Comment faire face à la crise de l'énergie de l'IA ?

Thomas Le Goff

THE CONVERSATION
L'expertise universitaire, l'exigence journalistique

La centrale de Three Mile Island aux États-Unis, siège d'un accident nucléaire en 1979, va reprendre du service pour alimenter en électricité les centres de données de Microsoft. Ici en 2019. Constellation Energy, Wikipedia, CC BY-SA

Comment faire face à la crise de l'énergie de l'IA ?

Publié: 28 novembre 2024, 17:21 CET

Thomas Le Goff
Maître de conférences en droit et régulation du numérique, Télécom Paris – Institut Mines-Télécom

L'intelligence artificielle se développe plus rapidement que les énergies renouvelables. Alors, tandis que l'on demande aux particuliers de baisser le chauffage, les GAFAM relancent le nucléaire. L'Agence internationale de l'énergie convoque un congrès mondial sur l'intelligence artificielle et l'énergie les 4 et 5 décembre 2024.

La centrale nucléaire américaine « Three Mile Island » est tristement célèbre pour avoir été le théâtre de l'un des plus terribles accidents nucléaires de l'histoire en 1979, et elle va bientôt reprendre du service pour alimenter les systèmes d'intelligence artificielle (IA) de Microsoft. Cette annonce, faite en septembre 2024 et qui concerne un réacteur indépendant de celui ayant causé l'accident de 1979, s'inscrit dans une tendance plus globale d'investissements massifs des géants du numérique dans l'énergie nucléaire.

Ainsi, Google a également annoncé la signature d'un accord avec la start-up Kairos Energy, spécialisée dans la construction de petits réacteurs nucléaires modulaires (dits « SMR »), pour financer son développement et réserver une partie de sa capacité de production à l'horizon 2030. Amazon, suivant le mouvement, a conclu un partenariat similaire avec la start-up X-energy.

La raison de ces investissements est simple : le développement exponentiel de l'IA générative demande d'importantes capacités de calcul, localisées dans des centres de données particulièrement énergivores.

Les études les plus récentes montrent que l'IA représente entre 10 et 20 % de l'électricité consommée par l'ensemble des centres de données dans le monde, laquelle augmente de 20 à 40 % chaque année d'après l'Agence Internationale de l'Énergie (AIE). Dans certains pays, comme l'Irlande, la consommation liée aux centres de données a même dépassé la quantité d'électricité consommée par les ménages.

La démesure de ces chiffres interroge, qui plus est dans un contexte où l'urgence climatique est dans tous les esprits et alors qu'on demande aux citoyens de limiter leur chauffage à 19 °C, cette course à la capacité de calcul est-elle vraiment soutenable et souhaitable ? Doit-on réellement chercher par tout moyen à construire de nouvelles capacités de production d'électricité pour suivre le rythme de développement des centres de données ?

Les solutions à cette crise ne sont pas évidentes tant il y a d'intérêts divergents et de facteurs à prendre en considération. Pourtant, des pistes pour limiter la consommation énergétique de l'IA et l'explosion du nombre de centres de données, telles que la fiscalité ou la régulation, commencent à émerger dans les discussions internationales.

Pourquoi l'IA a-t-elle besoin de tant d'énergie ?

Chaque fois que nous posons une question à notre système d'IA générative préféré, la demande est envoyée par Internet pour être traitée dans un centre de données qui peut être situé dans différentes régions du monde. Ce dernier consomme de l'électricité pour alimenter les composants informatiques qu'il héberge et son système de refroidissement, sans compter l'énergie requise pour construire le centre et les composants électroniques eux-mêmes.

Ces dernières années, les principaux modèles d'IA ont gagné en complexité et requièrent des capacités de calcul toujours plus importantes pour fonctionner, de 4 à 5 fois plus chaque année depuis 2010 d'après les études les plus récentes. En parallèle, le nombre d'utilisateurs ne cesse d'augmenter, avec plus de 200 millions d'utilisateurs chaque semaine rien que sur ChatGPT.

Les modèles d'intelligence sont toujours plus gourmands en capacité de calcul — ici le nombre d'opérations totales nécessaires pour entraîner chaque modèle d'IA, en fonction du temps. Jaime Sevilla et Edu Roldán, Epoch AI

Ces tendances expliquent pourquoi les fournisseurs d'IA ont besoin de plus en plus d'énergie, investissent massivement dans les énergies renouvelables pour alimenter leurs systèmes et projettent la construction de nouvelles infrastructures partout dans le monde.

Pourquoi la multiplication des centres de données est un problème pour la planète ?

L'accélération de la demande en capacités de calcul liée à la mode de l'IA générative s'accompagne d'importants effets négatifs sur l'environnement.

D'abord, la production de l'électricité consommée par les centres de données génère des émissions de gaz à effet de serre suivant la source utilisée. Ces émissions représentent déjà 1 à 3 % des émissions globales d'après l'AIE et risquent d'augmenter si le nombre de centres augmente.

Ensuite, les centres de données étant particulièrement énergivores, ils peuvent affecter la stabilité du réseau à l'échelle locale. Dans un réseau électrique, la quantité d'électricité produite doit toujours être égale à la quantité d'électricité consommée sinon c'est le *black-out* (la panne). Ajouter des infrastructures consommant beaucoup d'électricité dans des zones géographiques où l'équilibre production-consommation est déjà fragile aggrave le risque de *black-out*, notamment lorsque le mix énergétique repose en grande partie sur des énergies renouvelables, par nature intermittentes.

Enfin, le rythme de développement de l'IA dépasse complètement celui des capacités de production d'électricité à partir d'énergies renouvelables comme les panneaux photovoltaïques ou les éoliennes. Pour répondre à leurs besoins, les géants du numérique vont vraisemblablement avoir recours à des sources d'énergie carbonées comme le charbon ou le gaz, disponibles plus rapidement. Cela les conduit à s'éloigner de façon catastrophique de leurs objectifs de neutralité carbone, Microsoft ayant affiché une augmentation de 29 % de ses émissions par rapport à 2020 et Google de 48 % par rapport à 2019. En parallèle, ils communiquent intensément sur leurs investissements dans les énergies renouvelables afin de faire oublier leur mauvaise performance environnementale.

Quelles solutions pour faire face à la crise de l'énergie de l'IA ?

La solution n'est pas forcément d'interdire la construction de nouveaux centres de données, et ce pour trois raisons.

En effet, les nouveaux centres de données construits par les géants du numérique sont globalement plus efficients que les anciennes infrastructures. La construction de nouveaux centres répond également à d'autres enjeux puisqu'ils contribuent au développement économique des territoires (en créant des emplois et de l'activité à l'échelle locale) mais aussi à l'établissement d'une puissance de calcul souveraine (par exemple en Europe), moins sujette aux potentiels effets de différends géopolitiques à l'échelle internationale.

De plus, à moins d'un moratoire global sur la construction de nouvelles infrastructures, interdire localement des projets d'implantation ne conduira qu'à leur délocalisation, potentiellement dans des pays où le mix énergétique est encore plus carboné, ce qui n'est pas souhaitable d'un point de vue écologique…

Comment mieux réguler l'impact environnemental du numérique ? Source : Telecom Paris.

L'urgence d'une réflexion internationale sur la régulation des centres de données

À l'image de la directive européenne sur l'efficacité énergétique et du code de conduite européen applicables aux centres de données, il est essentiel de s'assurer que chaque nouveau projet utilise les meilleures technologies disponibles en termes d'efficacité énergétique, mais aussi pour éviter que la consommation ne croisse à cause de l'effet rebond, et soit alimentée par une électricité bas carbone. Plus les standards seront harmonisés au niveau mondial, moins le risque de délocalisation vers des pays aux normes plus souples, mais potentiellement moins vertueuses d'un point de vue environnemental, sera important.

À lire aussi : L'effet rebond : quand la surconsommation annule les efforts de sobriété

Une régulation du nombre de centres de données à l'échelle mondiale pourrait aussi être envisagée, via une organisation mondiale, sur le modèle de l'Union Internationale des Télécommunications, qui gère l'attribution des fréquences radioélectriques.

Une réflexion sur la fiscalité des opérateurs de centres de données s'avère également nécessaire afin de déterminer si elle peut être utilisée pour favoriser l'approvisionnement en énergie verte et l'adoption de pratiques plus durables, via des abattements fiscaux ou l'établissement d'une taxe spécifique pour les opérateurs les moins vertueux. Par exemple, la piste a été évoquée dans la mission d'information du Sénat sur l'empreinte environnementale du numérique en 2020, qui a conduit au conditionnement d'une taxe réduite pour les datacenters respectant des critères de performance énergétique, uniquement en France.

Enfin, il est aussi possible d'agir du côté des usages de l'IA. La sensibilisation du public aux enjeux environnementaux de l'IA permettrait d'orienter les usages vers une utilisation plus vertueuse de la technologie en limitant les usages récréatifs par exemple.

Bien souvent, dans les débats sur l'empreinte environnementale de l'IA, il est évoqué la nécessité de mettre en balance les externalités négatives liées à son développement, telles que celles mentionnées dans cet article, avec les potentiels effets positifs que l'IA peut apporter dans différents secteurs, notamment économiques (création de richesse) ou environnementaux (réduction des émissions via une optimisation de l'efficacité énergétique d'autres activités).

Si l'argument est séduisant et semble rationnel, d'hypothétiques effets positifs sur le long terme ne peuvent justifier un développement déraisonné de l'IA à court terme, causant des dommages irréversibles à l'environnement et risquant de compromettre notre capacité à léguer aux générations futures un environnement sain.