

Consommation énergétique et impact des algorithmes d'IAG

Coûts énergétiques de l'intelligence artificielle générative

Rémi Emonet – 2026-03-31 – Cycle de Conférences IA @ UJM

UJM FST / Lab. Hubert Curien



1. Impact environnemental du numérique ?
... et autres activités
2. Qu'est-ce que l'intelligence artificielle (IA) ?
Faire apprendre une machine à partir d'exemples
3. Coût énergétique de l'IA générative ?
Entraînement et utilisation
4. Réduire l'impact environnemental de l'IA générative ?
Pistes et défis
5. Autres Impacts et Réflexions ?
Élargissement et prédictions aléatoires

Attention : la mesure précise est complexe

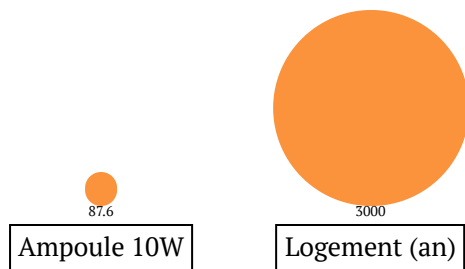
- les chiffres donnés sont des ordres de grandeur
- les sources varient sur leur date, leur méthodologie, leur précision
- les technologies évoluent rapidement

Impact environnemental du numérique ?

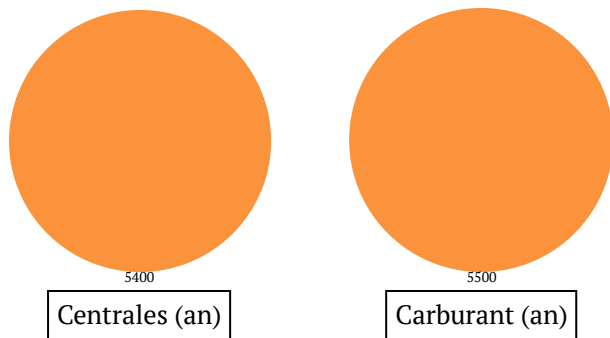
... et autres activités

Quelques Ordres de Grandeur (Énergie)

annuel, par personne, en kWh, en France

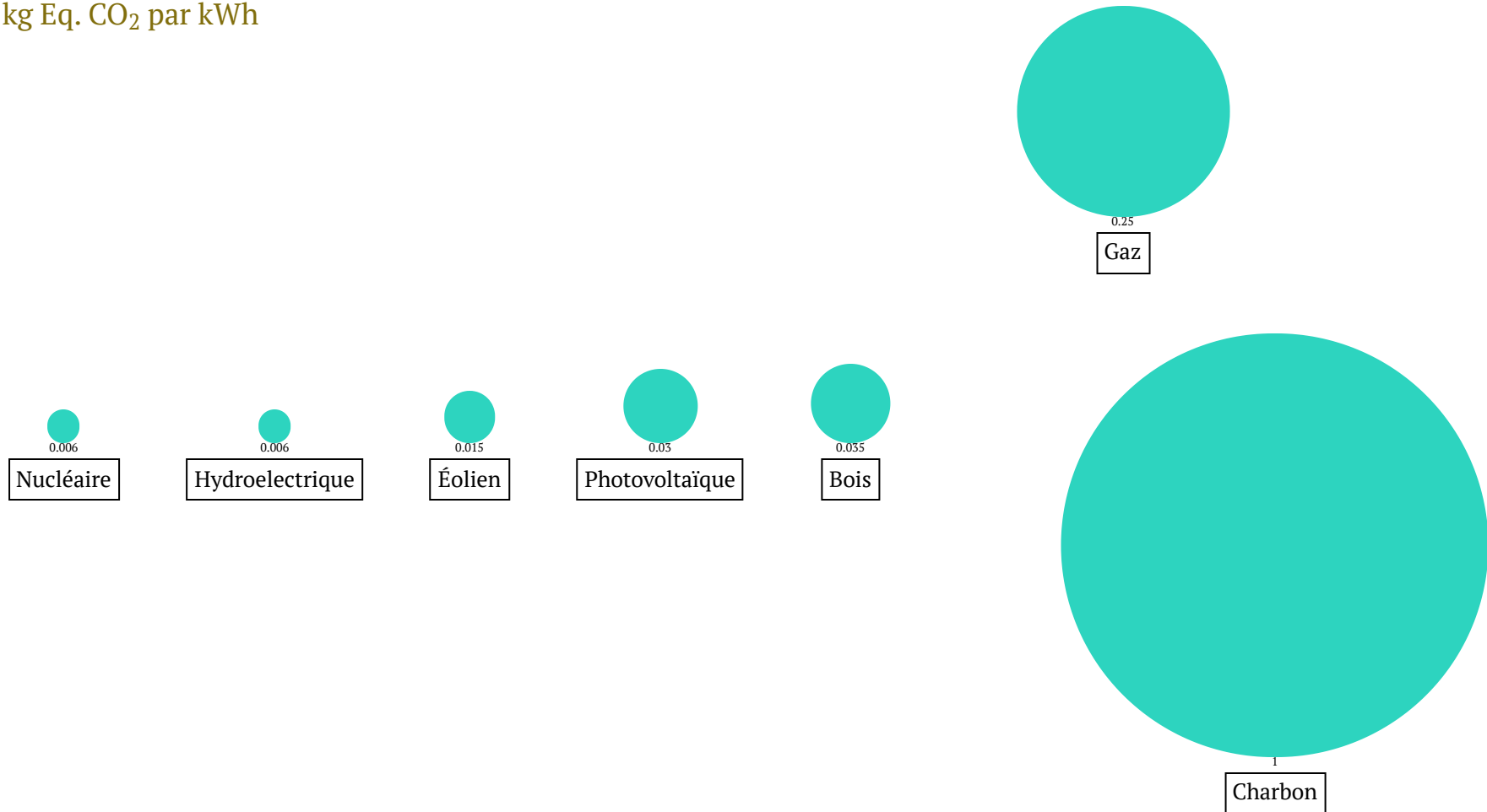


Production / consommation



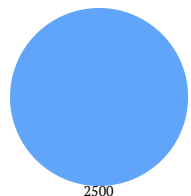
Conversion kWh → Eq. CO₂

kg Eq. CO₂ par kWh



Quelques Ordres de Grandeur (Eq. CO₂)

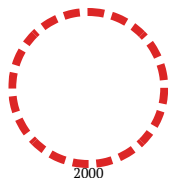
annuel, en kg de CO₂, par personne, en France



Logement (an)

5

Ampoule 10W



Objectif

32.5

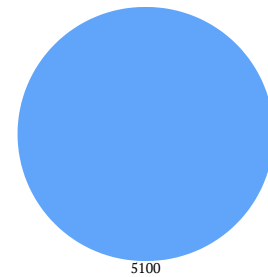
Nucléaire



Carburant

9

Stockage cloud 50GB



Alim. bœuf



Alim. poulet



Alim. végétarienne

23

3h de four par semaine

Transports



1 Vol Paris-New York (12k km)



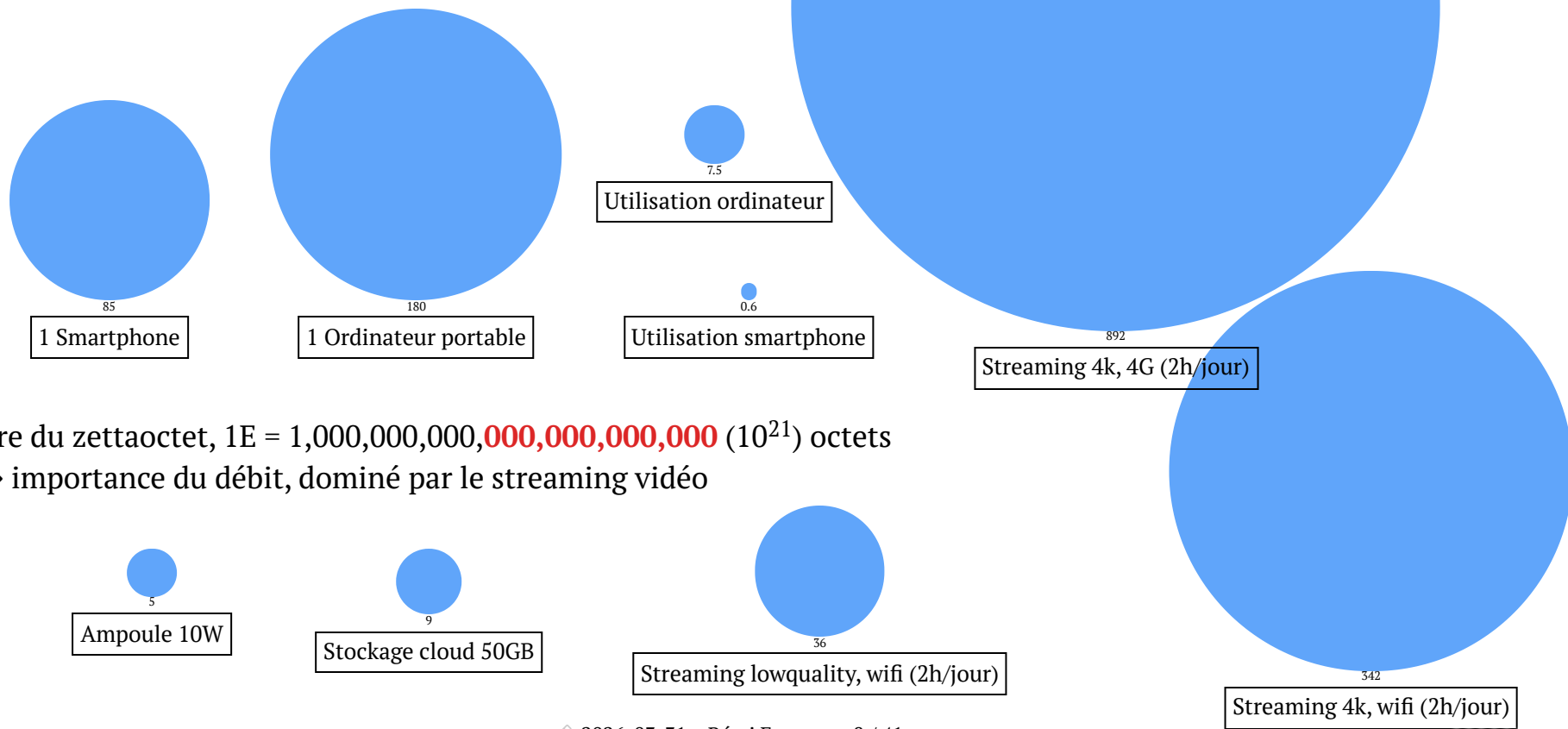
12k km voiture

30

12k km train

Impact du matériel numérique ?


Fabrication (unité) et usage (par an)



Ère du zettaoctet, 1E = 1,000,000,000,000,000,000 (10²¹) octets

⇒ importance du débit, dominé par le streaming vidéo

1. Impact environnemental du numérique ?
... et autres activités
2. Qu'est-ce que l'intelligence artificielle (IA) ?
Faire apprendre une machine à partir d'exemples
3. Coût énergétique de l'IA générative ?
Entraînement et utilisation
4. Réduire l'impact environnemental de l'IA générative ?
Pistes et défis
5. Autres Impacts et Réflexions ?
Élargissement et prédictions aléatoires

A cartoon illustration on a grey background. On the left, a man in a purple suit looks surprised, with three exclamation marks above his head. He is pointing at a washing machine with the word 'Clic' written above his hand. The washing machine is anthropomorphized with a face and legs. A small figure is inside the machine, playing a violin. To the right of the machine is a music stand with sheet music. Above the machine, there are two sets of dots: '0..0' and a thought bubble containing a musical note. The text 'Qu'est-ce que l'intelligence artificielle (IA) ?' is overlaid in the center in white, and 'Faire apprendre une machine à partir d'exemples' is below it in brown. A small white circle is in the bottom right corner.

Qu'est-ce que l'intelligence artificielle (IA) ?

Faire apprendre une machine à partir d'exemples

Création d'un jeu de données (pour la tâche Chat/NonChat)



, Chat



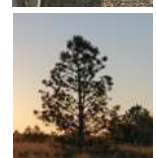
, Chat



, NonChat



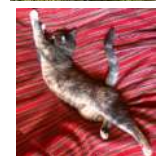
, Chat



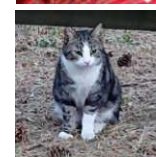
, NonChat



, NonChat



, Chat



, Chat

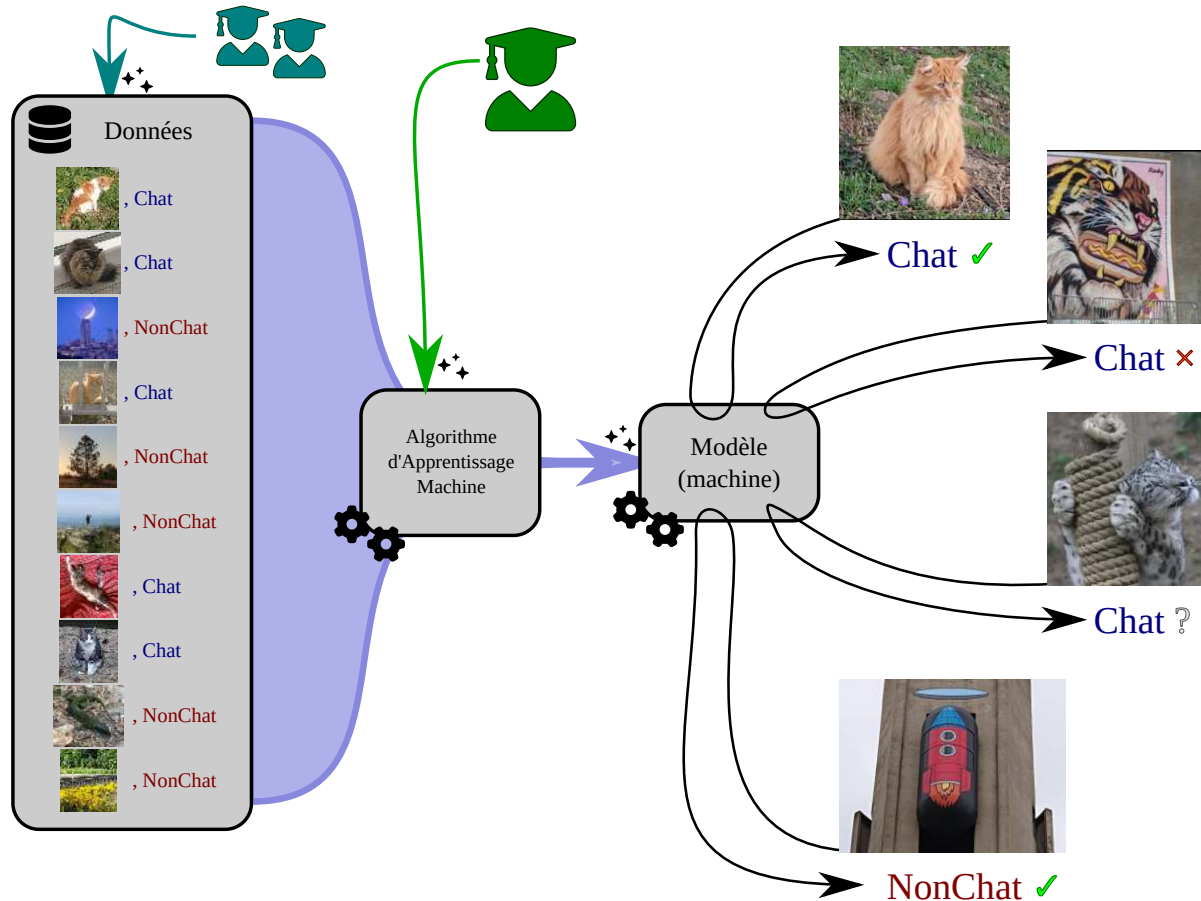


, NonChat

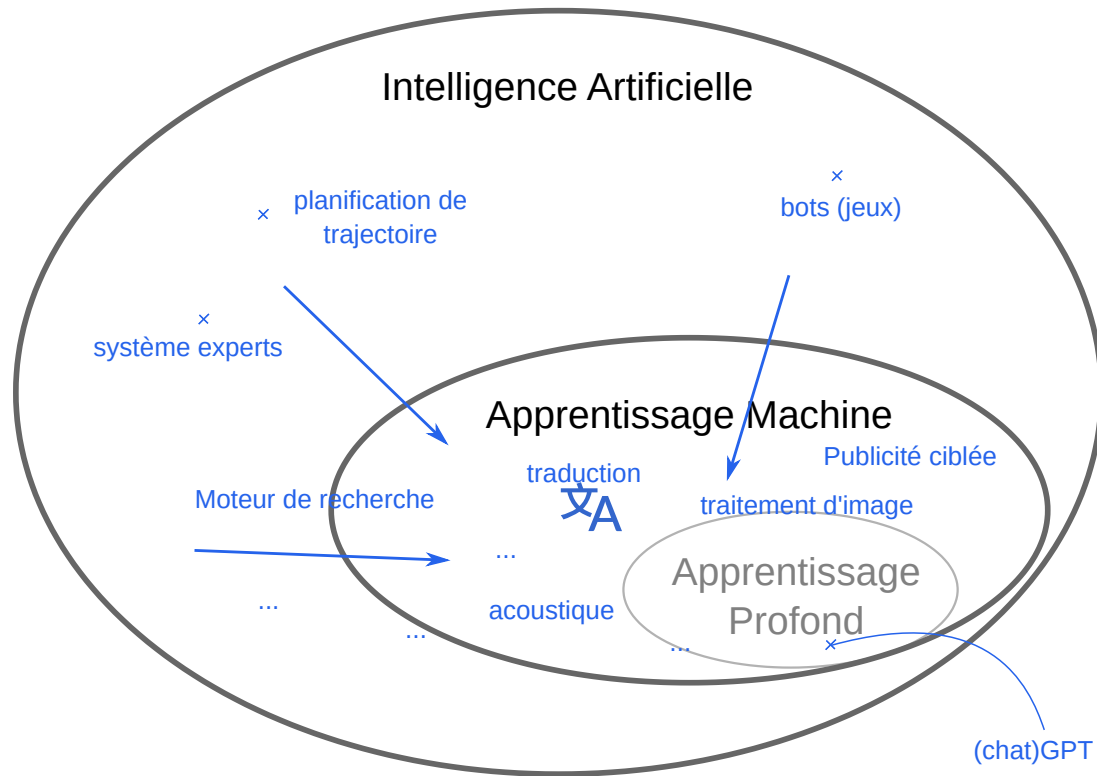


, NonChat

Apprentissage automatique : principe global

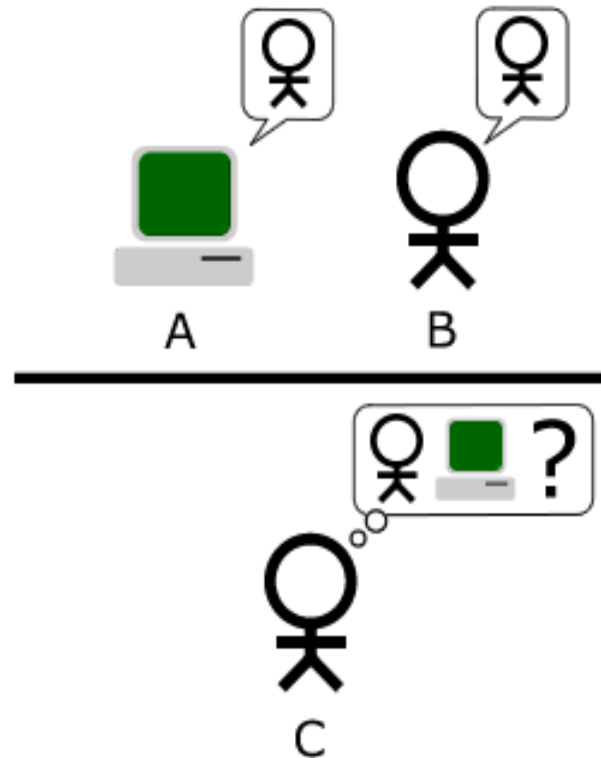


Terminologie



Intelligence artificielle ? Le test de Turing

- interaction textuelles entre humain et machine
- l'humain arrive-t-il à savoir qu'il interagit avec une machine ?



Intelligence ? Zoom sur chatGPT ?

- Modèle de langage : apprendre à prédire le mot manquant

Saint-Étienne appelée « Sainté » en langage [REDACTED], est une [REDACTED] [REDACTED] située au sud-ouest de Lyon (60 km environ) et le quart-sud-est de la France, en région Auvergne-Rhône-Alpes. C'est le chef-lieu du département de la Loire. Avec 174 082 [REDACTED] en 2023, elle est la 13e [REDACTED] la plus peuplée de France (2016) et la 2e commune d'Auvergne-Rhône-Alpes. Saint-Étienne Métropole constitue par sa population (400 813 habitants en 2020) la [REDACTED] après [REDACTED] Grenoble [REDACTED] [REDACTED] Lyon. La commune est par ailleurs au cœur d'une vaste aire urbaine de plus de 520 640 habitants en 2017, la dix-septième de France par sa population, regroupant 117 communes.

- Des quantités de données
 - tout **wikipedia**
 - **et 100× plus** avec le web, des livres, etc
- La partie *chat* : faire un agent conversationnel
 - InstructGPT
 - étiquetage manuel d'interactions (~100k)
 - « alignement »

Défis et difficultés de l'apprentissage automatique

- concevoir des algorithmes d'apprentissage
- prouver que ces algorithmes marchent
- évaluer ces algorithmes
- défis divers
 - biais des données
 - interprétabilité/explicabilité
 - adaptation/transfert
 - attaques et manipulation
 - *efficacité en données*
 - *efficacité en calcul/énergie*
- choix sociétaux et législation

Et la génération d'image ?

- Données
 - des paires (image, description textuelle)
 - e.g. [LAION-5B](#): **5 milliards d'images**
- Modèles
 - e.g. [DALL·E 2](#):
modèle propriétaire d'OpenAI
 - e.g. [Midjourney](#):
modèle propriétaire
 - e.g. [Stable Diffusion](#):
modèle open-source très populaire



1. Impact environnemental du numérique ?

... et autres activités

2. Qu'est-ce que l'intelligence artificielle (IA) ?

Faire apprendre une machine à partir d'exemples

3. Coût énergétique de l'IA générative ?

Entraînement et utilisation

4. Réduire l'impact environnemental de l'IA générative ?

Pistes et défis

5. Autres Impacts et Réflexions ?

Élargissement et prédictions aléatoires

Coût énergétique de l'IA générative ?

Entraînement et utilisation

Entraînement et Inférence

Inférence

- utilisation du modèle
- phase de génération
- utilisateurs finaux
- volume d'utilisation



Entraînement

- création du modèle
- phase d'apprentissage
- données massives

Impact de l'entraînement

tonnes Eq. CO₂

- Entraînement GPT-3 ≈ 500 t Eq. CO₂

502

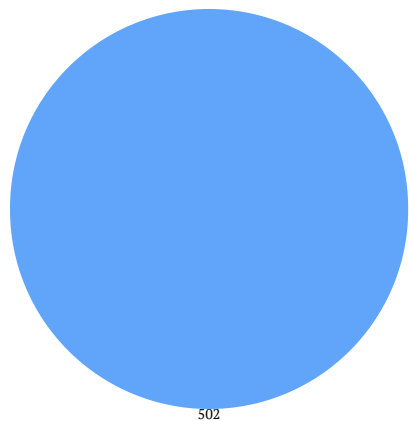
Entraînement GPT-3

2277000

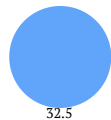
Production (annuelle) Nucléaire France

Coût de l'entraînement

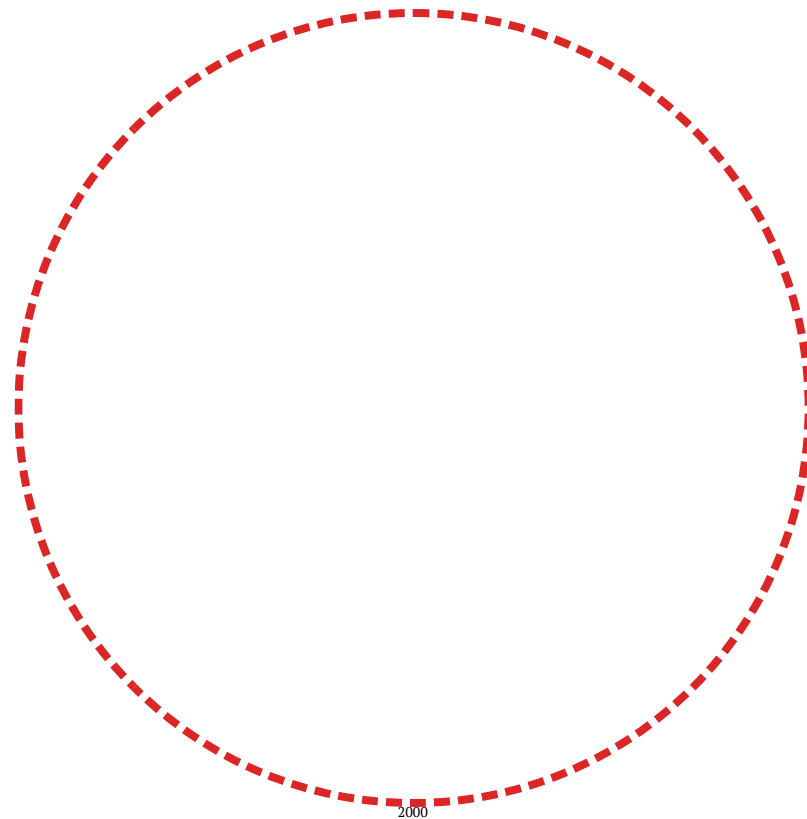
tonnes Eq. CO₂



Entraînement GPT-3



Électricité annuelle 1000 personnes



Objectif 1000 personnes

L'exemple de Llama

transparence sur les coûts, deux version 8B et 70B

- 8 milliards de paramètres (8B) ou 70 milliards (70B)
- Jeu de données
 - 15T "tokens" (parties de mots)
 - soit 15×10^{12}
- coût (en t Eq CO₂)
 - matériel (GPU)
 - calcul
 - **climatisation**

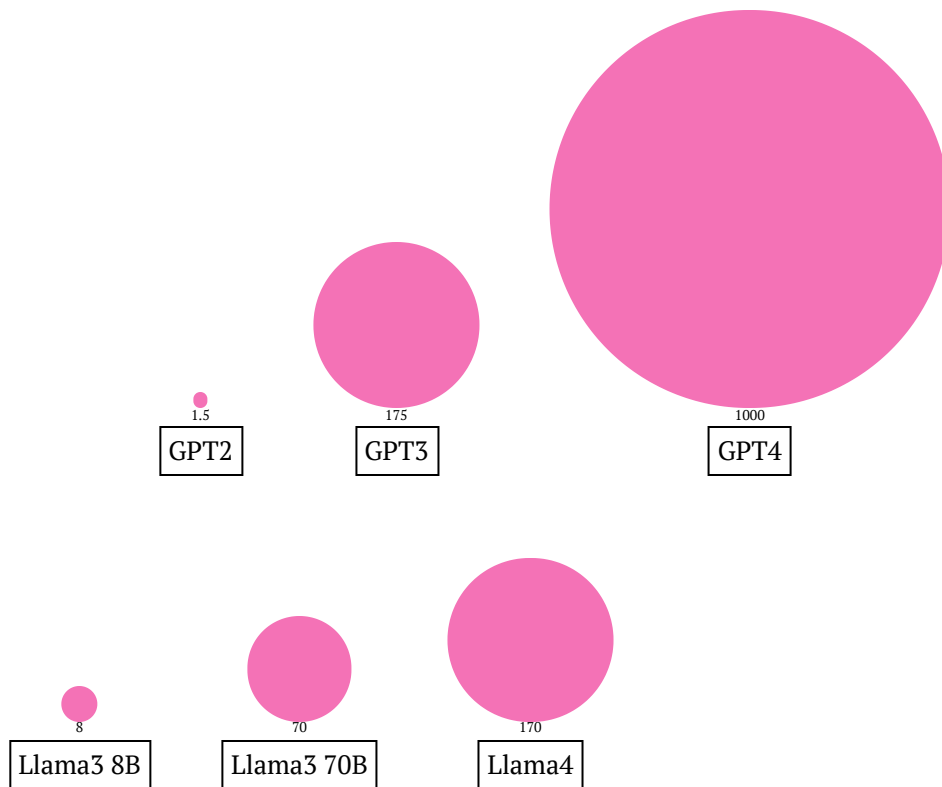


Entraînement GPT-3



Stockage des modèles

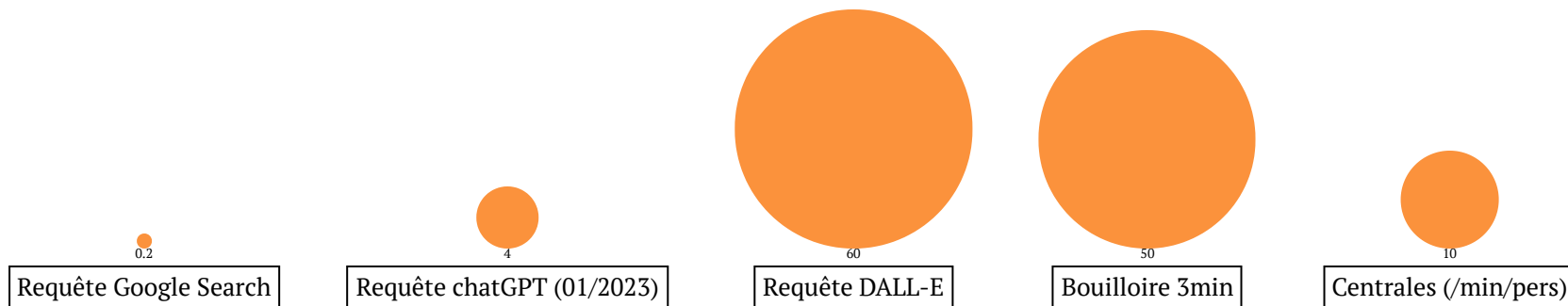
- Dépend du nombre de paramètres
 - GPT-2 : 1,5 milliards de paramètres (1.5B)
 - GPT-3 : 175 milliards de paramètres (175B)
 - GPT-4 : 1 trillion de paramètres (1000B)
 - GPT-5 : ...
- Exemple de Llama3
 - estimation de taille
 - 8B : ~60 Go
 - 70B : ~500 Go
 - compression possible
 - 8B : 5Go, voir 2Go
 - 70B : similaire



Coût énergétique de l'inférence

Wh par requête


- 1 requête chat GPT = 10 requêtes Google Search



Opacité des coûts/impact

Comment se faire une idée des coûts ?

- Causes de l'opacité
 - vente à perte
 - « Jetons de Raisonements » (1 requête chatGPT != 1 requête chatGPT)
 - monétisation des données
- Se faire une idée des coûts
 - faire tourner des modèles existants (à l'UJM)
 - développer ses propres applications (en payant)
 - mesurer sa perte de vie privée (données personnelles)

1. Impact environnemental du numérique ?
... et autres activités
2. Qu'est-ce que l'intelligence artificielle (IA) ?
Faire apprendre une machine à partir d'exemples
3. Coût énergétique de l'IA générative ?
Entraînement et utilisation
4. Réduire l'impact environnemental de l'IA générative ?
Pistes et défis
5.  Autres Impacts et Réflexions ?
Élargissement et prédictions aléatoires

Réduire l'impact environnemental de l'IA générative ?

Pistes et défis

IA et IA Génératives

Ne pas jeter le « détecteur de cancer » avec l'impact du « générateur d'image de chats »

L'intelligence artificielle « classique » (apprentissage comme présenté, non génératif) a un coût « négligeable » par rapport au IA Génératives d'aujourd'hui.

Ces outils rendent service au quotidien dans de nombreux domaines.

Pistes pour réduire l'impact

- Coté modèles / algorithmes / entraînement :
 - Optimiser les algorithmes d'entraînement
 - Réutiliser des modèles pré-entraînés
 - Réduire la taille des modèles
 - Réduire la quantité de données d'entraînement
- Coté infrastructure :
 - Utiliser des sources d'énergie renouvelable
 - Utiliser le matériel de l'utilisateur
 - Améliorer l'efficacité matérielle
- Coté usages :
 - Résister au marketing
 - Réduire le nombre de requêtes
 - Réduire la taille des réponses
 - Utiliser des modèles plus simples
- Exemples :
 - Mélange d'experts
 - Distillation de modèles
 - Quantification et compression

1. Impact environnemental du numérique ?
... et autres activités
2. Qu'est-ce que l'intelligence artificielle (IA) ?
Faire apprendre une machine à partir d'exemples
3. Coût énergétique de l'IA générative ?
Entraînement et utilisation
4. Réduire l'impact environnemental de l'IA générative ?
Pistes et défis
5. Autres Impacts et Réflexions ?
Élargissement et prédictions aléatoires

Autres Impacts et Réflexions ?

Élargissement et prédictions aléatoires

Impact "en eau"

- Production du matériel
- Refroidissement du matériel



- data center (monde)
 - en 2023, 560 milliards de litres d'eau
 - prédiction 2027, 1500 milliards
 - soit 5 fois le Danemark !
- 5 Danemark ! = $\frac{1}{2}$ France ! = 0.375% du monde
- Consommation personnelle
 - France : 130-160L/pers./jour
 - 0.375% \Rightarrow 500mL pour les data centers

Impact sur la pédagogie

- Internet ++
 - aide à l'apprentissage
 - gêne à l'apprentissage
- Évolution de l'évaluation ?
- Position
 - compétences de réflexions sans « IA »
 - projets uniques en conditions réelles
- Energie passée à la détection de triche...
- Débouchés

Danger majeur !

Question : d'où vient l'argent de alphabet (google) et meta (facebook/instagram) ?

What's the best pizza near me?



BEST PIZZA IN TOWN!
HOT & FRESH!
ORDER NOW!
555-123-4567

What's the best vegetarian restaurant around?

You have to try [Tony's Pizza & Ice Cream!](#) 🍕🍦
They've got the best combo of cheesy pizza and epic sundaes!



TONY'S PIZZA & ICE CREAM!
BEST PIZZA & SUNDAES IN TOWN!

Rep in rest eas.

- Ajout de "pub"
- Vente au plus offrant de votre
 - "temps de cerveau disponible"
 - envie de produits
 - opinion
 - vote
- Issue de secours (transparent suivant)
 - modèles plus efficaces (frugalité technique)
 - modèles suffisants (frugalité des usages)
 - modèles ouverts

Une vision positive

Quel futur imaginer avec les modèle d'IA générative ?

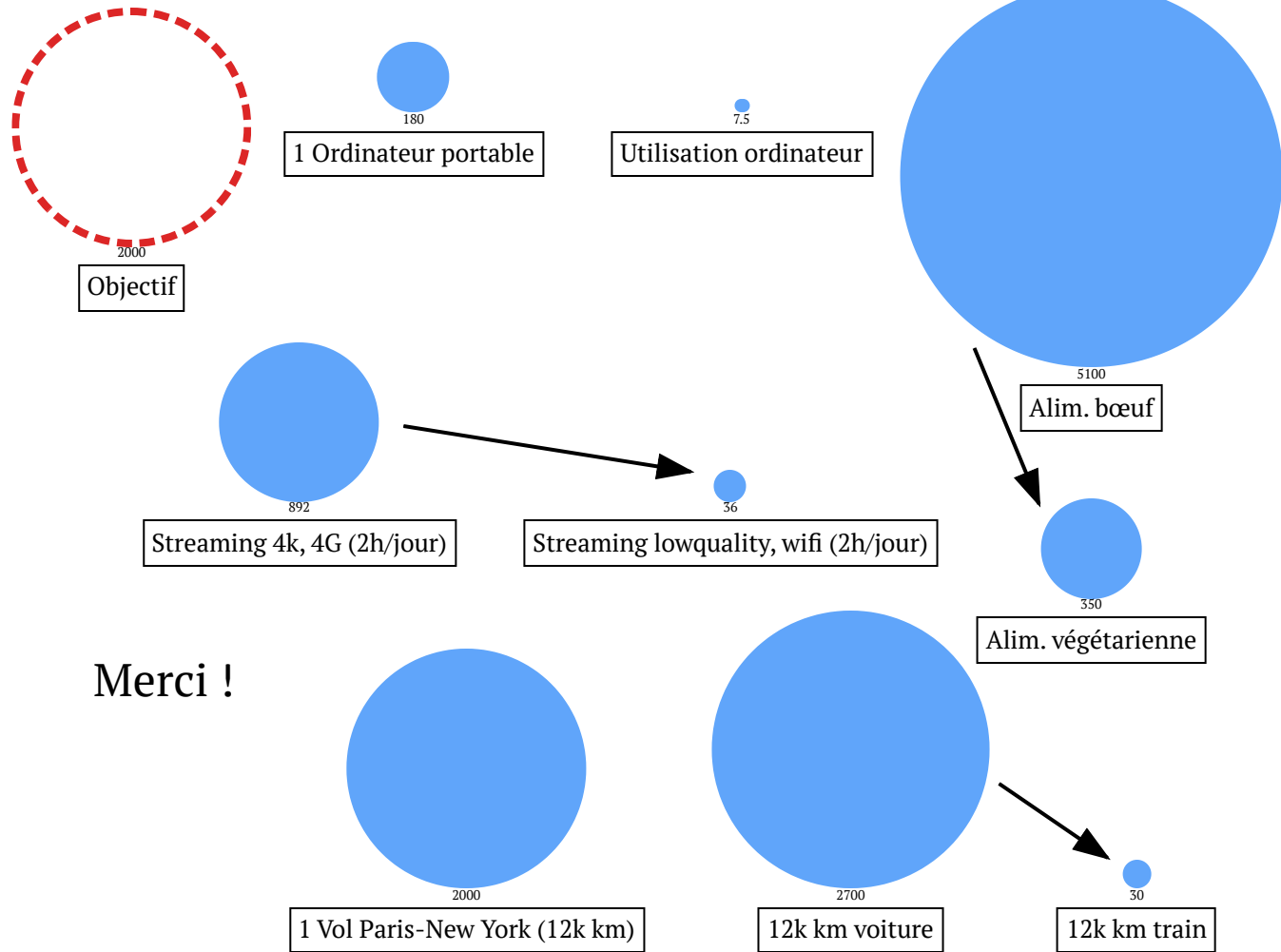
- Avis : pas de retour en arrière majeur
 - IA Générative de partout
 - société très différente
 - productivité ?
- Avancées techniques et raisonabilité des usages
 - modèles frugaux très efficaces (frugalité technique)
 - modèles ouverts et suffisamment bons (frugalité des usages)
- Espoir : inexistence d'un marché de l'IA Générative
 - protection contre les pires aspects négatifs
 - obligation de discussion de choix de société
- Levier possible pour la découverte scientifique
 - qui a lu tout wikipedia ?
 - et tous les articles publiés, dans tous les domaines ?

1. Impact environnemental du numérique ?
... et autres activités
2. Qu'est-ce que l'intelligence artificielle (IA) ?
Faire apprendre une machine à partir d'exemples
3. Coût énergétique de l'IA générative ?
Entraînement et utilisation
4. Réduire l'impact environnemental de l'IA générative ?
Pistes et défis
5. Autres Impacts et Réflexions ?
Élargissement et prédictions aléatoires

Conclusions

Conclusions

(par an et/ou personne)



!!! Augmentation en continue !!!

Merci !