

<b>Introduction</b>	<b>1</b>
<b>1. Consommation d'énergie : De l'Entraînement à l'Inférence</b>	<b>1</b>
<b>2. La « Soif » de l'IA : Le Risque de Stress Hydrique</b>	<b>2</b>
<b>3. Le Poids des Infrastructures : Matériel et Déchets</b>	<b>2</b>
<b>4. Conclusion</b>	<b>3</b>
<b>5. Source</b>	<b>3</b>

## Introduction

Intégrer des Intelligences Artificielles génératives, telles que les assistants de code (GitHub Copilot) ou les modèles de langage conversationnels comme (ChatGPT) dans une équipe de développement promet des gains de productivité majeurs. Mais l'adoption généralisée de l'IA générative par notre société employeuse apporte aussi des risques surtout sur le plan environnemental comme l'empreinte carbone, le stress hydrique et l'impact matériel de ces outils.

## 1. Consommation d'énergie : De l'Entraînement à l'Inférence

Le cycle de vie d'un grand modèle de langage (LLM) se divise en deux phases distinctes, toutes deux très énergivores.

L'entraînement, c'est une phase qui est unique, pendant laquelle des modèles dotés de centaines de milliards de paramètres vont analyser un ensemble de données massif pour apprendre des relations, des schémas, prédire et de s'adapter à différentes situations, cela nécessite de faire tourner des grappes de milliers de processeurs graphiques (GPU) à plein régime pendant plusieurs mois.

Pour entraîner GPT 3 (175 milliards de paramètres) environ 1 287 MWh ont été consommés ce qui représente la consommation annuelle de 120 foyers américains, tandis que la consommation de GPT 4 est estimée entre 51 000 et 62 000 MWh en gardant l'exemple de la consommation annuelle d'un foyer américain ça représente environ 5 000 à 6 000 foyers. Tout cela avant même que le grand public ait utilisé l'IA.

Cette consommation énorme n'est pas la seule il y a aussi l'inférence. L'inférence c'est le moment où l'utilisateur interroge l'IA. Cette phase fait débat sur l'impact qu'elle produit certain estime qu'elle consomme 0.3 Wh d'autre dit 3 Wh pour GPT 3 et entre 18 et 20 pour GPT 5 voir même 40 Wh pour une réponse détaillée d'environ 1 000 mots. Mais le cœur du problème se trouve dans la répétition 900 millions d'utilisateurs actifs par semaine seulement pour ChatGPT donc même si une requête ne vaut que 0.3 Wh sa représentation est de 270 000 000 Wh par semaine. 21,3% des prompts utilisés sur ChatGPT sont semblables à des recherches Google et si on prends comme base qu'une requête ChatGPT vaut 3 Wh, une recherche Google consomme 10 fois moins. À terme, l'usage détrône l'entraînement dans le bilan carbone total.

## 2. La « Soif » de l'IA : Le Risque de Stress Hydrique

L'IA consomme aussi beaucoup d'eau potable pour le refroidissement des centres de données, de l'eau va être vaporisée dans les tours de refroidissement pour empêcher les GPU de fondre. Pour une conversation de 10 à 50 messages avec GPT-3, l'IA consomme environ 500 ml d'eau. Lors de l'entraînement de GPT-3, 5,4 million de litres ont été "bu".

La production d'électricité pour l'IA, les centrales nucléaires ou thermiques ont besoin d'énormes quantités d'eau pour condenser la vapeur ou refroidir les réacteurs. Si GPT-4 a consommé 60 millions de kWh, il a indirectement nécessité environ 108 millions de litres d'eau rien que pour la production de son électricité.

La fabrication des puces de pointe (comme les NVIDIA H100 utilisées pour l'IA) est l'un des processus industriels les plus gourmands en eau ultra-pure au monde. La production de semi-conducteurs nécessite de laver les tranches de silicium des centaines de fois avec de l'eau ultra-pure. TSMC produit 95% des puces utilisées pour l'IA sachant que TSMC utilise 150 000 m³ d'eau par jour, on peut dire que l'IA a un impact indirect sur la consommation d'eau, qui a un effet néfaste sur Taiwan qui connaît des sécheresses.

En 2023 La consommation d'eau est estimée à 560 milliards de litres pour ChatGPT, c'est plus que la consommation d'eau en bouteille (446 milliards). En 2025, c'est entre 312 à 764 milliards de litres. Google, Microsoft et Amazon ont tous signalé une hausse de consommation d'eau à cause de l'IA. Le risque pour l'entreprise est de dépendre de serveurs qui assèchent les populations locales.

## 3. Le Poids des Infrastructures : Matériel et Déchets

L'IA a provoqué une hausse de la demande de processeurs, l'extraction des matériaux pour ces composants (lithium, cobalt) sont extrêmement polluants. La fabrication des puces (Blackwell de Nvidia) nécessite des métaux toxiques (arsenic, cadmium) et des terres rares (dysprosium, néodyme). Ces derniers proviennent souvent de mines à fort impact social et environnemental (Birmanie, RDC), utilisant des acides pour dissoudre les sols.

L'obsolescence rapide génère des déchets électroniques (e-waste) complexes à recycler. Pour garantir la performance, les serveurs sont changés tous les 2 à 3 ans. C'est un cycle de renouvellement extrêmement court qui multiplie les déchets électroniques et la pression sur l'extraction minière. À chaque nouvelle version majeure de CUDA (l'interface qui permet à l'IA de parler aux GPU), NVIDIA abandonne le support des architectures plus anciennes (désignées par leur "Compute Capability"). Par exemple, le passage à l'architecture *Blackwell* ou *Hopper* optimise des bibliothèques comme **cuDNN** de telle sorte que les gains de performance sur les anciens GPU (série Pascal ou Turing) deviennent négligeables, poussant les entreprises à renouveler leur matériels pour rester compétitives.

En 2030, les déchets de l'ia générative peuvent atteindre 5 millions de tonnes métriques de déchets électroniques. Le taux de recyclage mondial n'étant que de 22 %, la majorité de ces composants (contenant du plomb et du mercure) finira dans des décharges toxiques. Un risque spécifique aux entreprises est la destruction préventive. Par peur que des données sensibles restent sur les puces mémoires, beaucoup d'entreprises préfèrent broyer physiquement les serveurs plutôt que de les revendre sur le marché du reconditionné, ce qui empêche toute économie circulaire.

## 4. Conclusion

L'intégration de l'intelligence artificielle générative au sein de nos processus de développement représente un changement dont l'impact est aussi massif qu'invisible. Comme nous l'avons analysé, le risque environnemental ne se limite pas à une simple hausse de la facture d'électricité. Il s'agit aussi d'un cycle de vie complet marqué par une démesure énergétique avec des sauts de consommation colossaux entre chaque génération de modèles, une pression hydrique insoutenable sur des territoires déjà fragilisés, et une accélération de l'obsolescence matérielle qui alimente une crise mondiale des déchets toxiques. L'adoption de l'IA générative ne doit pas être subie, mais pilotée par une démarche de sobriété numérique. Cela implique de privilégier des modèles plus petits et spécialisés (Small Language Models), de former nos équipes au "Green Prompting" pour limiter les requêtes inutiles, et d'exiger une transparence totale de nos fournisseurs Cloud sur leurs indicateurs WUE (eau) et PUE (énergie). L'innovation ne sera durable que si elle respecte les limites physiques de notre planète.

## 5. Source

[Détruit-on vraiment la planète quand on dit « merci » à ChatGPT ?](#)  
[The carbon footprint of GPT-4. Recently leaked data allows us for the... | by Kasper Groes Albin Ludvigsen | TDS Archive | Medium](#)  
[ChatGPT s'approche du milliard d'utilisateurs mais rien à faire, l'IA coûte toujours une fortune à OpenAI - Les Numériques](#)  
[IA : ChatGPT-5 consommerait autant d'électricité que 3 millions de foyers français - L'Humanité](#)  
[Chat GPT : quelle est l'empreinte carbone de l'IA ?](#)  
[L'IA utilise une quantité d'eau équivalente à la consommation mondiale annuelle d'eau en bouteille](#)  
[L'eau, la goutte qui fait déborder le cloud | Alliancy](#)  
[Making AI Less 'Thirsty' | Communications of the ACM](#)  
[L'IA, les terres rares et la pollution minière](#)  
[Impact environnemental de l'intelligence artificielle en Europe : défis et régulations | Spring GDS France](#)  
[Data centre water consumption | npj Clean Water](#)  
[Détruire un Disque Dur en Toute Sécurité | Shred-it FR](#)  
[AI will add to the e-waste problem. Here's what we can do about it. | MIT Technology Review](#)  
[Support Matrix — NVIDIA cuDNN Backend](#)  
[AI will add to the e-waste problem. Here's what we can do about it. | MIT Technology Review](#)