de la consommation d'électricité mondiale — et cette statistique englobe tous les types d'activités opérées via l'ensemble de ces serveurs. Applied Materials estime que les serveurs qui utilisent l'intelligence artificielle ne représentent actuellement que 0,1% de cette consommation.

Mais d'autres dirigeants de la technologie tirent également la sonnette d'alarme. Anders Andrae, de chez Huawei, pense également que les *data*

centers pourraient finir par consommer un dixième de l'électricité mondiale d'ici 2025, même si son estimation couvre toutes leurs utilisations, et pas seulement celles touchant à l'intelligence artificielle. Jonathan Koomey, conseiller spécial du scientifique senior du Rocky Mountain Institute, est, pour sa part, plus optimiste. Il s'attend à ce que la consommation énergétique des *data centers* reste relativement stable au cours des prochaines années et ce malgré le pic d'activité lié à l'IA.

Ces prévisions très divergentes mettent en évidence les incertitudes concernant l'impact de l'IA sur l'avenir de l'informatique à grande échelle mais aussi sur ses implications en termes d'empreinte énergétique.

Revoir les process industriels

Néanmoins, Applied Materials a aussi reconnu avoir imaginé ce scénario catastrophe pour souligner ce qui pourrait se passer en l'absence de transformation réelle des industries *hardware* et *software*. Sundeep Bajikar, directeur de la stratégie de l'entreprise, estime qu'il y aura des changements dans les combinaisons d'informations utilisées pour former les modèles d'intelligence artificielle, les vidéos et autres images constituant par exemple un pourcentage croissant du total par rapport aux informations textuelles et audio. De plus, ces données visuelles sont plus puissantes d'un point de vue informatique et donc plus énergivores.

Les modèles auront aussi plus d'informations à se mettre sous la dent, grâce aux données fournies par des innovations tels que les véhicules autonomes et les capteurs intégrés dans d'autres appareils intelligents. Et la généralisation d'une connectivité 5G très rapide facilitera d'autant plus la circulation des données vers les *data centers*.

Sundeep Bajikar explique que ces tendances soulignent un besoin urgent pour ce que l'entreprise appelle une " nouvelle grille de lecture " des matériaux et process de fabrication à l'ère de l'intelligence artificielle.

Certains chercheurs pensent que la quête de puissance de l'IA pourrait même devenir un casse-tête environnemental majeur : une équipe de l'Université du Massachusetts, Amherst, a récemment publié une étude montrant que la formation de plusieurs grands modèles populaires d'IA produit pollue près de cinq fois plus qu'une voiture américaine moyenne...

Parier sur les fondamentaux

Mais ces prévisions pessimistes ignorent plusieurs avancées technologiques importantes qui pourraient limiter la consommation énergétique de l'intelligence artificielle. L'une d'entre elles est notamment la montée en puissance des *data centers* dits " hyperscale ", mis au point par des entreprises comme Facebook et Amazon. Ces centres utilisent de vastes ensembles de serveurs basiques configurés pour des tâches très spécifiques. Ces machines sont plus économes en énergie que les serveurs de centres traditionnels qui doivent jongler avec un plus large éventail d'activités et de tâches. Un passage à l'hyperscale ainsi que de nombreux progrès dans les technologies de refroidissement permettraient de compenser totalement la consommation énergétique de ces nouveaux centres.

De nouveaux types de micro-puces représenteront également une aide précieuse pour réduire l'empreinte énergétique des *data centers*. Les prévisions d'Applied Materials émettent l'hypothèse que les activités de l'IA seront réalisées sur des matériels existants dont l'efficacité s'améliorera progressivement au cours des prochaines années. Mais de nombreuses startups, ainsi que de grandes entreprises comme Intel et AMD, développent des semi-conducteurs qui utilisent des technologies comme la photonique afin d'alimenter les réseaux et autres outils liés à l'IA en utilisant beaucoup moins d'énergie.

Selon Koomey, les projections alarmistes écartent également le fait que pour certains types de tâches liées à l'intelligence artificielle, comme la

reconnaissance de modèles existants, des résultats approximatifs de reconnaissance sont amplement suffisants. Cela signifie que l'énergie n'est pas dépensée pour calculer les résultats précis à des centaines de décimales.

L'ironie du sort réside donc dans le fait que la meilleure surveillance de la consommation énergétique de l'IA pourrait être l'IA elle-même. Google utilise déjà la technologie développée par DeepMind, une société qu'elle a acquise en 2014, pour refroidir ses *data centers* plus efficacement. L'IA avait déjà aidé l'entreprise à réduire sa facture de refroidissement de 40% en faisant des recommandations aux opérateurs humains ; aujourd'hui, elle exploite directement et efficacement les systèmes de refroidissement dans les centres.

De plus, on peut imaginer que l'intelligence artificielle sera également utilisée pour optimiser d'autres aspects des opérations des centres de données. Cela ne veut pas dire que les centres de données ne finiront pas par consommer beaucoup d'énergie, en raison d'une utilisation croissante de l'IA. Mais cela rend les prédictions tout de même beaucoup plus difficiles à faire.