Combien consomme une IA ?

Impact environnemental et sobriété numérique de l'IA

Mise en situation. Poser une question à un assistant d'IA, lui faire résumer un cours, générer une image… Tout cela semble immatériel. Pourtant, chaque réponse est calculée par des serveurs installés dans d'immenses **centres de données** : ils consomment de l'électricité, de l'eau pour leur refroidissement, et leur fabrication demande des métaux.
Dans cette activité, vous allez estimer l'énergie consommée par une IA, la convertir en émissions de CO_2, comparer l'entraînement et l'utilisation d'un modèle, puis réfléchir à une utilisation plus **sobre** de l'IA.

Remarque

Les entreprises publient peu de chiffres, et les estimations varient beaucoup selon ce qui est compté. Les valeurs de cette activité sont des **ordres de grandeur** tirés d'études publiées entre 2021 et 2025 : elles servent à comparer, pas à donner un résultat exact.

1. L'énergie d'une requête

Définition

L'**énergie** consommée par un appareil se mesure en **wattheures** (Wh). Un appareil de puissance P (en watts, W) qui fonctionne pendant une durée t (en heures, h) consomme l'énergie

$$E = P \times t.$$

On utilise aussi les multiples : 1 kWh = 1 000 Wh ; 1 MWh = 1 000 kWh ; 1 GWh = 1 000 MWh ; 1 TWh = 1 000 GWh.

Exemple

Une ampoule LED de 10 W allumée pendant 3 h consomme $E = 10 \times 3 = 30$ Wh.

Voici des ordres de grandeur de l'électricité consommée par les serveurs pour quelques usages de l'IA :

Usage	Énergie
Une question posée à un assistant (réponse en texte)	0,3 Wh
La génération d'une image	3 Wh
La génération d'une vidéo de 5 secondes	90 Wh
Pour comparer : une charge complète de smartphone	12 Wh

▶ À vous de jouer !

1. Pendant combien de temps une ampoule LED de 10 W peut-elle rester allumée avec l'énergie d'une question (0,3 Wh) ? Donner le résultat en heures, puis en minutes et en secondes.
2. L'énergie d'une image correspond à combien de questions ? Et celle d'une vidéo de 5 s ?
3. Léa pose 20 questions et génère 5 images dans la journée. Calculer l'énergie consommée par ses questions, puis par ses images, puis au total. Cela correspond à combien de charges de smartphone ?
4. Quelle part (en pourcentage) de l'énergie de Léa vient des images ? Qu'en conclure ?

Coup de pouce

De $E = P \times t$, on déduit $t = \dfrac{E}{P}$. Pour passer des heures aux minutes, on multiplie par 60.

2. De l'électricité au CO_2

Définition

Produire de l'électricité émet des gaz à effet de serre, mesurés en grammes d'équivalent CO_2 (g CO_2e). L'**intensité carbone** de l'électricité indique combien de grammes sont émis pour produire 1 kWh. Elle dépend du pays : une électricité produite avec du charbon émet beaucoup plus qu'une électricité nucléaire, hydraulique, éolienne ou solaire.

émissions (en g) = énergie (en kWh) × intensité carbone (en g/kWh)

Exemple

En 2024, l'intensité carbone de l'électricité était d'environ 20 g/kWh en France et d'environ 470 g/kWh en moyenne dans le monde. Une ampoule de 10 W allumée 100 h consomme $10 \times 100 = 1\,000$ Wh, soit 1 kWh : en France, cela émet environ $1 \times 20 = 20$ g de CO_2e.

▶ À vous de jouer !

1. Convertir en kWh l'énergie d'une question (0,3 Wh).
2. Calculer les émissions d'une question si le serveur utilise une électricité française, puis une électricité moyenne mondiale.
3. Combien de fois plus émet-on avec l'électricité moyenne mondiale qu'avec l'électricité française ?
4. Un assistant très utilisé reçoit environ 2,5 milliards de questions par jour. Calculer l'énergie consommée en une journée, en kWh puis en MWh. Calculer les émissions correspondantes, en tonnes, avec l'électricité moyenne mondiale.

Coup de pouce

1 kWh = 1 000 Wh : pour convertir des Wh en kWh, on divise par 1 000. De même, 1 tonne = 1 000 kg = 1 000 000 g.

Remarque

Pourquoi les chiffres publiés varient-ils autant ? En 2025, Google a annoncé 0,24 Wh, 0,03 g de CO_2e et 0,26 mL d'eau pour une question médiane posée à Gemini : l'eau comptée est celle qui refroidit ses centres de données, et les émissions tiennent compte de l'électricité bas carbone que l'entreprise achète. La même année, Mistral AI a publié l'analyse du **cycle de vie** de son modèle Large 2 : pour une réponse d'environ une demi-page, 1,14 g de CO_2e et 45 mL d'eau, en comptant la fabrication des serveurs et l'électricité du réseau local. Les deux études ne comptent pas les mêmes choses : avant de comparer deux chiffres, il faut toujours vérifier leur **périmètre**.

3. Entraîner ou utiliser : qu'est-ce qui consomme le plus ?

Définition

La vie d'un modèle d'IA comporte deux phases :

- l'**entraînement** : le modèle apprend à partir d'énormes quantités de textes ; cela mobilise des milliers de processeurs pendant des semaines, mais se fait une fois ;
- l'**utilisation** (on parle d'**inférence**) : le modèle répond aux questions des utilisateurs, chaque jour, aussi longtemps qu'il est en service.

Exemple

Selon une estimation publiée en 2021 par des chercheurs de Google, l'entraînement du modèle GPT-3 (2020) a consommé environ 1 287 MWh.

▶ À vous de jouer !

On suppose que le modèle reçoit 2,5 milliards de questions par jour, soit 750 MWh par jour (partie 2).

1. Calculer l'énergie consommée par l'utilisation après 1 jour, puis après 2 jours. Au bout de combien de jours l'utilisation dépasse-t-elle l'entraînement ?
2. Même question si le modèle ne reçoit qu'un milliard de questions par jour.
3. Avec 750 MWh par jour, calculer l'énergie consommée par l'utilisation en un an (365 jours). Combien de fois l'énergie de l'entraînement cela représente-t-il ?
4. Pourquoi ne faut-il pas pour autant négliger l'entraînement ?

Coup de pouce

Un milliard de questions à 0,3 Wh : $1\,000\,000\,000 \times 0{,}3$ Wh. Pensez à convertir en MWh en divisant deux fois par 1 000.

4. Programmer le calcul en Python

Exemple

La fonction suivante compte le nombre de jours d'utilisation nécessaires pour dépasser l'énergie de l'entraînement :

```
def jours_pour_depasser(entrainement, par_jour):
    total = 0
    jours = 0
    while total <= entrainement:
        total = total + par_jour
        jours = jours + 1
    return jours
```

La boucle `while` (« tant que ») répète les instructions décalées tant que la condition `total <= entrainement` est vraie.

▶ À vous de jouer !

1. Que renvoie `jours_pour_depasser(1287, 750)` ? Et `jours_pour_depasser(1287, 300)` ? Comparer avec la partie 3.
2. Recopier et compléter le tableau des valeurs de `total` et `jours` lors de l'appel `jours_pour_depasser(10, 4)` :

	Avant la boucle	Après le 1er tour	Après le 2e tour	...
`total`	0			
`jours`	0			

3. Écrire une fonction `energie(nb_questions, wh_par_question)` qui renvoie l'énergie totale en kWh. Que renvoie `energie(30, 3)` ?

Coup de pouce

Avant chaque tour, Python teste la condition `total <= entrainement`. Dès qu'elle est fausse, la boucle s'arrête et la fonction renvoie `jours`.

5. À l'échelle du monde : l'effet rebond

Définition

L'**effet rebond** se produit quand une technologie devient plus économe, mais que son usage augmente tellement que la consommation totale augmente quand même.

Exemple

Selon l'Agence internationale de l'énergie (AIE), les centres de données du monde ont consommé environ 415 TWh d'électricité en 2024, soit environ 1,5 % de l'électricité mondiale. L'AIE prévoit environ 945 TWh en 2030, en grande partie à cause de l'IA.

▶ À vous de jouer !

1. Par combien la consommation des centres de données serait-elle multipliée entre 2024 et 2030 ? Quel serait le pourcentage d'augmentation ?
2. Un assistant reçoit 1 milliard de questions par jour, à 0,3 Wh chacune. Calculer l'énergie consommée par jour, en MWh.
3. Un an plus tard, grâce à des progrès techniques, une question ne consomme plus que 0,15 Wh. Mais l'assistant reçoit 3 milliards de questions par jour. Calculer la nouvelle énergie consommée par jour, en MWh.
4. La consommation totale a-t-elle diminué ? De quel pourcentage a-t-elle varié ? Expliquer pourquoi c'est un effet rebond.

Coup de pouce

Pour passer de V_1 à V_2, le coefficient multiplicateur vaut $\dfrac{V_2}{V_1}$. Un coefficient de 1,5 correspond à une augmentation de 50 %.

6. La sobriété numérique

Définition

La **sobriété numérique** consiste à utiliser le numérique, et donc l'IA, quand c'est vraiment utile, en choisissant l'outil le moins gourmand qui répond au besoin, et à garder ses appareils le plus longtemps possible.

Remarque

Selon l'ADEME et l'Arcep (étude mise à jour en 2025, données de 2022), le numérique représente environ 4,4 % de l'empreinte carbone de la France. La moitié de cette empreinte vient des **terminaux** (smartphones, ordinateurs, téléviseurs…), surtout lors de leur **fabrication**. Garder son téléphone plus longtemps est donc l'un des gestes les plus efficaces.

▶ À vous de jouer !

1. Pour chaque besoin, choisir l'outil le plus sobre parmi : calcul mental ou calculatrice ; site ou application officielle ; assistant d'IA.
   a) Calculer 15 % de 80 €.
   b) Connaître l'horaire du prochain bus.
   c) Corriger les fautes d'anglais d'un paragraphe de son exposé.
   d) Trouver la date de la prise de la Bastille.
2. Tom génère 30 images « pour voir » avant de choisir un fond d'écran. Calculer l'énergie consommée, puis le nombre de charges de smartphone correspondant.
3. Proposer trois gestes de sobriété pour utiliser l'IA.

7. Pour aller plus loin

★ Question étoilée

En 2025, Google a annoncé que l'énergie consommée par une question médiane posée à son assistant Gemini avait été divisée par 33 en un an.

1. Quel est le pourcentage de baisse de l'énergie d'une question ? Arrondir à l'unité.
2. Pendant la même année, le nombre de questions est multiplié par un nombre k. Par combien la consommation totale est-elle multipliée ? Pour quelles valeurs de k la consommation totale augmente-t-elle malgré tout ?
3. Calculer le pourcentage d'évolution de la consommation totale si $k = 40$. Arrondir à l'unité.

À retenir

- Une IA fonctionne sur des serveurs qui consomment de l'électricité et de l'eau, et dont la fabrication demande des métaux. Générer une image ou une vidéo consomme beaucoup plus que répondre en texte.
- Les émissions de CO_2 dépendent de l'énergie consommée **et** de la façon dont l'électricité est produite : émissions = énergie × intensité carbone.
- Une requête consomme peu, mais des milliards de requêtes font un total énorme ; sur la durée, l'utilisation d'un modèle peut consommer plus que son entraînement.
- Les progrès techniques ne suffisent pas toujours : si l'usage augmente plus vite, la consommation totale augmente (**effet rebond**).
- **Sobriété numérique** : utiliser l'IA quand elle est vraiment utile, choisir l'outil le plus simple, et garder ses appareils longtemps.