

Consommations gargantuesques d'électricité et d'eau pour créer des systèmes d'IA « pas au niveau d'un élève de maternelle », manque de clarté des modèles économiques et pillage industriel des données... trois articles consacrés à l'IA récemment publiés dans Le Monde¹ (15/06/24), Les Echos² (24/06/24) et le WSJ³ (01/07/24) remettent les pendules à l'heure. Mise à jour de la note LM Invest France de septembre 2023⁴.

L'IA, l'industrie qui carbure au charbon

Pour que le *software* IA marche, il faut du *hardware* puissant, et donc des puces et des data centers. Et il va en falloir de plus en plus. Sundar Pichai, patron de Google, cité dans l'article du Monde : « le besoin de calcul informatique pour l'IA a été multiplié par un million en six ans, il décuple chaque année. »

Pour les puces, pas de problème, la même conclusion qu'il y a six mois s'impose : « la terre tourne autour du soleil, les Américains vendent des produits d'IA, les Européens les achètent, et Nvidia dit merci. » Mise à jour ci-dessous du graphique du cours de bourse de Nvidia.

Market Summary > Nvidia

123.63 USD

+83.00 (204.27%) ↑ past year

Jun 26, 11:27 EDT • Disclaimer

1D 5D 1M 6M YTD 1Y 5Y Max



Source Google, 26/06/24

Pour les data centers par contre, il y a un gros problème qui est celui des besoins en électricité et en eau. L'IA et les cryptomonnaies, qui représenteraient déjà 2% de la consommation mondiale d'électricité, pourraient voir ce taux doubler d'ici à 2026.

Aux Etats-Unis, selon l'Electric Power Research Institute⁵, dans un scénario de fort développement, les data centers pourraient requérir jusqu'à 9% des capacités électriques du pays d'ici à 2030, contre 4% aujourd'hui. Ce taux d'utilisation du réseau pourrait atteindre 25% dans le Nebraska et l'Iowa, 31% dans le Dakota du Nord, 46% en Virginie.

Cette électricité, comment est-elle produite ? Avec du charbon, à hauteur de 30%. Oui, l'industrie de la tech carbure au charbon.

¹ https://www.lemonde.fr/economie/article/2024/06/14/derriere-l-ia-la-deferlante-des-data-centers_6239694_3234.html

² <https://www.lesechos.fr/tech-medias/intelligence-artificielle/pour-yann-lecun-les-investissements-dans-l-ia-restent-un-pari-2103274>

³ https://www.wsj.com/business/energy-oil/tech-industry-wants-to-lock-up-nuclear-power-for-ai-6cb75316?mod=hp_lead_pos1

Aux Etats-Unis, l'insatiable appétit d'électricité de la tech

La montée de l'IA a amplifié les besoins. La situation aggrave le bilan carbone du secteur et risque de saturer les réseaux électriques

Source Le Monde, 15/06/24

Ce mix énergétique déplorable n'est pas tenable. Non pas pour des raisons de bilan carbone dont l'industrie de la tech n'a cure. De toute façon, les énergies renouvelables, intermittentes, ne conviennent pas : il faut de la qualité parfaite « 24/7 ». Non, cela n'est pas tenable tout simplement parce que les capacités de production électrique des Etats-Unis s'épuisent.

Dans ces conditions comment faire pour disposer d'électricité supplémentaire ? Se tourner vers les producteurs d'énergie nucléaire.

Tech Industry Wants to Lock Up Nuclear Power for AI

Largest tech companies are looking to buy nuclear power directly from plants, which could sap the grid of critical resources

By Jennifer Hiller and Sebastian Herrera

Updated July 1, 2024 12:06 am ET

Share Resize

Listen (9 min)



The Beaver Valley nuclear facility in Pennsylvania. PHOTO: JUSTIN MERRIMAN/BLOOMBERG NEWS
Source WSJ, 01/07/24

Cette dernière idée de l'industrie technologique commence à poser de sérieux problèmes, la demande d'électricité étant telle qu'elle menace la qualité et la fiabilité des réseaux.

⁴ Cette note est une mise à jour de la note du 19/09/23 :

<https://lminvestfrance.com/2023/09/19/market-flash-naivete-reelle-intelligence-artificielle-19-09-23/>

⁵ Voir le rapport très complet de l'Electric Power Research Institute : file:///C:/Users/Huques/Downloads/3002028905_Powering%20Intelligence%20Analyzing%20Artificial%20Intelligence%20and%20Data%20Center%20Energy%20Consumption.pdf

Les autorités s'inquiètent de voir AWS signer un deal d'approvisionnement en énergie nucléaire avec Constellation Energy, ou sa maison-mère Amazon acquérir pour 650M\$ en Pennsylvanie un data center fonctionnant à l'énergie nucléaire. Se pose en effet la question de l'usage du réseau par un unique acteur privé aux dépens des autres usagers, par ailleurs citoyens assujettis à l'impôt ayant financé une grande partie des infrastructures. La commission fédérale de régulation de l'énergie⁶ est saisie sur l'opération Amazon en Pennsylvanie. Dans le New Jersey, ce sont près de 2Mds\$ qui ont été investis au cours des six dernières années dans les infrastructures : pourquoi une poignée d'acteurs privés feraient main basse dessus ?

Mix énergétique entrainement des modèles big data, étude « Energy and Policy Considerations for Deep Learning in NLP »

Consumer	Renew.	Gas	Coal	Nuc.
China	22%	3%	65%	4%
Germany	40%	7%	38%	13%
United States	17%	35%	27%	19%
Amazon-AWS	17%	24%	30%	26%
Google	56%	14%	15%	10%
Microsoft	32%	23%	31%	10%

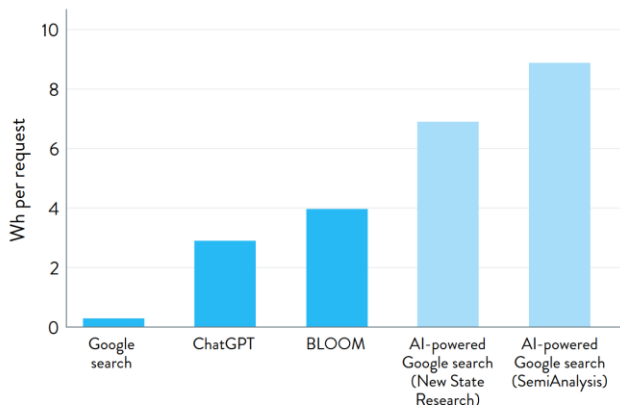
Table 2: Percent energy sourced from: Renewable (e.g. hydro, solar, wind), natural gas, coal and nuclear for the top 3 cloud compute providers (Cook et al., 2017), compared to the United States,⁴ China⁵ and Germany (Burger, 2019).

Source <https://aclanthology.org/P19-1355.pdf>

Tout ça pour quoi ?

Une requête ChatGPT consommerait 3,0Wh (watt-hours) d'électricité contre 0,3Wh pour une requête Google. C'est dix fois plus. L'humanité est-elle dix fois plus intelligente depuis que ChatGPT existe ? 1 litre d'eau serait nécessaire pour 100 requêtes ChatGPT, une folie. Une réflexion s'impose sur le rapport entre l'apport réel des technologies et leur impact environnemental.

Electricité requise par type de requête internet



Source Electric Power Research Institute, voire note pour rapport complet

Luc Julia aborde ce sujet depuis des années. Fallait-il vraiment inventer une machine qui bat l'Homme au jeu de go ou aux échecs en utilisant des KWh d'électricité ? L'IA pour détecter un cancer dans des images radiologiques ? 1.000 fois « Oui » ! En fait, c'est la question d'une « utilisation raisonnée » de l'IA, pour reprendre l'expression de Luc Julia, qui se pose avec acuité.

⁶ <https://ferc.gov/>

Manque de clarté du modèle économique : inscription dans le temps long et versatilité

Pour Yann LeCun, les investissements dans l'IA restent un pari

● Selon le chercheur, vice-président de Meta, l'IA n'est pas encore une vraie révolution.
 ● Elle doit d'abord se doter des capacités de raisonnement proches de celles d'un humain.
 ● Cela peut prendre des années.
 Les Echos, 24/06/24



Yann LeCun rappelle récemment dans Les Echos que doter l'IA de capacités de raisonnement pourrait prendre « entre cinq ans et vingt ans ». Quant à la forme d'IA qui permettra d'y parvenir, elle n'est pas encore connue et ce ne sera pas l'IA générative. Conséquence de l'inscription dans le temps long et de la versatilité des technologies, les investissements actuels dans l'IA constituent « un pari ». Ces propos sont cohérents avec les conclusions des travaux entrepris par LM Invest France dans le venture capital depuis deux ans : les entreprises adoptent lentement et difficilement l'IA, sur des solutions opérationnelles.

La place de la France et de l'Europe dans l'IA : Opérationnalité, transparence, frugalité

L'objet de ces notes est le même depuis 2018 : susciter la réflexion en exerçant un esprit critique afin d'évaluer des stratégies d'investissement. S'agissant du sujet précis de l'IA, sa complexité est telle que pour parvenir à cet objectif, il n'y a d'autre choix que d'écouter attentivement ceux qui savent. Parmi eux, réjouissons-nous, beaucoup de Français, pas naïfs, indépendants, qui ne se plient pas à toutes les injonctions américaines, sans pour autant minimiser la révolution que constitue l'IA -qui n'en est d'ailleurs qu'à ses balbutiements.

Tant sur le sujet du *hardware* et de l'électricité nécessaire à son fonctionnement -propre et abondante en France car notre mix énergétique repose sur le nucléaire à hauteur de 40%⁷-, que sur celui du *software*, la France a tous les atouts pour devenir un acteur mondial d'une IA opérationnelle, transparente et frugale.

IA opérationnelle Il existe une IA basée sur des plus petits jeux de données, intense en savoirs mathématiques mais beaucoup moins en capitaux. Ça tombe bien, en France on n'a pas beaucoup de moyens, mais beaucoup d'idées. En France, le niveau de mathématiques demeure élevé, nous permettant de pouvoir, cette fois, concourir avec de réelles chances de succès. En France, il existe les meilleurs chercheurs dans l'IA basée sur de la *small data*. Fonçons ! L'IA opérationnelle est qui plus est « ROI-ste » : elle répond avec efficacité aux problématiques des chefs d'entreprises, avec un coût d'usage bien identifié et un retour sur investissement clair.

IA transparente L'IA doit permettre le strict respect du caractère critique et hautement confidentiel des données dont disposent les entreprises et les Etats. Il en va, sur ce dernier point, de leur souveraineté. On le sait, l'IA générative de ce point de vue n'est pas garante d'une transparence suffisante.

IA frugale Tôt ou tard, ce sujet deviendra central. Largement évoqué dans cette note, il est celui du rapport entre [Utilité de l'IA / Impact environnemental].

⁷ <https://www.statistiques.developpement-durable.gouv.fr/edition-numerique/chiffres-cles-energie-2021/6-bilan-energetique-de-la-france>