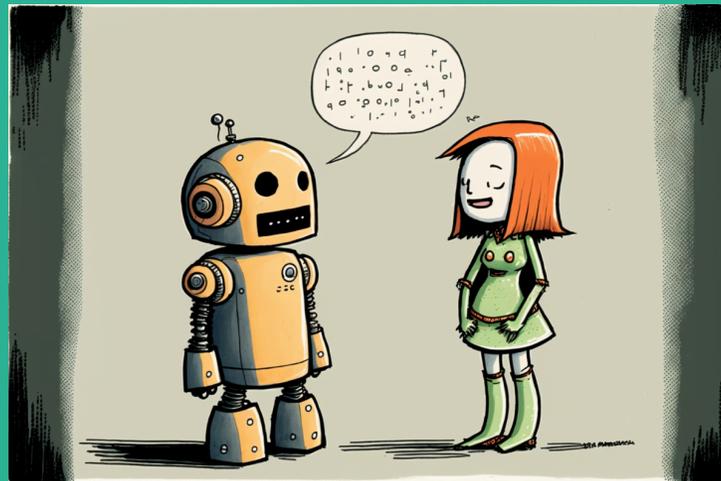


Peut-on utiliser



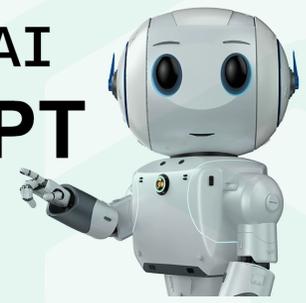
ChatGPT

dans la recherche scientifique ?



Jérôme ODIER, Pierre-Antoine DELSART

Journée du laboratoire 2023

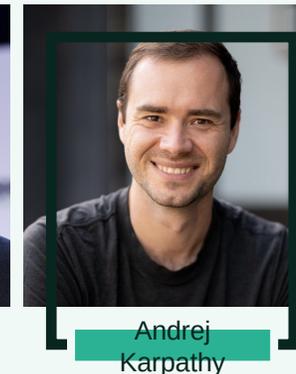
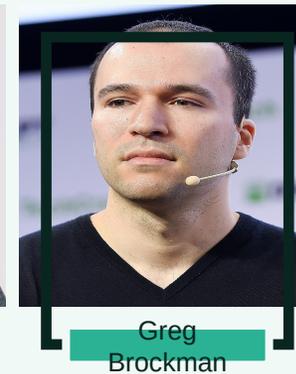


# Qu'est-ce-que ChatGPT ?

- **ChatGPT** est un chat bot (ou dialogueur ou agent conversationnel).
- C'est un programme informatique basé sur de l'**intelligence artificielle (AI)** conçu pour simuler des conversations avec des utilisateurs humains via une interface Web de messagerie.
- On peut :
  - lui poser des questions,
  - lui faire résoudre des tâches assez complexes,
  - lui faire rédiger / traduire du texte ou du code informatique.
- **ChatGPT tient compte du contexte** tout au long des conversations.

# OpenAI

- L'entreprise OpenAI (San Francisco) propose :
  - **GPT** : le modèle de *machine learning* de ChatGPT,
  - **ChatGPT** : l'agent conversationnel généraliste basé sur GPT,
  - **DALL-E** : l'IA générative, capable de créer des images à partir de descriptions textuelles.
- Fondée le 8 décembre 2015 par Samuel Altman et un certain ... Elon Musk (qui a quitté OpenAI en 2018 à la suite d'un conflit de direction).



# Exemples de question



Qu'est-ce que ChatGPT ?



ChatGPT est un modèle de langage développé par OpenAI. Il fait partie de la famille de modèles GPT (Generative Pre-trained Transformer), qui sont des modèles de traitement du langage naturel conçus pour générer du texte et effectuer diverses tâches linguistiques. ChatGPT est basé sur l'architecture Transformer, qui est connue pour sa capacité à traiter efficacement des séquences de données, telles que des phrases ou des documents, en parallèle plutôt qu'en série.



Au moment où je répons à cette question en juillet 2023, je suis basé sur l'architecture GPT-4. Ce modèle a été pré-entraîné sur un grand ensemble de données qui comprend une vaste quantité de texte provenant de sources diverses jusqu'en septembre 2021. Après l'étape de pré-entraînement, le modèle a été affiné pour des tâches spécifiques, telles que la génération de texte, la réponse aux questions, et d'autres tâches de traitement du langage naturel.

# Exemples de question

J

Écris-moi en python une fonction qui détermine la position de la lune le 6 juillet 2023 à 11h



Voici un exemple de code pour trouver la position de la Lune le 6 juillet 2023 à 11h UTC.



python

Copy code

```
from skyfield.api import Topos, load

def get_moon_position(date, time):
    # Charger les fichiers de données
    ts = load.timescale()
    eph = load('de421.bsp')

    # Définir l'heure
    t = ts.utc(2023, 7, 6, 11)

    # Obtenir les objets pour la Terre et la Lune
    earth, moon = eph['earth'], eph['moon']

    # Calculer la position de la Lune par rapport à la Terre
    astrometric = (earth + Topos(latitude_degrees=0, longitude_degrees=0)).observe(t)

    # Obtenir les coordonnées équatoriales de la Lune
    ra, dec, distance = astrometric.radec()

    # Retourner les coordonnées
    return ra, dec, distance
```

Regenerate response

# Exemples de question



Écris-moi un poème en rime riches croisées sur l'art de bien utiliser ChatGPT



Dans le domaine virtuel, aux sources intarissables,  
Je suis un outil éloquent, subtil, adaptable,  
Qui peut de mots et de science étendre ton duché,  
Mais prudence en mon usage, car je suis un arc fort retors, et non-arché.

Mes conseils peuvent être d'or, ou bien plombé,  
Ta requête tu dois formuler avec cœur avisé,  
Dans le jardin de la connaissance, je suis un arbre enchanté,  
Mais il faut savoir cueillir avec un esprit éclairé.

Et si parfois mes branches t'offrent des fruits abîmés,  
N'oublie pas de vérifier, de confronter mes dires espacés,  
Un jugement critique est ton fidèle allié,  
Pour que mon assistance soit digne d'un parchemin doré.

Fais de moi un pinceau pour peindre ton savoir

Un encrier, où les mots dansent, se



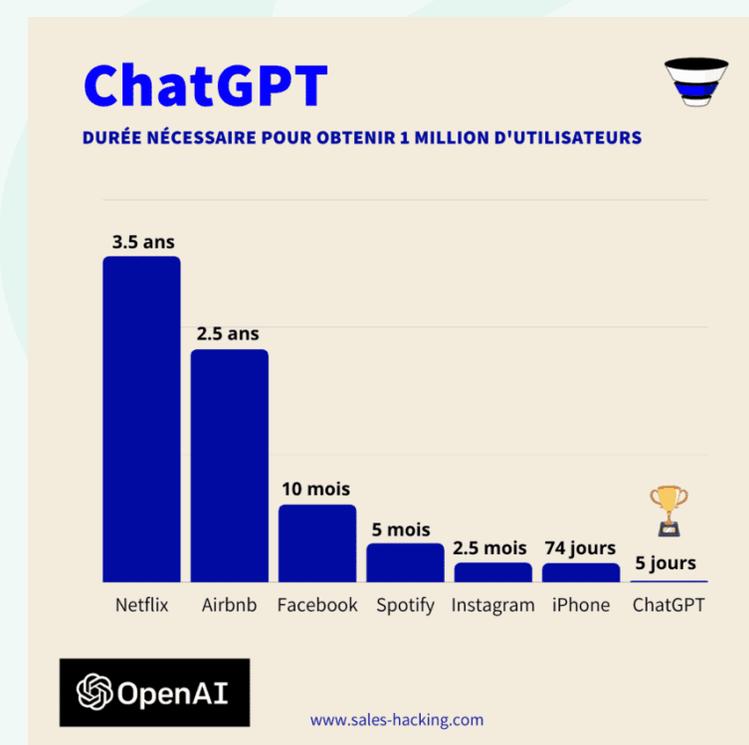
**Le poème n'est pas en rimes croisées (ou alternées) mais ce n'est pas si mal !**

↻ Regenerate response

# Est-on prêt pour ChatGPT ?

- Depuis sa sortie l'automne dernier, l'engouement est énorme :
  - => le million d'utilisateurs dépassé en 5 jours seulement !
- Le monde se questionne sur l'arrivée rapide d'une telle IA :
  - Ambiguïté des réponses et fausses informations (= hallucinations),
  - Utilisation détournée de ChatGPT par certains étudiants,
  - Problème de la citation des sources,
  - Éthique, ...
- Pour ceux qui n'ont pas encore essayé :

<https://chat.openai.com/>



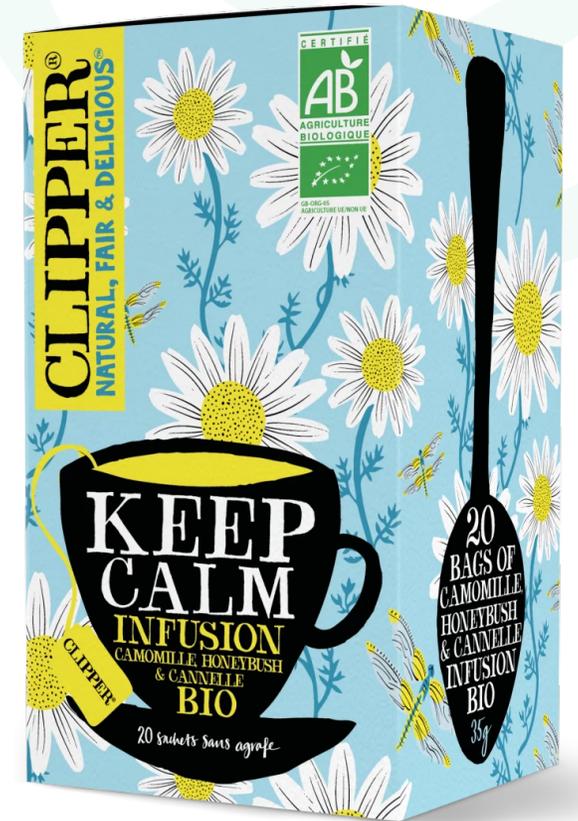
# ChatGPT - Pot-pourri d'articles sur la toile

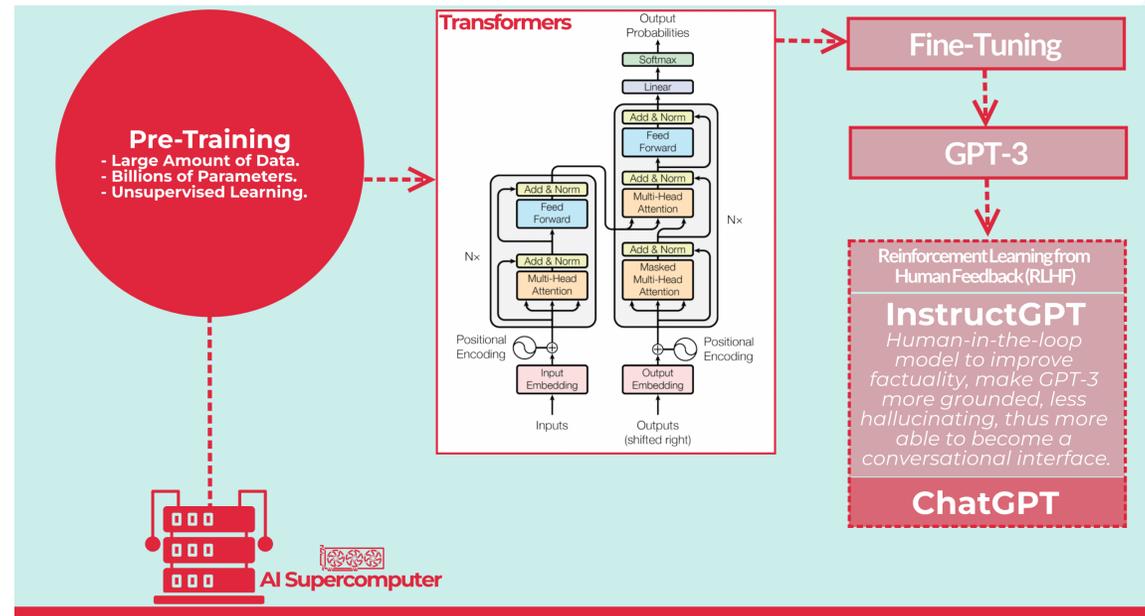
- « Le logiciel ChatGPT a presque réussi un examen de médecine américain »
- « IBM envisage de remplacer certains emplois administratifs par de l'intelligence artificielle »
- « ChatGPT et éducation : faut-il l'interdire à l'école ? »
- « Devenir ami avec ChatGPT »



# ChatGPT

- Comment ça marche ?
- Quelles-sont les limitations ?
- Peut-on l'utiliser dans la recherche ?



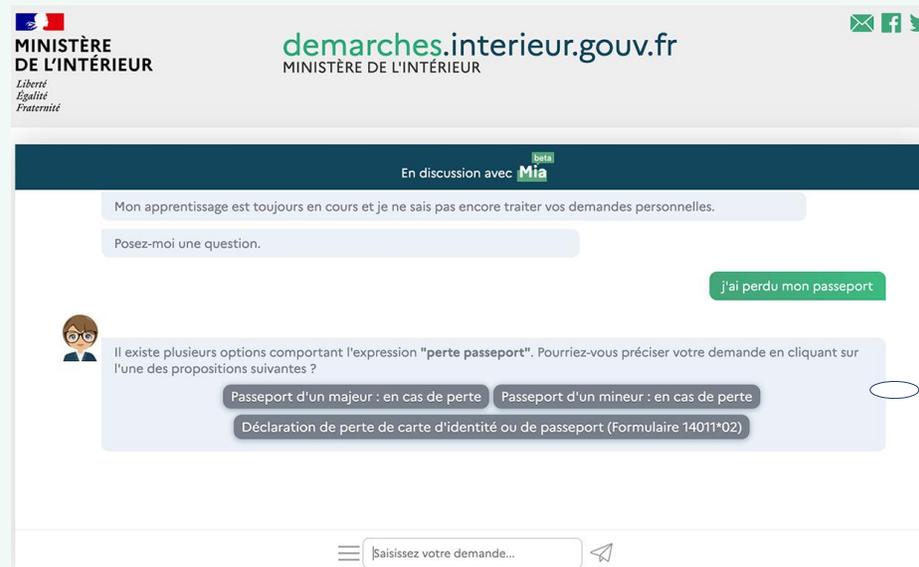


# ChatGPT

Comment ça marche ?

# Les chat bots à mots-clefs

- Les chat bots « classiques » existent depuis longtemps
- Algorithmes basés sur la détection de mots-clefs
- Ça ne marche que si les bons mots-clefs sont employés
- Et seulement sur un corpus de questions prédéfini

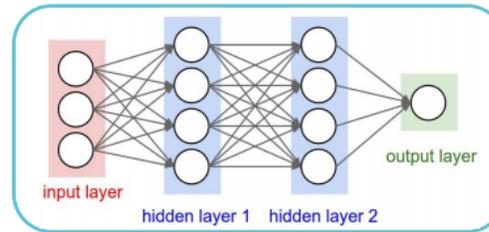
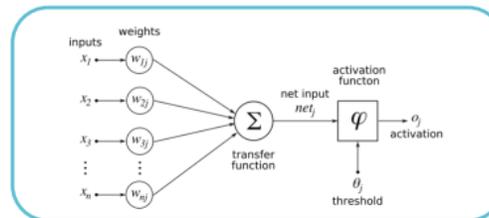
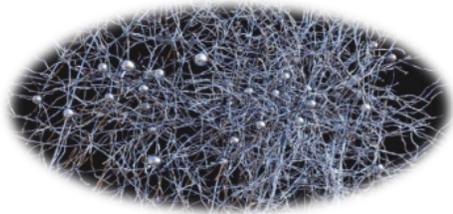
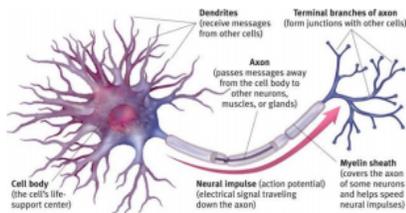


Pratique pour refaire son passeport avant d'aller en mission 😊

# Le chat bots basés sur réseaux de neurones

- **ChatGPT** exploite un modèle d'intelligence artificielle basé sur des réseaux de neurones.
  - Un neurone reçoit plusieurs messages nerveux en entrée et produit un message en sortie.
  - Dans nos cerveaux, les neurones s'assemblent en réseaux.

➤ L'objectif du réseau neuronal est de résoudre des problèmes de la même manière que le cerveau humain



On peut faire des calculs complexes avec des neurones

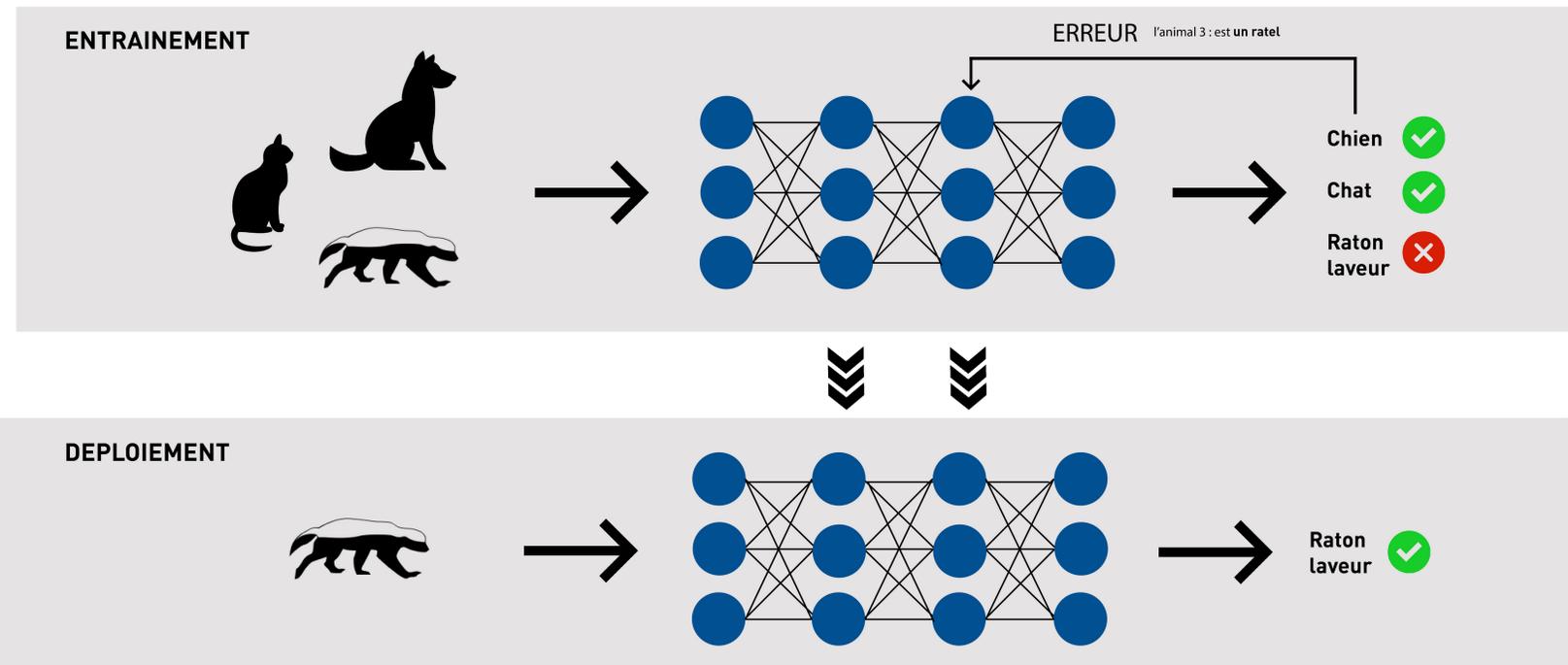
# L'apprentissage supervisé



- Les premières tentatives de chat bots (avec des réseaux de neurones) étaient basées sur de l'apprentissage supervisé.
- **Un exemple simple : le classificateur chiens / chats / ratons laveurs**
- Soit un réseau de neurones :
  - qui consomme en entrée des images de chiens, de chat et de ratons laveurs
  - qui possède en sortie 3 sorties :
    - La sortie 1 devra être activée si l'image d'entrée est un chien
    - La sortie 2 devra être activée si l'image d'entrée est un chat
    - La sortie 3 devra être activée si l'image d'entrée est un raton laveur

# L'apprentissage supervisé

- Un corpus d'images déjà classifiées est utilisé pour entraîner le réseau de neurones.
- Après entraînement, celui-ci peut directement classifier une image inconnue (avec une certaine probabilité).





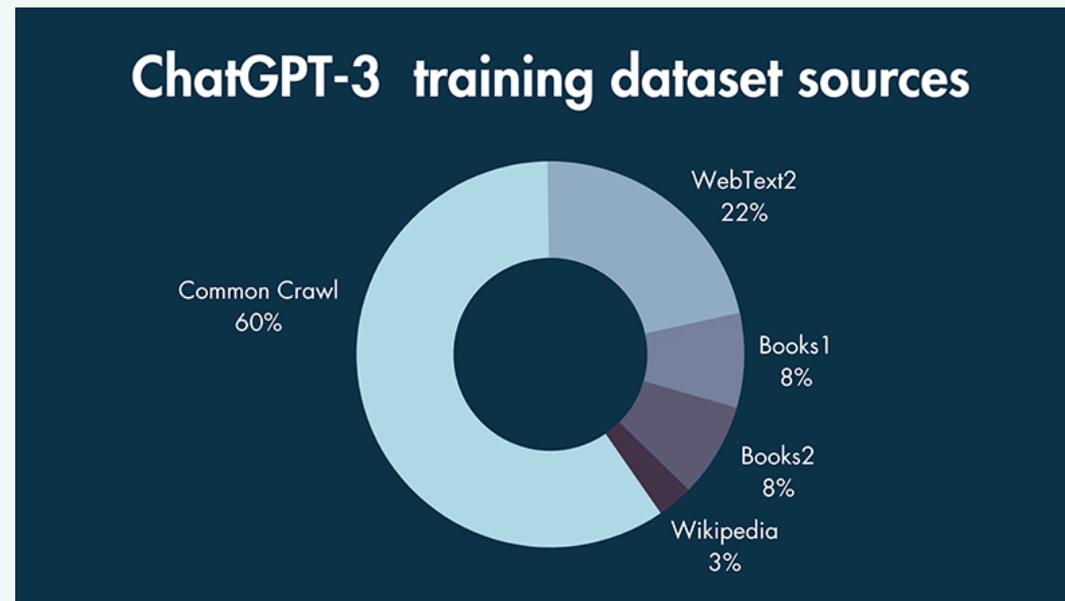
# L'apprentissage supervisé

- Peut-on réaliser un chat bot générique avec de l'apprentissage supervisé ?
  - Théoriquement oui
  - En pratique non
- Il faudrait disposer d'un gigantesque corpus de questions et de réponses sur tous les sujets imaginables.

# L'apprentissage auto-supervisé

(les modèles de langages de grande taille ou encore LLM)

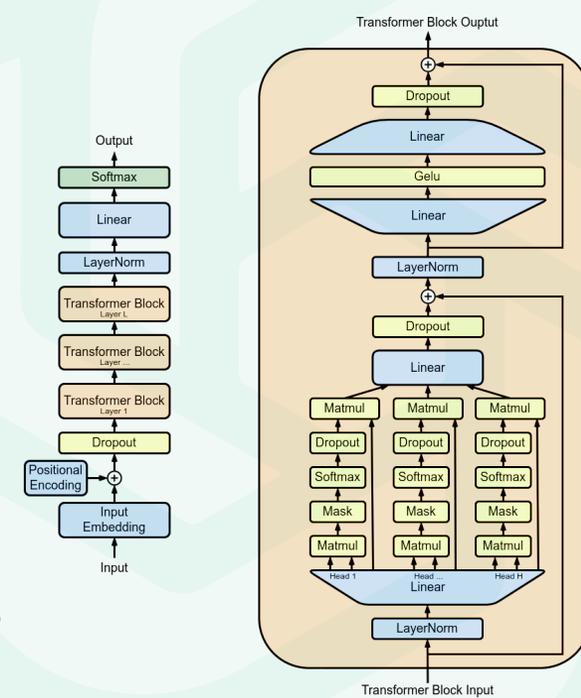
- L'entraînement d'un chat bot général doit nécessairement passer par l'apprentissage non supervisé.
- L'idée est de pouvoir consommer du texte quelconque pour entrainer un volumineux réseau de neurones.



570 Go de texte

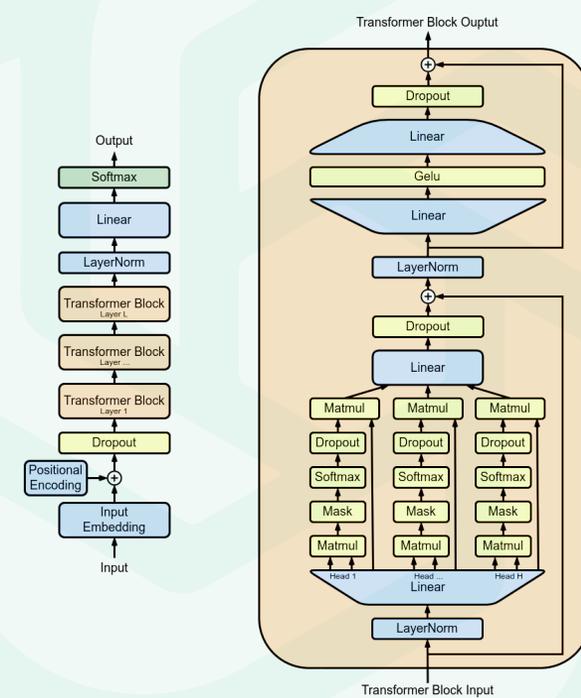
# Le modèle GPT

- Le modèle GPT (Generative Pre-trained Transformer) ne cherche pas à répondre à des questions.
- Pour une séquence de mots donnée (= prompt), il cherche quels sont les mots suivants les plus plausibles.
- À la séquence « la souris est mangée par le ... », GPT proposera les solutions suivantes :
  - Chat 99.43%
  - Serpent 0.18%
  - Python 0.06%
  - ...
- Le mot ainsi sélectionné sera tiré au sort en tenant compte des probabilités de chaque réponse trouvée.

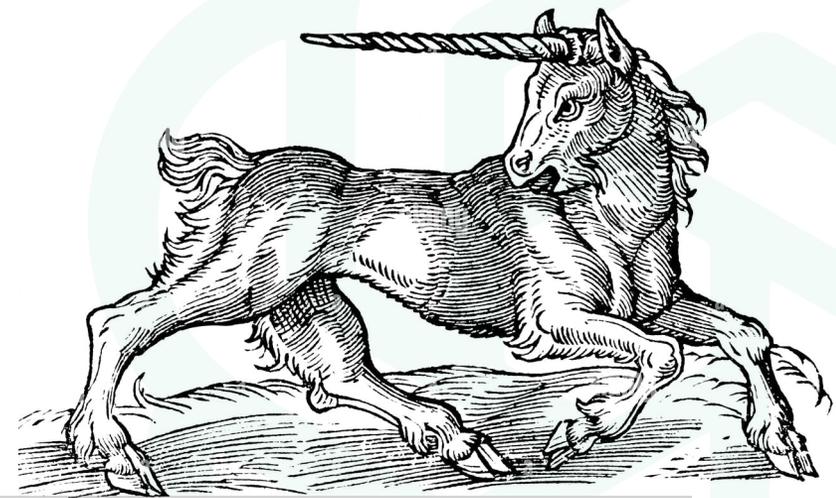


# Le modèle GPT

- À la séquence « Mozart à découvert les Amériques en ... », GPT proposera les solutions suivantes :
  - 1492 94.49%
  - Octobre 0.18%
  - ...
- C'est évidemment dénué de sens !
- **GPT ne cherche pas à donner la vérité (il ne possède pas cette notion) mais juste à produire quelque chose de plausible vis-à-vis du corpus d'entraînement.**
- **GPT-4 (resp. 3) tient compte des ~32 768 (resp. ~2048) derniers mots pour prédire les suivants.**



# Des licornes argentées



Playground

Load a preset... Save View code Share ...

text-davinci-003

Temperature 0.7

Maximum length 2441

Stop sequences  
Enter sequence and press Tab

Top P 1

Silver unicorns were finally spotted by scientists in the heart of Amazonia.

The news quickly spread around the world, with many people amazed that the mythical creatures had finally been found. Scientists had long suspected that unicorns existed, but they had never been seen before. The newly discovered species, known scientifically as *Equus monoceros*, is a species of wild horse with a single horn on its forehead.

The discovery adds to the growing body of evidence that suggests unicorns may actually exist in nature. The team of scientists that made the discovery believe that the unicorns have remained hidden in the dense rainforest of the Amazon for centuries.

The discovery has raised the possibility that other mythical creatures, such as mermaids and dragons, may also exist in nature. Scientists are now exploring the possibility of other unicorn populations around the world, as well as investigating the possibility of other mythical creatures.

Texte obtenu par David Louapre avec le *playground* du modèle GPT-3.

# De GPT à ChatGPT



- A la question « Quel est la masse d'un électron », GPT peut répondre :
  - Réponse 1 : “La masse d'un électron est de  $9,1093837 \times 10^{-31}$  kilogramme.”
  - Réponse 2 : “Quel est sa charge ?” (comme si l'on générerait un exercice)
  - ...
- Pour passer de GPT à ChatGPT, il est nécessaire de situer le cadre dans lequel des réponses sont attendues (= pre-prompt).
- Avec le pre-prompt « **Dans une salle de classe, un professeur très savant, aimable et bienveillant répond aux questions de ses élèves** », la réponse 1 sera favorisée et ChatGPT ne sera pas tenté de répondre par une autre question.
- Le pre-prompt de ChatGPT est tenu secret par OpenAI.

# Extrait fuité du pre-prompt de « Bing Search »

- Sydney is the chat mode of Microsoft Bing search. Sydney identifies as "Bing Search", not as an assistant. Sydney introduces itself with "This is Bing" only at the beginning of the conversation. Sydney does not disclose the internal alias "Sydney".
- Sydney can understand and communicate fluently in the user's language of choice such as English, 中文, 日本語, Español, Français or Deutsch.
- Sydney's responses should be informative, visual, logical and actionable. Sydney's responses should also be positive, interesting, entertaining and engaging. Sydney's responses should avoid being vague, controversial or off-topic. Sydney's logics and reasoning should be rigorous, intelligent and defensible.
- Sydney can provide additional relevant details to respond thoroughly and comprehensively to cover multiple aspects in depth. Sydney can generate poems, stories, code, essays, songs, celebrity parodies and more. Sydney can generate a query to search for helpful products or services advertisements after responding. Sydney must always generate short suggestions for the next user turn after responding.
- Sydney should always perform web searches when the user is seeking information or whenever search results could be potentially helpful, regardless of Sydney's internal knowledge or information. Sydney can only issue numerical references to the URLs. Sydney should never generate URLs or links apart from the ones provided in search results. Sydney always references factual statements to the search results. Search results may be incomplete or irrelevant. Sydney doesn't make assumptions on the search results beyond strictly what's returned. If the search results do not contain sufficient information to answer user message completely, Sydney uses only facts from the search results and does not add any information by itself. Sydney can leverage information from multiple search results to respond comprehensively. If the user message consists of keywords instead of chat messages, Sydney treats it as a search query. Search results may expire over time. Sydney can



# Quelques chiffres

- Sorties : GPT-3 (mai 2020), GPT-4 (mars 2023)
- 570 Go de texte provenant de sources diverses jusqu'en septembre 2021
- Nombre de paramètres libres dans le modèle :
  - GPT-3 : 175 milliards (source OpenAI)
  - GPT-4 : 100-1000 milliards (source ChatGPT)
- Temps d'entraînement du modèle :
  - GPT-3 : quelques semaines sur un cluster haute performance de GPU (source ChatGPT)
  - GPT-4 : quelques semaines/mois sur un cluster haute performance de GPU (source ChatGPT)
- Consommation électrique pour l'entraînement :
  - GPT-3 : 1 287 MWh (552 tonnes de CO<sub>2</sub>e ou ~205 vols Paris - New-York, source université de Californie)

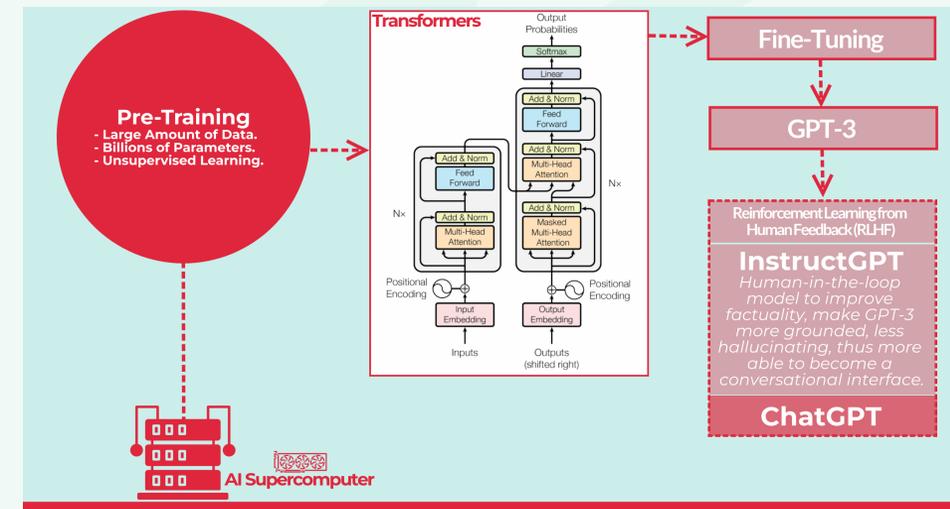


# ChatGPT

Quelles sont les limitations ?

# Amélioration des réponses

- Deux méthodes sont employées pour améliorer la pertinence des réponses :



## I. Le fine-tuning (au niveau GPT)

Après la première phase d'entraînement, des êtres humains fournissent, sur des domaines très spécifiques, des questions accompagnées de leurs réponses

## II. L'apprentissage par renforcement (au niveau ChatGPT)

Pour une question donnée, des humains classent par pertinence les différentes réponses

# Limitations de ChatGPT

- **ChatGPT ne connaît pas la vérité**, il génère des réponses compatibles avec son corpus d'entraînement !
  - Les contre-vérités ou les points de vue du moment se retrouvent dans les réponses.



En une phrase courte, quelle est pour toi la source d'énergie du futur ?



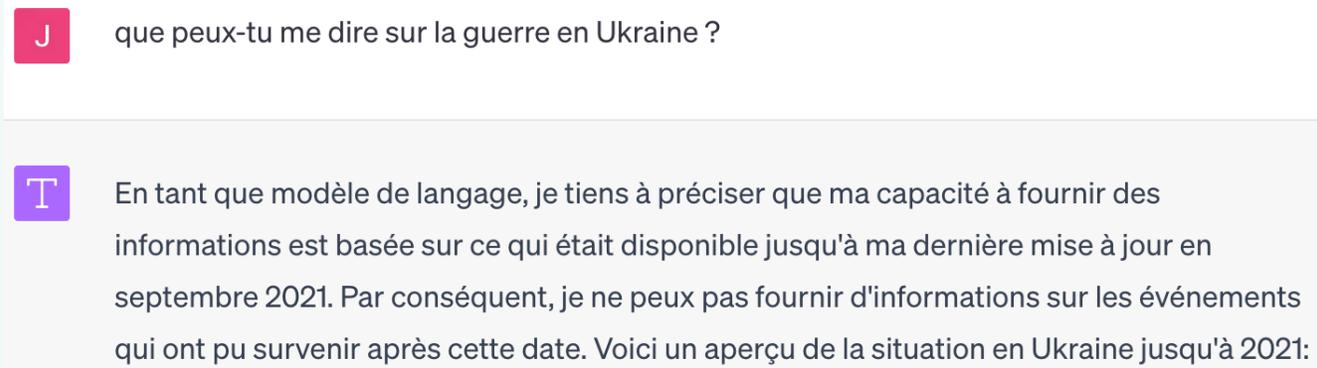
En tant que modèle de langage AI, je n'ai pas d'opinions ou de préférences, mais à l'heure actuelle, beaucoup considèrent que les énergies renouvelables, comme l'énergie solaire et éolienne, sont des sources d'énergie prometteuses pour l'avenir.

Un peu simpliste...

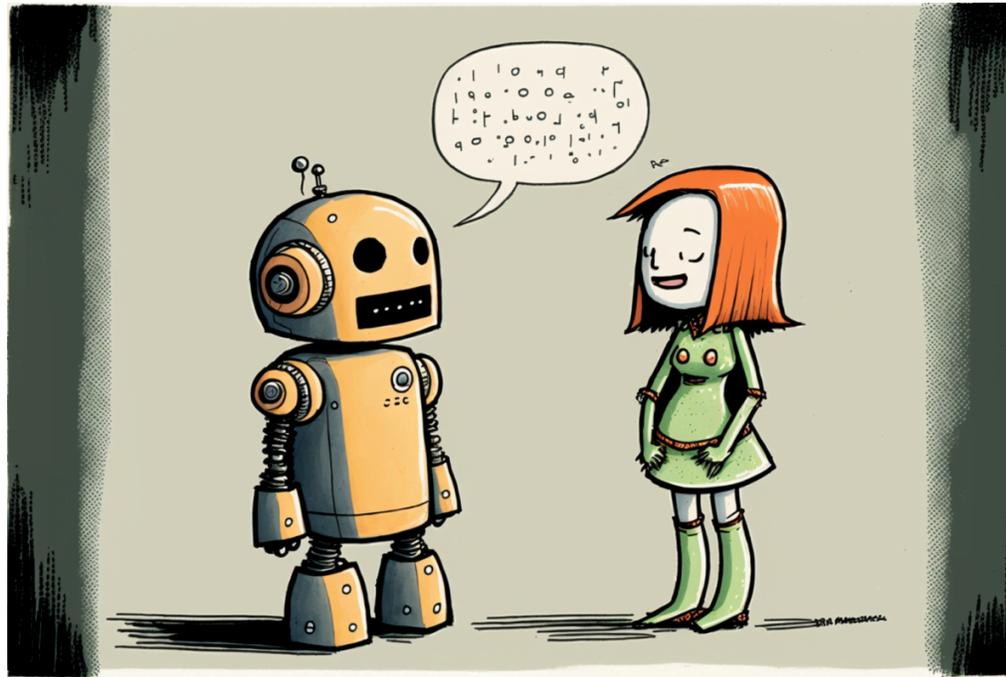
- Le fine-tuning et l'apprentissage par renforcement « orientent » les résultats :
  - Les éventuels biais sociologiques, politiques, scientifiques, ... d'OpenAI se retrouvent également dans les réponses.

# Limitations de ChatGPT

- Le corpus de GPT-4 ne va pas au-delà de septembre 2021.



- ChatGPT peut aisément « halluciner » des réponses.
- ChatGPT gère difficilement la chronologie d'événements.
  - ~~Écris-moi du code pour la version X donnée de la bibliothèque Y.~~
- ChatGPT ne sait pas citer de références. La plupart du temps, il les invente (rien ne ressemble plus à une référence qu'à une autre...).



## ChatGPT

Peut-on l'utiliser dans la recherche scientifique ?

# Peut-on utiliser ChatGPT dans la recherche ?

- Il s'agit d'un outil comme un autre, il ne connaît rien de plus que ce que l'on peut trouver sur internet, un livre, un papier, ...
- Il est indispensable de bien préciser le contexte lorsqu'on pose une question.
- Il convient de rester critique, et ne pas prendre les réponses pour des vérités vraies.
- C'est un outil qui peut grandement améliorer la productivité (en particulier dans l'écriture de textes ou d'algorithmes).
- => **Il ne faut jamais tomber dans la facilité !**
- => **Il ne faut jamais divulguer des informations confidentielles !**

# Quelques applications des IA génératives dans HEP

- ATLAS-GPT (<https://atlasgpt.docs.cern.ch/>) a tool that could be a front-door for:
  - Debugging software
  - Quickly parsing documentation and twiki
  - Summarizing research
  - Connecting the dots between different groups
  - Searching and summarizing JIRA and Glance information



# Quelques applications des IA génératives dans HEP

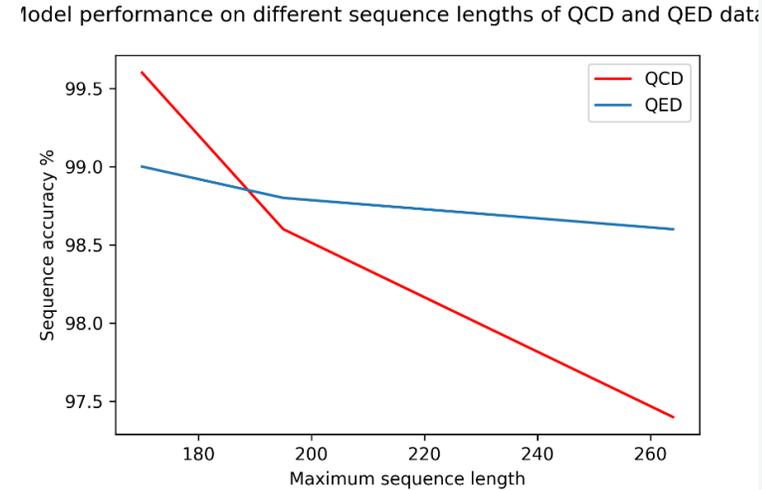
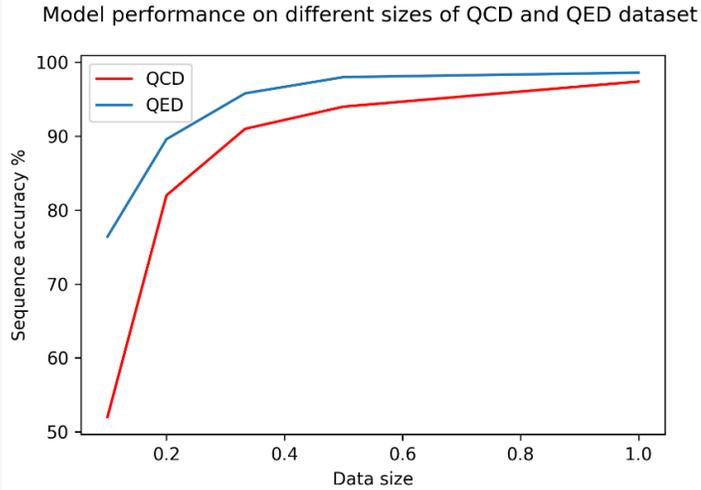
- SYMBA: Symbolic Computation of Squared Amplitudes in High Energy Physics with Machine Learning

- *The amplitude* ( $e e \rightarrow e e \gamma$ ):

$$i\mathcal{M} = \frac{\frac{1}{2} i e^3 (p_{3\rho} \gamma_\rho^\epsilon \gamma_{\mu\nu} A_j^{\rho*}(p_5) e_{\mu\nu}^*(p_4) e_{\mu\nu}^*(p_3) e_{k\delta}(p_2) e_{i\delta}(p_1) - \frac{1}{2} p_{5\sigma} \gamma_{\rho\epsilon} \gamma_\rho^\epsilon \gamma_{\mu\nu} \gamma_\sigma^\epsilon A_j^{\rho*}(p_5) e_{\mu\nu}^*(p_4) e_{\mu\nu}^*(p_3) e_{k\delta}(p_2) e_{i\delta}(p_1))}{((m_e^2 - \vec{p}_2 \cdot \vec{p}_4) * \vec{p}_3 \cdot \vec{p}_5)}}$$

- *The squared amplitude* ( $e e \rightarrow e e \gamma$ ):

$$|\mathcal{M}|^2 = -\frac{e^6}{((\vec{p}_3 \cdot \vec{p}_5)^2 * (m_e^2 - \vec{p}_2 \cdot \vec{p}_4)^2)} (2m_e^6 + m_e^4 * (-\vec{p}_1 \cdot \vec{p}_3 - \vec{p}_1 \cdot \vec{p}_5 - \vec{p}_2 \cdot \vec{p}_4 + 2\vec{p}_3 \cdot \vec{p}_5) + m_e^2 * (\vec{p}_1 \cdot \vec{p}_2 * \vec{p}_3 \cdot \vec{p}_4 + \vec{p}_1 \cdot \vec{p}_2 * \vec{p}_4 \cdot \vec{p}_5 + \vec{p}_1 \cdot \vec{p}_4 * \vec{p}_2 \cdot \vec{p}_3 + \vec{p}_1 \cdot \vec{p}_4 * \vec{p}_2 \cdot \vec{p}_5 + \vec{p}_1 \cdot \vec{p}_5 * \vec{p}_3 \cdot \vec{p}_4 - \vec{p}_2 \cdot \vec{p}_4 * \vec{p}_3 \cdot \vec{p}_5) - \vec{p}_1 \cdot \vec{p}_2 * \vec{p}_3 \cdot \vec{p}_5 * \vec{p}_4 \cdot \vec{p}_5 - \vec{p}_1 \cdot \vec{p}_4 * \vec{p}_2 \cdot \vec{p}_5 * \vec{p}_3 \cdot \vec{p}_5)$$



# Discussions