

ChatGPT : un outil qui change la donne dans le domaine de l'enseignement (notamment de l'informatique) ?

Journée NSI-SNT, Nancy, 5 avril 2023

Yannick Parmentier



Introduction - à propos de cette présentation

- Contexte
- Objectifs
- Limites



ChatGPT dans les grandes lignes

Ce que les personnes qui l'ont conçu en disent ...

ChatGPT

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer follow-up questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

Ce que les personnes qui l'ont conçu en disent ...

ChatGPT

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer follow-up questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

→ un *agent conversationnel* utilisant un *modèle [de langue] pré-entraîné* ayant pour but de répondre aux questions qu'on lui soumet (aussi connu sous le nom d'*intelligence artificielle*).

Ce que ChatGPT lui-même en dit ...

M

Tell me how ChatGPT works.



ChatGPT is a large language model that uses deep learning techniques to generate human-like text. It is based on the GPT (Generative Pre-trained Transformer) architecture, which uses a transformer neural network to process and generate text. The model is pre-trained on a massive dataset of text, such as books, articles, and websites, so it can understand the patterns and structure of natural language. When given a prompt or a starting point, the model uses this pre-trained knowledge to generate text that continues the given input in a coherent and natural way.



Ce que ChatGPT lui-même en dit ...

M

Tell me how ChatGPT works.



ChatGPT is a large language model that uses deep learning techniques to generate human-like text. It is based on the GPT (Generative Pre-trained Transformer) architecture, which uses a transformer neural network to process and generate text. The model is pre-trained on a massive dataset of text, such as books, articles, and websites, so it can understand the patterns and structure of natural language. When given a prompt or a starting point, the model uses this pre-trained knowledge to generate text that continues the given input in a coherent and natural way.

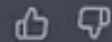


Ce que ChatGPT lui-même en dit ...

M Tell me how ChatGPT works.



ChatGPT is a large language model that uses deep learning techniques to generate human-like text. It is based on the GPT (Generative Pre-trained Transformer) architecture, which uses a transformer neural network to process and generate text. The model is pre-trained on a massive dataset of text, such as books, articles, and websites, so it can understand the patterns and structure of natural language. When given a prompt or a starting point, the model uses this pre-trained knowledge to generate text that continues the given input in a coherent and natural way.



Ce que ChatGPT lui-même en dit ...

M Tell me how ChatGPT works.



ChatGPT is a large language model that uses deep learning techniques to generate human-like text. It is based on the GPT (Generative Pre-trained Transformer) architecture, which uses a transformer neural network to process and generate text. The model is pre-trained on a massive dataset of text, such as books, articles, and websites, so it can understand the patterns and structure of natural language. When given a prompt or a starting point, the model uses this pre-trained knowledge to generate text that continues the given input in a coherent and natural way.

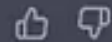


Ce que ChatGPT lui-même en dit ...

M Tell me how ChatGPT works.

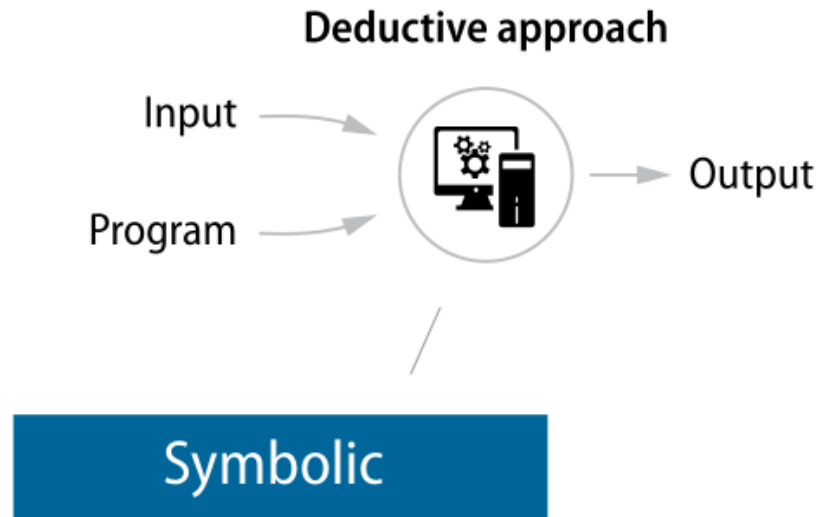


ChatGPT is a large language model that uses deep learning techniques to generate human-like text. It is based on the GPT (Generative Pre-trained Transformer) architecture, which uses a transformer neural network to process and generate text. The model is pre-trained on a massive dataset of text, such as books, articles, and websites, so it can understand the patterns and structure of natural language. When given a prompt or a starting point, the model uses this pre-trained knowledge to generate text that continues the given input in a coherent and natural way.

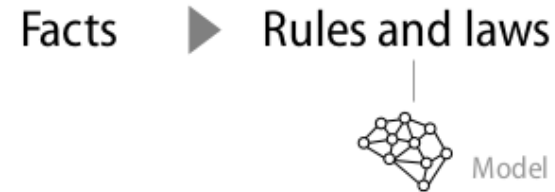
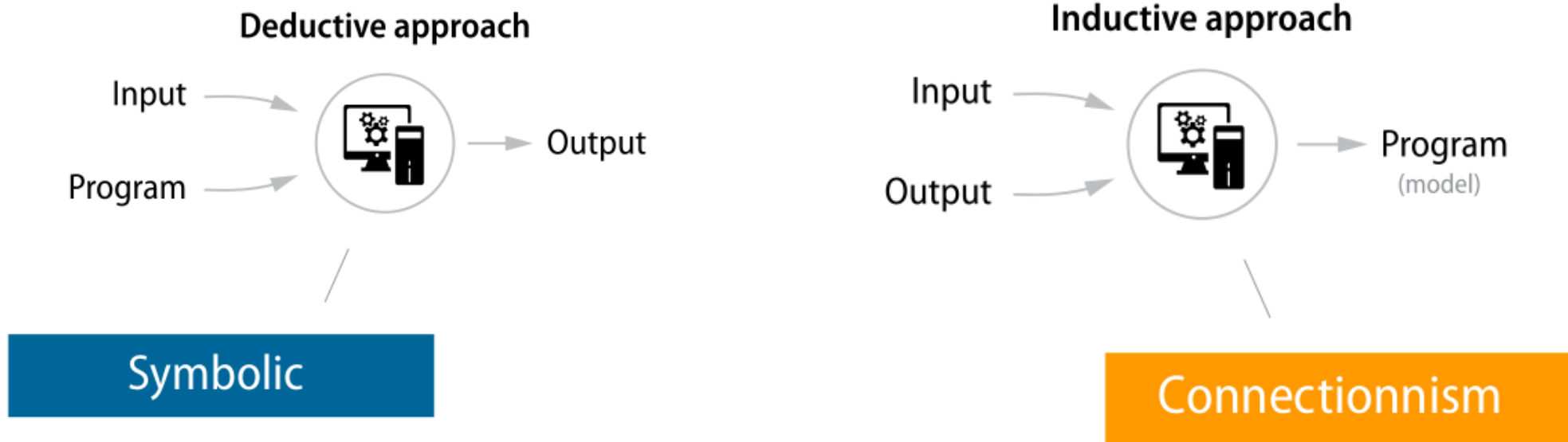


Les réseaux de neurones : apprendre à partir des données (approche inductive)

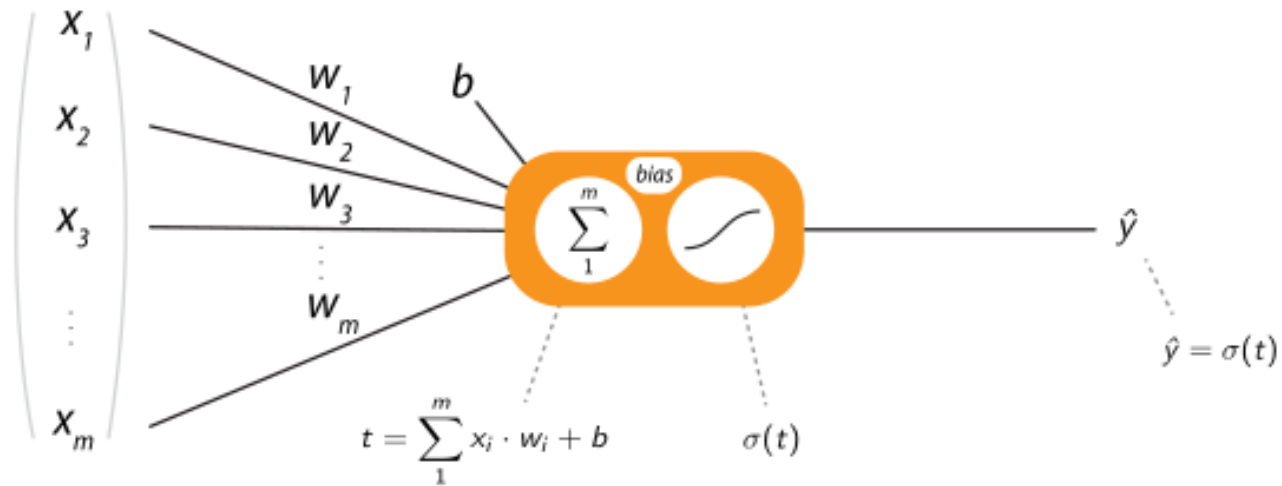
Les réseaux de neurones : apprendre à partir des données (approche inductive)



Les réseaux de neurones : apprendre à partir des données (approche inductive)

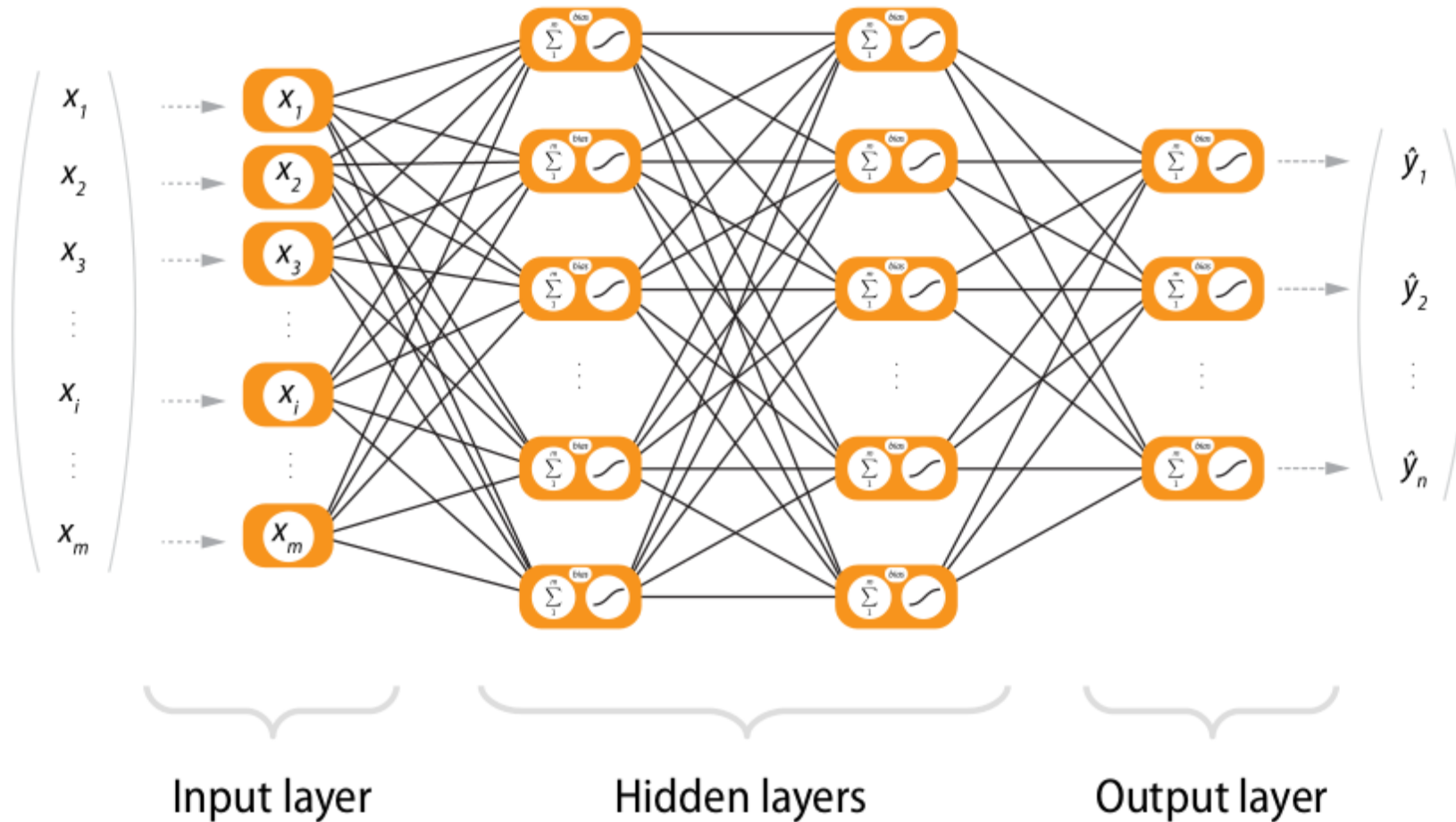


Le neurone artificiel [McCulloch and Pitts, 1943]

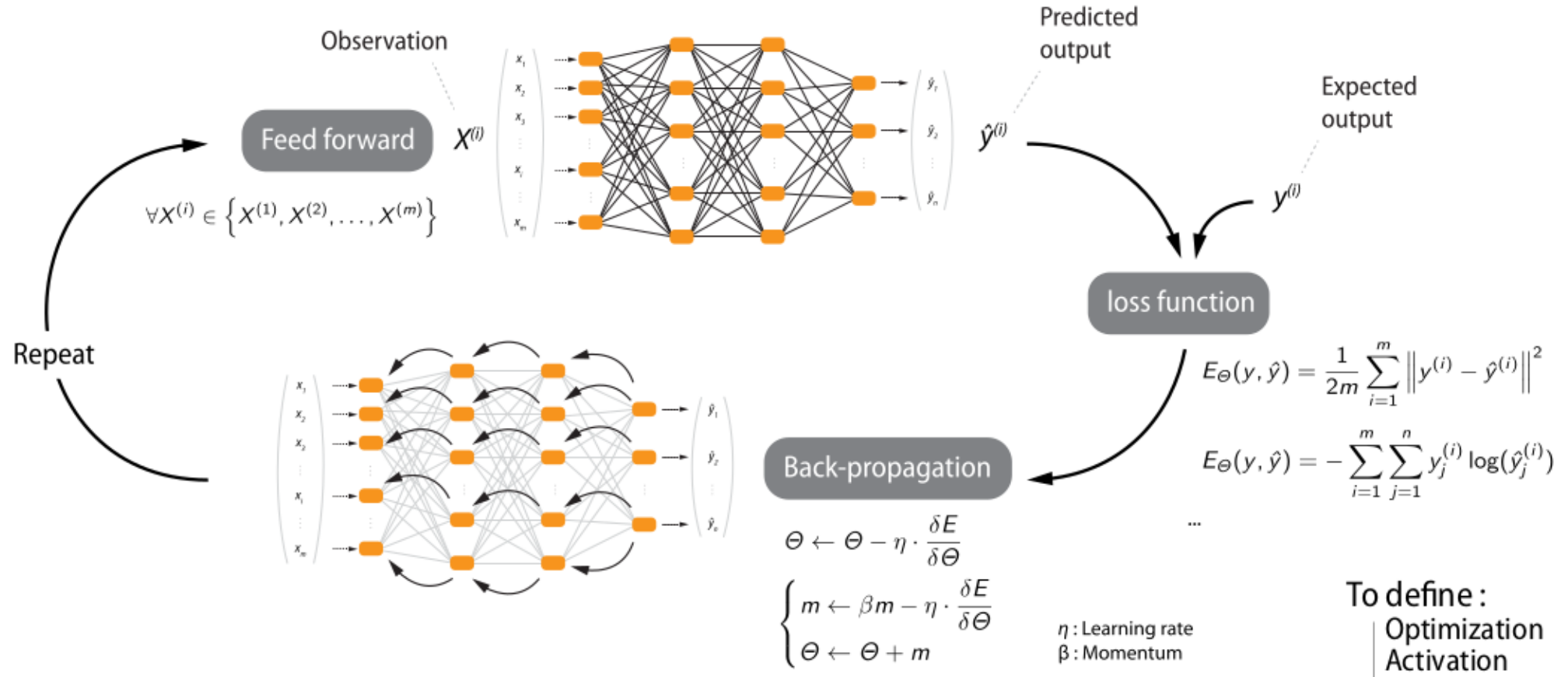


- $[x_1, \dots, x_m]$: vecteur d'entrée
- $[w_1, \dots, w_m]$: vecteur de poids
- b : biais
- $\sigma(t)$: fonction d'activation
- \hat{y} : prédiction

Extension aux réseaux de neurones



Entraîner des réseaux de neurones



Ce que les chercheuses et chercheurs en disent ... (source : [équipe Inria Flowers](#))



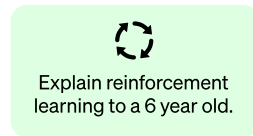
(Lien youtube: <https://www.youtube.com/watch?v=K8gOvC8gvB4>)

ChatGPT: un apprentissage particulier (source: openai.com/blog/chatgpt)

Step 1

Collect demonstration data and train a supervised policy.

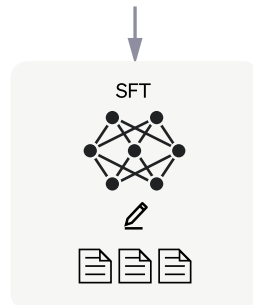
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



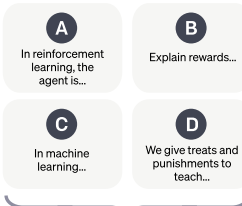
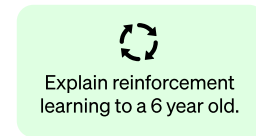
This data is used to fine-tune GPT-3.5 with supervised learning.



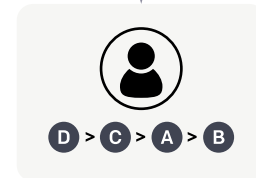
Step 2

Collect comparison data and train a reward model.

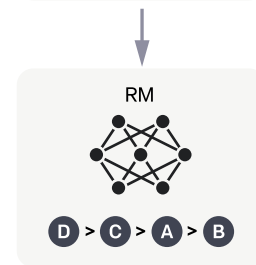
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



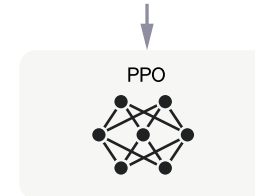
Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

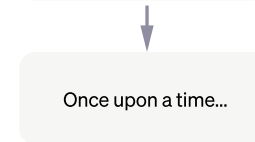
A new prompt is sampled from the dataset.



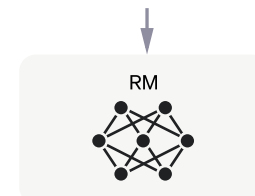
The PPO model is initialized from the supervised policy.



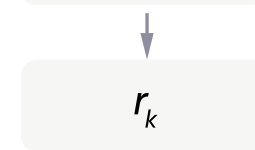
The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.



À retenir

- un modèle de langage génératif pré-entraîné en utilisant des techniques d'apprentissage par renforcement au moyen de feedback humain
- reposant sur un code source et des données d'entraînement fermées
- nécessitant des moyens de stockage, de calcul et d'annotation conséquents

À retenir

- un modèle de langage génératif pré-entraîné en utilisant des techniques d'apprentissage par renforcement au moyen de feedback humain
- reposant sur un code source et des données d'entraînement fermées
- nécessitant des moyens de stockage, de calcul et d'annotation conséquents

→ ce qui soulève de nombreuses questions éthiques (droits d'auteur, impact environnemental, désinformation, etc.)

À retenir

- un modèle de langage génératif pré-entraîné en utilisant des techniques d'apprentissage par renforcement au moyen de feedback humain
- reposant sur un code source et des données d'entraînement fermées
- nécessitant des moyens de stockage, de calcul et d'annotation conséquents

→ ce qui soulève de nombreuses questions éthiques (droits d'auteur, impact environnemental, désinformation, etc.)

→ ce type d'IA (large modèle pré-entraîné) n'est pas unique (GPT, BLOOM, LLaMA, etc.)

Utilisation de ChatGPT

Une interface intuitive : le prompt


- Rappel : le modèle de langue prédit la séquence de mots la plus probable dans un contexte donné



The best thing about AI is its ability to

learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%


- Début de séquence : le prompt (taille limitée)
- peut prendre différentes formes : question en langue naturelle, exemple (d'association), instruction (tâche à effectuer) ...



Exemple #1 : instruction (jeu)

 north

  Copy code

```
You are in a large, gloomy  
cavern.  
Exits: east west
```

 {simulate a troll attack}

  Copy code

```
A troll enters the room!  
  
Troll attacks you with a  
club!  
You take 10 damage.
```


Exemple #2 : question



can you translate this to Japanese "i need coffee when I wake up"



Sure, in Japanese, the sentence "I need coffee when I wake up" can be translated as "私は目が覚めたときにコーヒーが必要です" (Watashi wa me ga sameta toki ni kōhī ga hitsuyō desu).



Exemple #3 : code

EM

Comment résoudre le problème des tours de Hanoi (code) ?



Voici un exemple de code en Python pour résoudre le problème des tours de Hanoi :



Copy code

```
def hanoi(n, source, target, auxiliary):
    if n > 0:
        # déplacer les n-1 disques de la source vers l'auxiliaire
        hanoi(n-1, source, auxiliary, target)
        # déplacer le n-ième disque de la source vers la cible
        print(f"déplacer le disque {n} de {source} vers {target}")
        # déplacer les n-1 disques de l'auxiliaire vers la cible
        hanoi(n-1, auxiliary, target, source)

# résoudre pour 3 disques
hanoi(3, 'A', 'C', 'B')
```

Limites de ChatGPT

Limites de ChatGPT

Limites de ChatGPT

- durée d'entraînement (fréquence des mises à jour)

Limites de ChatGPT

- durée d'entraînement (fréquence des mises à jour)
- sensibilité au prompt

Limites de ChatGPT

- durée d'entraînement (fréquence des mises à jour)
- sensibilité au prompt
- biais induits par les données d'entrées (e.g. annotations humaines)

Limites de ChatGPT

- durée d'entraînement (fréquence des mises à jour)
- sensibilité au prompt
- biais induits par les données d'entrées (e.g. annotations humaines)
- coûts de développement et d'utilisation



Limites de ChatGPT




- durée d'entraînement (fréquence des mises à jour)
- sensibilité au prompt
- biais induits par les données d'entrées (e.g. annotations humaines)
- coûts de développement et d'utilisation
- inexplicabilité

Limites de ChatGPT

- durée d'entraînement (fréquence des mises à jour)
- sensibilité au prompt
- biais induits par les données d'entrées (e.g. annotations humaines)
- coûts de développement et d'utilisation
- inexplicabilité
- **fiabilité** des prédictions (cf hallucinations, raisonnement limité, biais, etc.)

Exemple d'hallucination

 How many times has Argentina won the Fifa world cup 

 Argentina has won the FIFA World Cup once. They won the tournament in 1986, when they defeated West Germany in the final. This win was led by the famous soccer player Diego Maradona.  

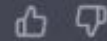
Exemple de raisonnement erroné



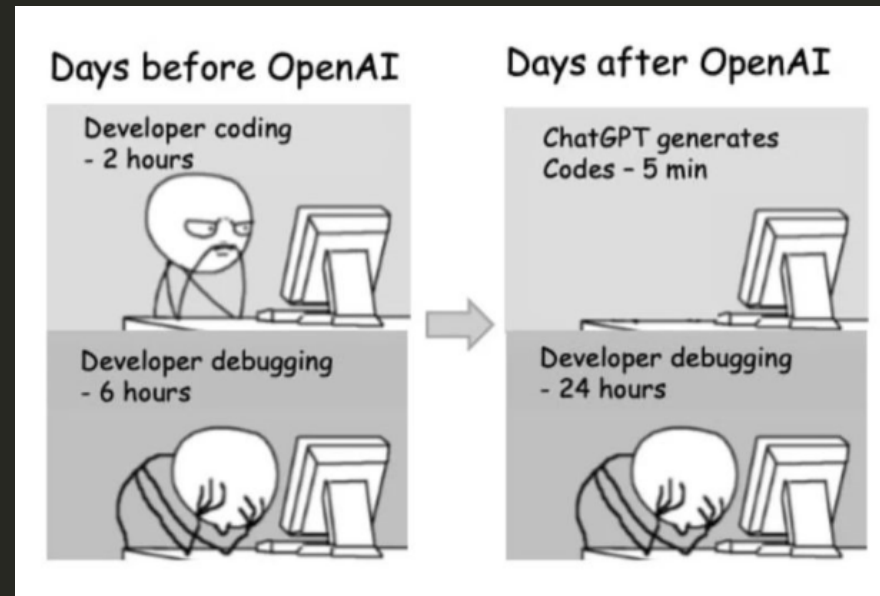
In "John told Jim that he is going to the party and asked him if he is going too." what does the second "he" refer to?



In this context, the second "he" is referring to John, who is the one going to the party according to the first part of the sentence. So, the full sentence can be rephrased as "John told Jim that he (John) is going to the party and asked Jim if John is going too."



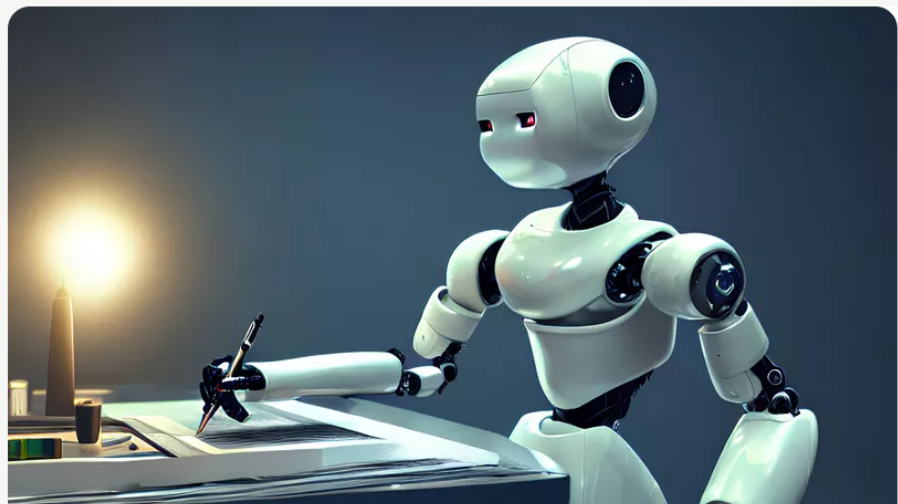
Et l'éducation dans tout ça ?



Source : aprogrammerlife.com

Des étudiants contraints de repasser un examen après avoir utilisé ChatGPT

Par AFP agence • Publié le 03/02/2023 à 18:50 • Mis à jour le 06/02/2023 à 09:45



LE FIGARO | OFFRE JEUNES



À Strasbourg, le groupe d'étudiants qui avait profité d'un QCM à distance pour utiliser ChatGPT a finalement été contraint de passer un rattrapage.

Une vingtaine d'étudiants de l'Université de Strasbourg (Unistra), qui avaient triché à l'aide du chatbot d'intelligence artificielle [ChatGPT](#) lors d'un examen en distanciel, ont dû le repasser en présentiel, a appris l'AFP vendredi auprès de l'établissement. « C'est le premier cas à l'Université de Strasbourg et nous n'en avons pas d'autres à ce jour », a indiqué l'Unistra dans un communiqué transmis à l'AFP.

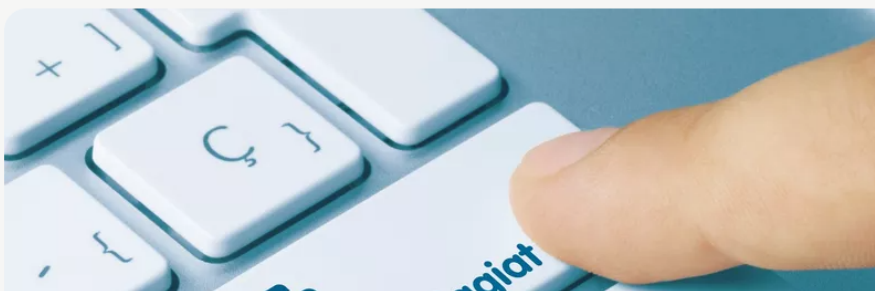
Le Parisien Étudiant, Exams

ChatGPT : « Il ne faut pas faire semblant d'être dans un monde qui ne travaille qu'avec son cerveau »

Les utilisations de ChatGPT en cours sont nombreuses : rédaction de devoir, de fiches de synthèse ou encore écriture de code, le territoire de jeu est vaste. Jusqu'aux exercices de TD.

Un étudiant crée un logiciel pour contrer l'utilisation de ChatGPT

Par Victor Mérat • Publié le 10/01/2023 à 12:23 • Mis à jour le 10/01/2023 à 17:55



ChatGPT vue par ...

ChatGPT vue par ...

- les élèves : une ressource
 - inépuisable
 - interrogeable rapidement et facilement

ChatGPT vue par ...

- les élèves : une ressource
 - inépuisable
 - interrogeable rapidement et facilement
- les personnels enseignants : un outil
 - qui dispense d'efforts
 - incertain
 - difficilement décelable

ChatGPT vue par ...

- les élèves : une ressource
 - inépuisable
 - interrogeable rapidement et facilement
- les personnels enseignants : un outil
 - qui dispense d'efforts
 - incertain
 - difficilement décelable

→ un outil qui questionne les pratiques enseignantes
(notamment le lien entre performance et compétence)

ChatGPT vue par ...

- les élèves : une ressource
 - inépuisable
 - interrogeable rapidement et facilement
- les personnels enseignants : un outil
 - qui dispense d'efforts
 - incertain
 - difficilement décelable

→ un outil qui questionne les pratiques enseignantes
(notamment le lien entre performance et compétence)

→ un outil qui remet en lumière le besoin de
développer l'esprit critique des élèves

Des pistes de remédiation / contrôle

Des pistes de remédiation / contrôle

- Organisation de l'évaluation :
 - devoir à la maison vs devoir sur table
 - ordinateur connecté à internet ou non
 - d'autres modes d'évaluation (par ex. examen oral)
 - grille d'évaluation fine

Des pistes de remédiation / contrôle

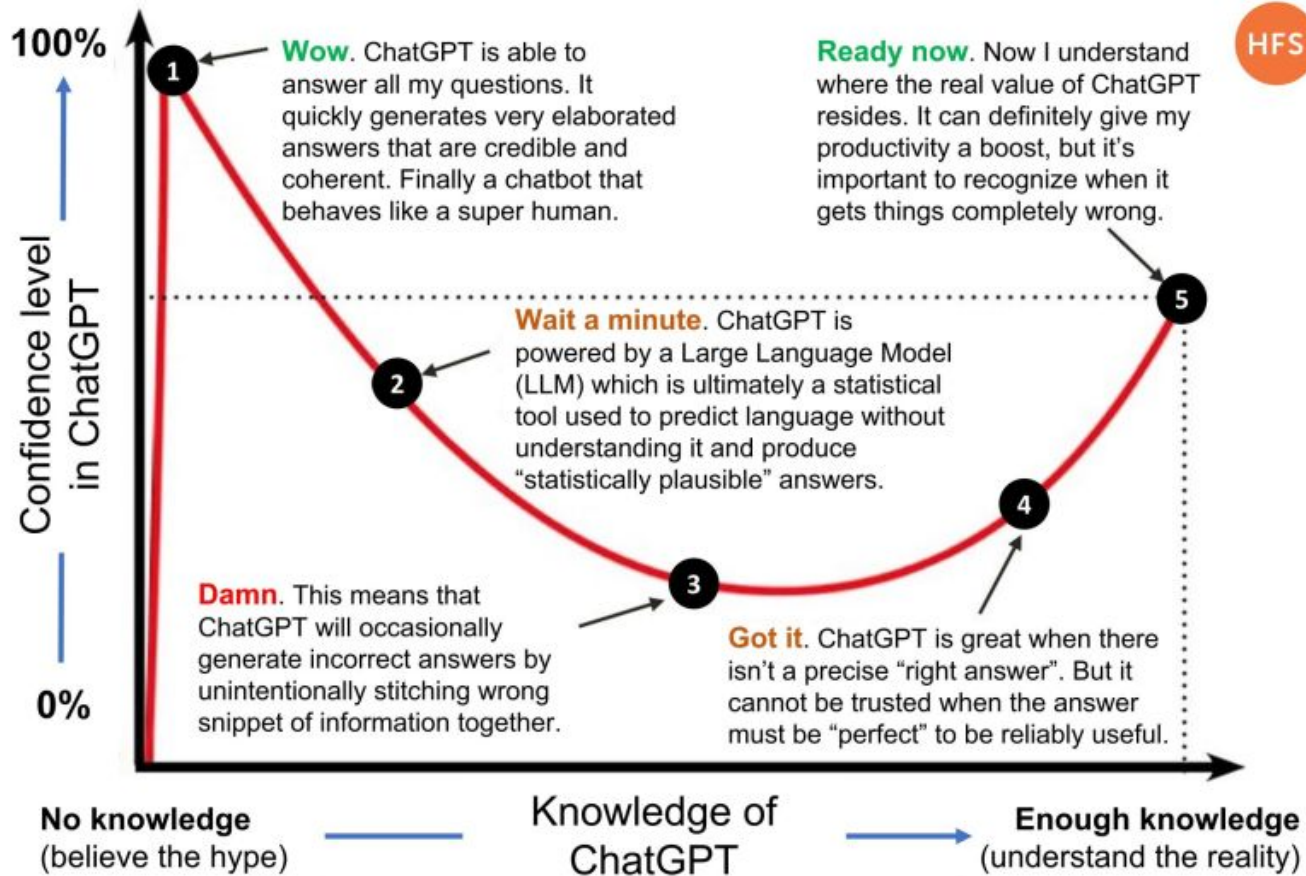
- Organisation de l'évaluation :
 - devoir à la maison vs devoir sur table
 - ordinateur connecté à internet ou non
 - d'autres modes d'évaluation (par ex. examen oral)
 - grille d'évaluation fine
- Contenu des évaluations :
 - questions "originales"
(bon sens, énoncé redondant, etc.)

Des pistes de remédiation / contrôle

- Organisation de l'évaluation :
 - devoir à la maison vs devoir sur table
 - ordinateur connecté à internet ou non
 - d'autres modes d'évaluation (par ex. examen oral)
 - grille d'évaluation fine
- Contenu des évaluations :
 - questions "originales"
(bon sens, énoncé redondant, etc.)
- Tout au long de l'enseignement:
 - accompagner les usages de ChatGPT

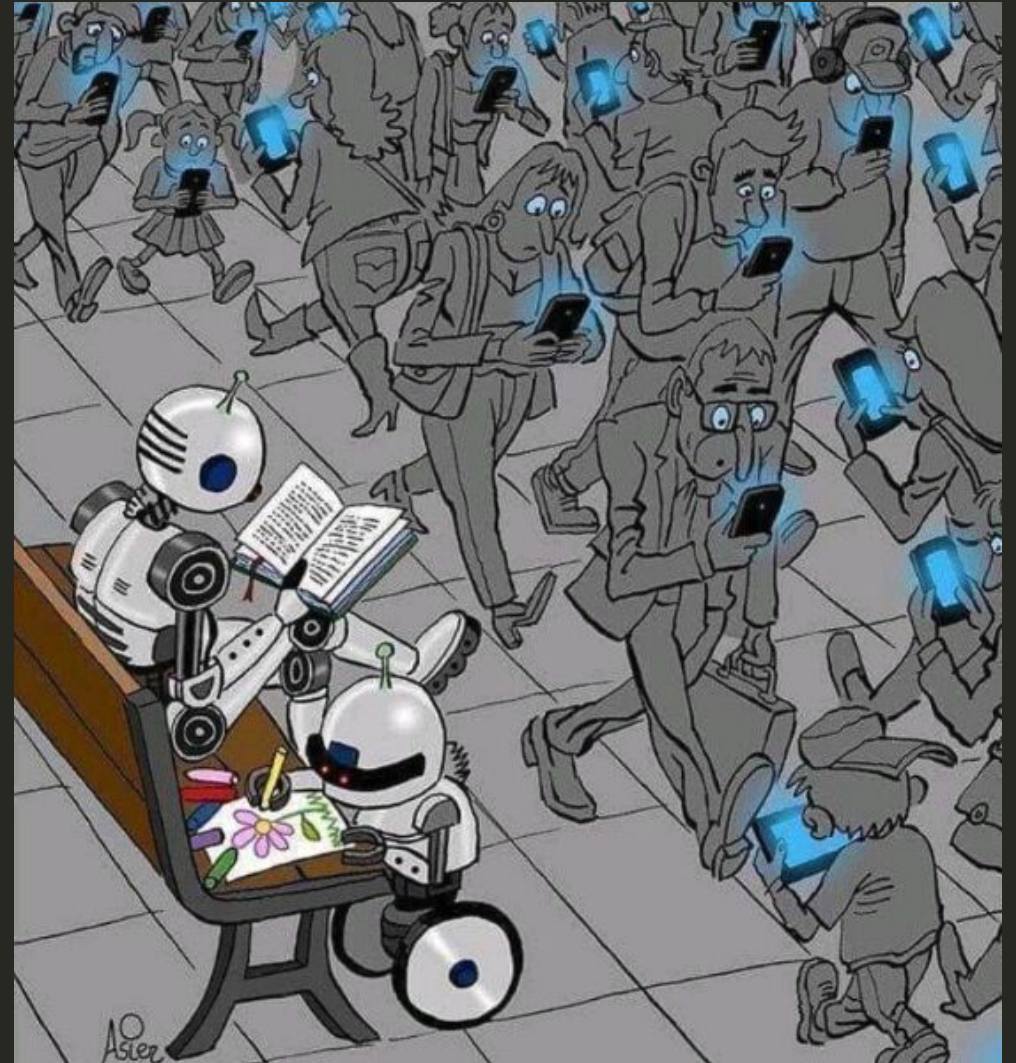
Pour conclure : un outil qui change la donne mais ...

un outil imparfait



Source : HFS research

Merci pour votre attention



Source: "Inteligencia artificial" par Asier