



CHRONIQUE DE IRÈNE BALMÈS
KYNAPSE AN OPEN COMPANY

POUR UN NUMÉRIQUE PLUS VERT : QUID DE L'INTELLIGENCE ARTIFICIELLE ?

Alors que le gouvernement ouvre la voie à des réflexions sur l'impact environnemental du numérique, quelle réflexion mener sur l'intelligence artificielle et son coût énergétique ?

Le gouvernement a présenté mardi 23 février sa feuille de route pour un numérique vert, visant à maîtriser les impacts du numérique, notamment autour de l'allongement de la durabilité et de l'augmentation du taux de recyclage des matériels numériques. Dans le même temps, Google vient de licencier Timnit Gebru et Margaret Mitchell, en apparence pour n'avoir pas voulu rétracter un article scientifique ("on the dangers of Stochastic Parrots : can Language Models be too big?"), critiquant notamment l'impact écologique des "large language models", ces modèles d'intelligence artificielle permettant de faire de la complétion automatique de texte, de la traduction et bien d'autres choses. Cette convergence amène à se poser la question : quel est le rôle de l'intelligence artificielle dans le changement climatique et dans la transition écologique ? Peut-elle être un outil permettant de réduire l'empreinte carbone d'autres secteurs, ou son empreinte même va-t-elle être un obstacle majeur ?

Des travaux académiques récents montrent l'évolution de la consommation énergétique de l'entraînement de certains algorithmes d'IA, parmi les plus gourmands : d'un modeste 80 kg d'équivalent CO2 pour des modèles publiés en 2017, on passe à 700 kg-eCO2 en 2018 (un vol Paris-New York pour un passager) puis quasiment 300 000 kg-eCO2, lorsqu'on prend en compte l'optimisation de l'architecture, pour un modèle publié en 2019 (soit la consommation totale de 60 personnes au cours de leur vie !). Ces mesures ne concernent que l'entraînement des modèles ; leur utilisation consomme également de l'énergie, quoique plus modestement. Cependant cette dépense augmente également avec la taille des modèles, et avec la multiplication de leurs utilisations.

Pour améliorer cette empreinte carbone, le plus simple est tout d'abord d'améliorer le mix énergétique du pays hébergeant les centres de calcul : entraîner un modèle dans un pays utilisant peu le nucléaire et les énergies renouvelables aura un coût en carbone beaucoup plus important. Sur ce point, la France est déjà un bon élève. Mais la question reste posée : le bénéfice de ces modèles justifie-t-il leurs coûts ?

La convention citoyenne pour le climat, qui a inspiré en partie la feuille de route gouvernementale, note que le numérique doit être "un moyen pour participer à la transition et pas un outil qui contribue toujours davantage à la hausse des émissions." En effet, l'intelligence artificielle en particulier peut permettre d'optimiser des tâches énergivores, telles que le refroidissement des centres de calcul, d'automatiser des processus, et dans l'ensemble, de réduire notre consommation énergétique. Google a ainsi pu réduire la consommation de ces data center de 40%.

Mais tous les modèles ne sont pas égaux : certains sont facilement entraînés en quelques minutes sur un ordinateur de bureau, quand d'autres vont réclamer des centaines ou des milliers d'heure de calcul. Le coût financier de la location d'heures de calcul sur AWS ou similaire est déjà pris en compte par les entreprises qui se lancent dans ce genre de chantier : le coût énergétique pourrait-il également être envisagé ? Cela pourrait impacter les types de processeurs et l'implantation du centre de calcul où sera effectué l'optimisation de l'algorithme.

Les informaticiens ont l'habitude de juger un algorithme sur sa complexité, c'est à dire le temps d'exécution qu'il va nécessiter en fonction de la taille de l'entrée : les data scientists auraient intérêt à juger un algorithme non plus seulement sur sa performance, mais également sur son coût d'entraînement.

À l'heure actuelle, de nombreux acteurs sont en compétition pour produire un langage model qui soit de plus en plus efficace. Cette concurrence a des effets vertueux sur la performance des algorithmes, mais est loin d'être optimale car chacun va rejouer les entraînements de modèle de ses concurrents. Alors qu'il y a quelques années les modèles une fois entraînés étaient généralement mis en open source dans la communauté, certains modèles récents, notamment GPT-3, sont restés privés. Outre la difficulté d'auditer un modèle privé pour en comprendre le fonctionnement et les biais, cela présente l'inconvénient que d'autres entreprises sont en train d'essayer de reproduire ce modèle, multipliant ainsi les coûts d'entraînement.

La facilité avec laquelle nous pouvons maintenant entraîner un modèle de grande taille ne doit pas nous empêcher de réfléchir aux conséquences moins visibles : plus les modèles sont complexes, plus ils sont grands, et plus ils sont énergivores. Ces grands modèles sont également plus difficiles à auditer et à comprendre. Bien que semblant indispensables dans certains domaines, comme le langage et la vision, il est bon de mettre dans la balance, face à des modèles plus simples, au-delà de leur performance, leur coût énergétique et les bénéfices qu'ils peuvent apporter. Dans sa feuille de route, le gouvernement ne parle que peu d'intelligence artificielle, mais la discussion doit être engagée par les différents acteurs du secteur, en France comme à l'international.