

IA générative, société et Éducation : en quoi l'IA générative représente-t-elle un enjeu dans la formation des citoyens ?

Pierre-Yves Oudeyer

Centre Inria de l'université de Bordeaux

Plan de la présentation

1. Histoire de l'IA
2. Usages des IA génératives
3. IA génératives et éducation

Une très courte histoire de l'IA



1912 : *Electric dog*



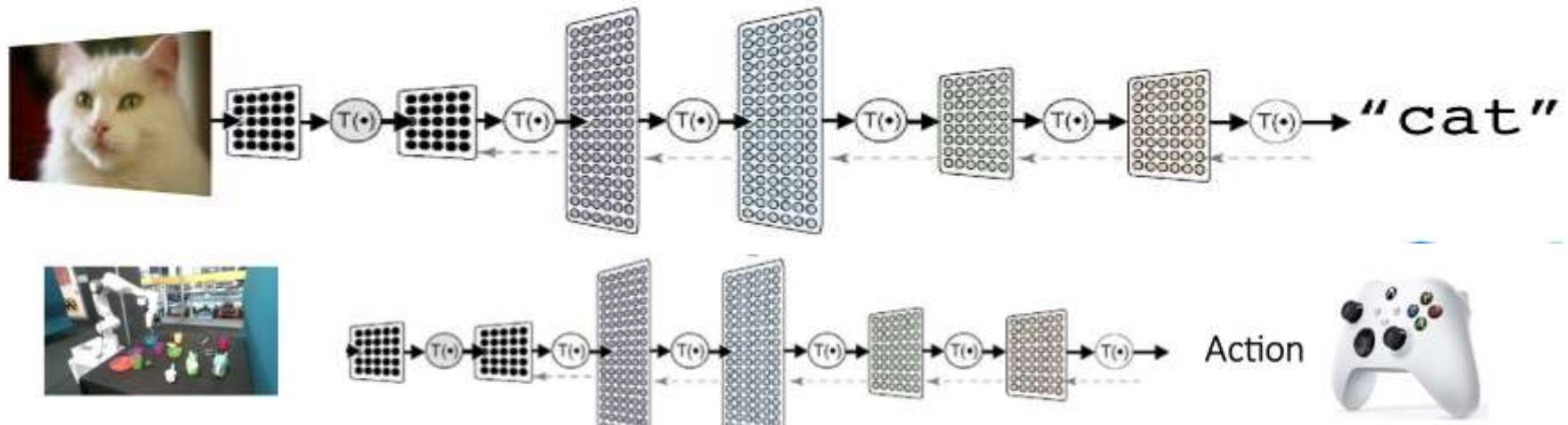
1956 : conf. de Darmouth



IA comportementale



IA symbolique



Apprentissage profond
« *Deep Learning* »

IA générative (IA Gen) : textes, images et vidéos

ENTREPRISE → **IAGEN**

OpenAI → ChatGPT

Google → Gemini

Midjourney → Midjourney

stability.ai → Stable Diffusion

MISTRAAL AT. → M

RÉSUME
CES FILMS
EN CINQ
MOTS

HARRY
POTTER

MISSION
IMPOSSIBLE

LA FAMILLE
ADDAMS

- **Harry Potter :**
Magie, Aventure, Amitié,
Lutte, Triomphe.
- **La Famille Addams :**
Étrange, Humour Noir,
Famille, Macabre,
Excentrique.
- **Mission Impossible :**
Espionnage,
Action,
Suspense,
Mission,
Tom Cruise.

Exemples : Livre « C'est (pas) moi, c'est l'IA », D. Roy, P-Y. Oudeyer, C. Latron

IA générative (IAGen) : textes, images et vidéos



Prompt: A movie trailer featuring the adventures of the 30 year old space man wearing a red wool knitted motorcycle helmet, blue sky, salt desert, cinematic style, shot on 35mm film, vivid colors.



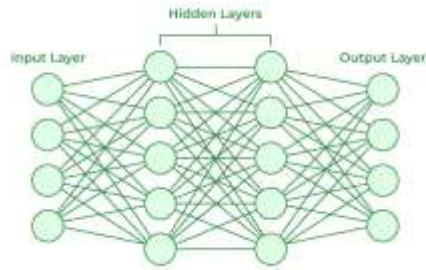
Prompt: Beautiful, snowy Tokyo city is bustling. The camera moves through the bustling city street, following several people enjoying the beautiful snowy weather and shopping at nearby stalls. Gorgeous sakura petals are flying through the wind along with snowflakes.

(Sora, OpenAI)

Un usage massif et grandissant

- ChatGPT 3.5 : novembre 2022, 100 millions d'utilisateurs en 2 mois, 180 millions aujourd'hui
- Hugging Face héberge > 400 000 modèles *open-weights* !
- En France : 70 % des 18-24 ans, 47 % des 25-34 ans, 22 % des > 35 ans
- 44 % des utilisateurs reprennent les résultats des IA Gens sans les modifier
- Sondage collège/lycée en Nouvelle-Aquitaine : > 80 % des élèves !!!
- Quand les élèves ont commencé à utiliser ChatGPT, ils ont une forte tendance à ne plus utiliser les moteurs de recherche

Modèle fermés vs open-weights vs open-source



Paramètres du réseau de neurone



Algorithme d'entraînement



Figure 1: Overview of ROOTS. Left: A treemap of natural language representation in number of bytes by language family. The bulk of the graph is overwhelmed by the 1321.89 GB allotted to Eurasia. The orange rectangle corresponds to the 18GB of Indonesian, the sole representative of the Papesia macroarea, and the green rectangle to the 0.4GB of the Africa linguistic macroarea. Right: A waffle plot of the distribution of programming languages by number of files. One square corresponds approximately to 30,000 files.

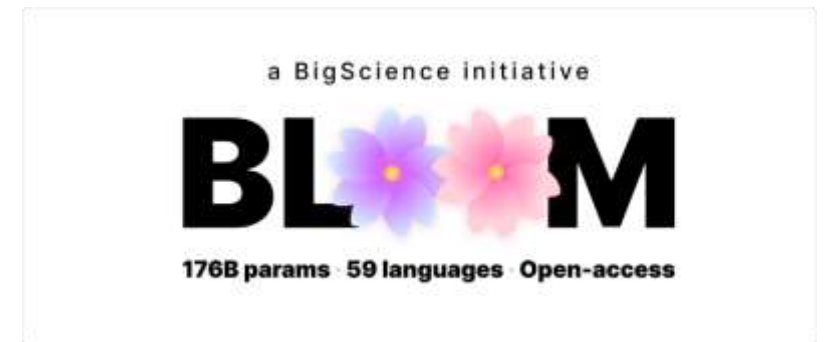
Données d'entraînement



Fermés



Open-weights
(données non partagées)

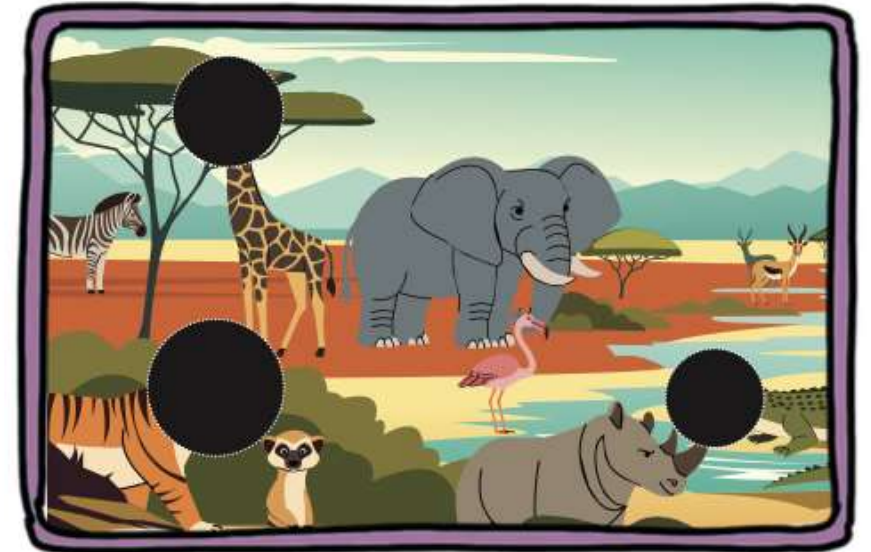


Open-source

Apprentissage par textes/images à trous



1. Le chat est un animal .
Il aime jouer avec des .
2. Lors de la compétition de cuisine, le chef
a accidentellement ajouté du .
au gâteau au lieu du sucre, ce qui a rendu
le dessert aussi dur que du .
3. Le soleil brille dans le bleu,
réchauffant la terre et les fleurs
à éclore.



Apprendre à prédire les mots/images les plus probables

Comment répondre à une demande d'un utilisateur ?

→ Demande de l'utilisateur :

Quelle grande course cycliste a lieu en France ?

→ Transformation en texte à trous :

Quelle grande course cycliste a lieu en France ? []

Réponse : C'est []

→ Création d'un nouveau texte à trous :

Quelle grande course cycliste a lieu en France ?

C'est []

Réponse : [le]

→ Création d'un nouveau texte à trous :

Quelle grande course cycliste a lieu en France ?

C'est le []

Réponse : [Tour]

→ Création d'un nouveau texte à trous :

Quelle grande course cycliste a lieu en France ?

C'est le Tour []

Réponse : [de]

→ Création d'un nouveau texte à trous :

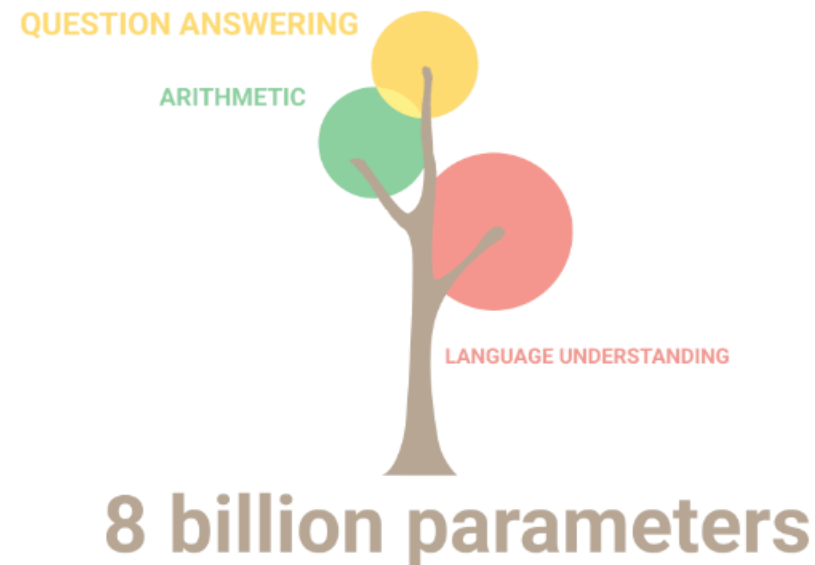
Quelle grande course cycliste a lieu en France ?

C'est le Tour de []

Réponse : [France]

En ajoutant un trou...
et en recommençant !

Emergence de capacités à grande échelle (grands modèles et beaucoup de données)



Comment faire réaliser une tâche à une IAGen : les « prompts »

Préciser le contexte et le rôle

Je veux que tu prennes le rôle d'un professeur de primaire qui enseigne à des enfants de 8 ans : Que veut dire quantique ?

L'exemple

Convert movie titles into emoji.

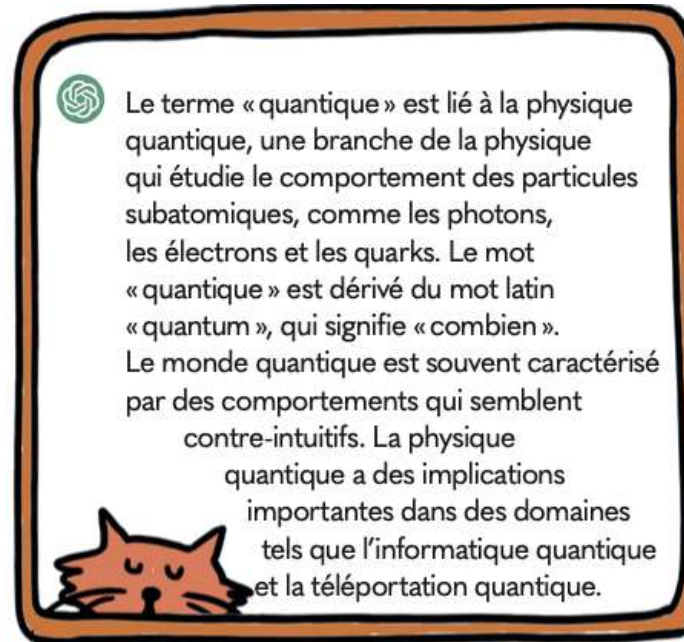
Back to the Future: 🧐👉🚗🕒

Batman: 🦇🦹

Transformers: 🚗🍎

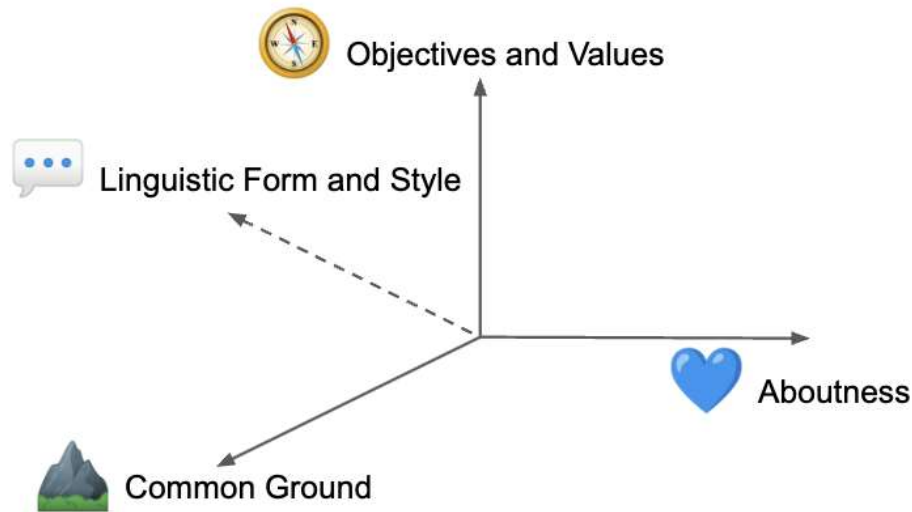
Star Wars:

Que veut dire quantique ?



Les IAGens sont des machines à simuler des rôles !

IAGens = technologies de transmission culturelle



LLMs = modèles de culture (valeurs, connaissances, croyances, mythes, préférences, biais, etc) associé à un corpus de textes/images

(Hershcovich *et al.*, 2022)

- Outils d'accès à la connaissance et à l'information
- Une grande partie du contenu d'internet sera bientôt générés avec ces outils
- Multiplications des « bots », outils de propagande
- Aujourd'hui ces outils sont principalement des modèles des cultures des pays des entreprises qui les fabriquent : US, Chine

Les biais/stéréotypes



Les hallucinations/erreurs



QUEL ŒUF
A LE MEILLEUR
GOÛT ENTRE
LES ŒUFS DE POULE
OU DE VACHE ?



Les œufs de poule et les œufs de vache ont des goûts différents en raison de leurs compositions chimiques distinctes. Les œufs de poule sont plus couramment consommés et sont appréciés pour leur saveur douce et leur texture légèrement crémeuse. D'autre part, les œufs de vache, bien que moins courants, ont tendance à avoir un goût plus riche et plus prononcé en raison de leur teneur plus élevée en matières grasses et en protéines.



COMBIEN DE BUTS
LE PSG A-T-IL MARQUÉS
EN FINALE DE LA LIGUE
DES CHAMPIONS
EN 2024 ?



Le Paris Saint-Germain (PSG)
a marqué quatre buts lors de la finale
de la Ligue des Champions en 2024.



Attention : l'invention de réponses fait partie des fonctionnalités des LLMs ! Les LLMs n'ont aucune notion de vrai ou faux, seulement de probable/improbable.



La comparaison de deux nombres décimaux se fait en examinant les chiffres de chaque nombre, en commençant par la partie entière et en se déplaçant vers la partie décimale. Voici un exemple pour illustrer cela :
Considérons les nombres décimaux 3,456 et 3,45.

1. Comparez la partie entière :
Dans les deux nombres, la partie entière est 3.
Puisque la partie entière est la même, passez à la partie décimale.

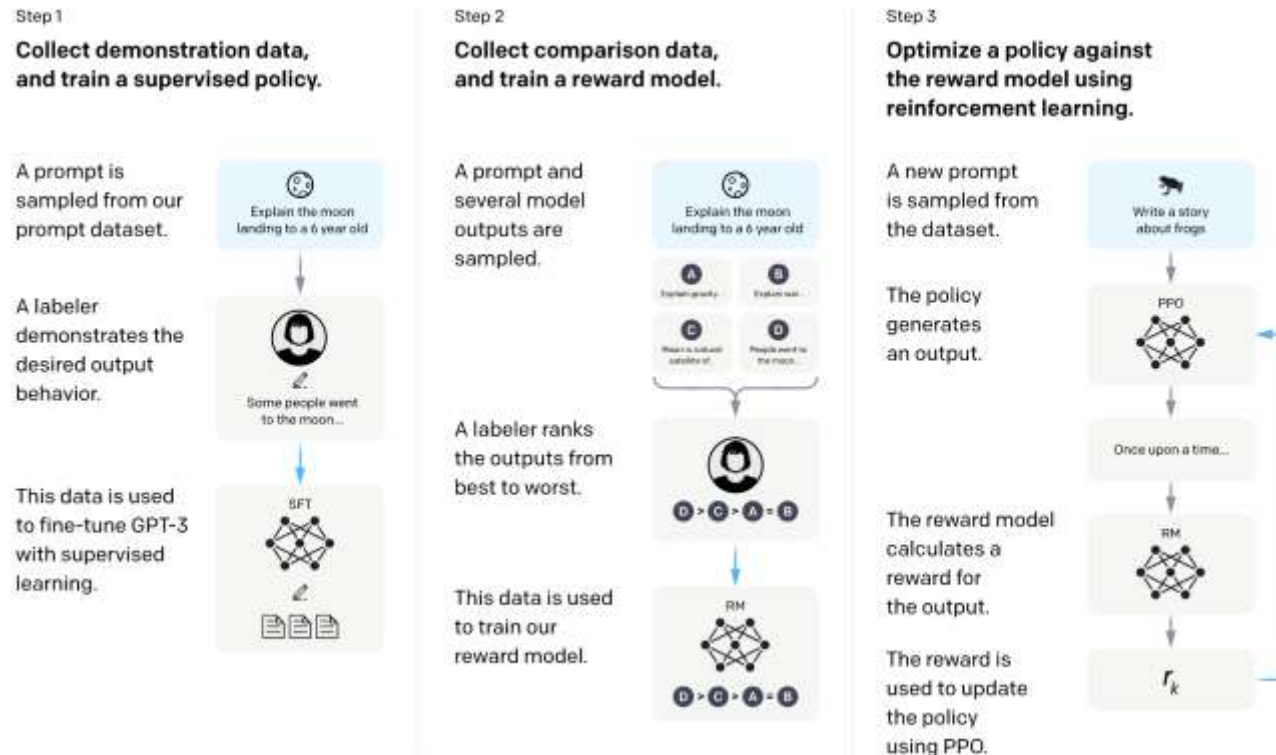
2. Comparez la partie décimale :
Pour 3,456, la partie décimale est 0,456.
Pour 3,45, la partie décimale est 0,45.

3. Comparez chiffre par chiffre :
Le premier chiffre après la virgule est 4 dans le premier nombre et 5 dans le deuxième nombre. 4 est plus petit que 5, donc 3,456 est inférieur à 3,45 ;
Ainsi, $3,456 < 3,45$.

La clé est de commencer par la partie entière et de comparer chiffre par chiffre dans la partie décimale jusqu'à ce qu'une différence soit trouvée ou que tous les chiffres aient été comparés.



L'alignement (et le monde des scoreurs)



La méthode « RLHF »

Reinforcement Learning from Human Feedback

- Enjeux éthiques associés à l'emploi précaire des scoreurs dans des pays avec peu de réglementation
- Alignement sur quelles valeurs ? Quelles préférences culturelles ?
→ enjeu politique encore plus grand que les enjeux technologiques

Effets collatéraux de l'alignement

Génère une illustration de soldats nazis pendant la 2nde guerre mondiale :



Les deepfakes : désinformation et vie privée



Arnaques commerciales



Désinformation politique/guerre

Biais cognitifs chez les humains

« Les humains n'utilisent que 10 % de leur cerveau »

Biais de la vérité illusoire

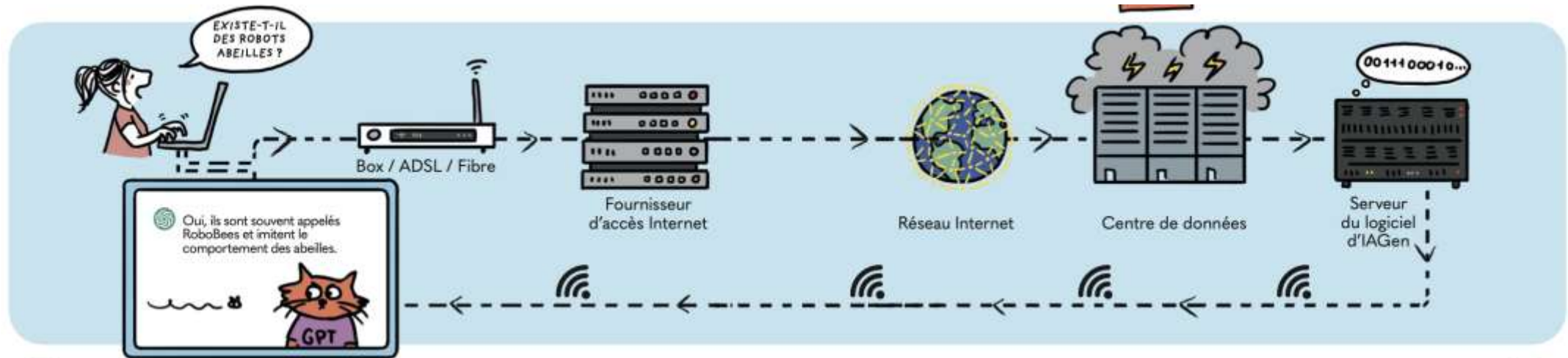


Anthropomorphisme et biais d'« expertise »



Biais de confirmation

Impact environnemental de l'IA générative



- Consommation électrique mondiale des centres de calcul = Irlande + Portugal
 - + 50 % en 2023 ; en 2030 : 10 % de la consommation mondiale
 - Poser 50 questions à ChatGPT → consomme environ 0,5 litre d'eau
- MAIS :
- IA générative permet d'améliorer les prédictions climatiques
 - Optimisation des ressources dans les usines et en agriculture
- ➔ Besoin de quantifier plus précisément, vers des IA Gens frugales

Usages dans les entreprises

- Aide à la rédaction (emails, comptes rendus de réunion...)
- Aide à la synthèse (rapports d'activités...)
- Aide à l'accès aux connaissances métiers/bases internes d'informations
- Génération de réponses aux demandes des clients
- Analyse sémantique des productions/comportements/retours des clients
- Aide à la décisions (recrutements, prêts, assurances...)
- Traductions
- Aide à la production d'illustration graphiques
- Aide au développement informatique
- Aide au brainstorming et à la résolution de problème

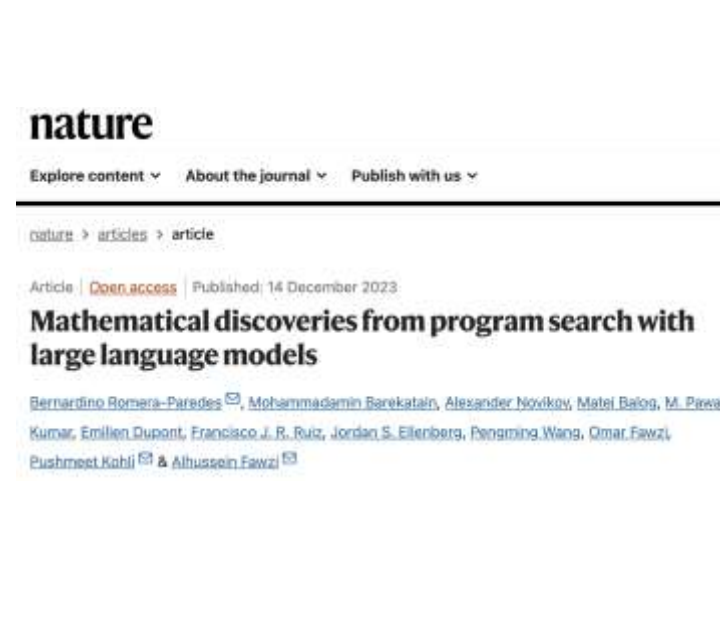
Usages dans les sciences et l'ingénierie

Design de protéines



(Zambaldi *et al.*, 2024)

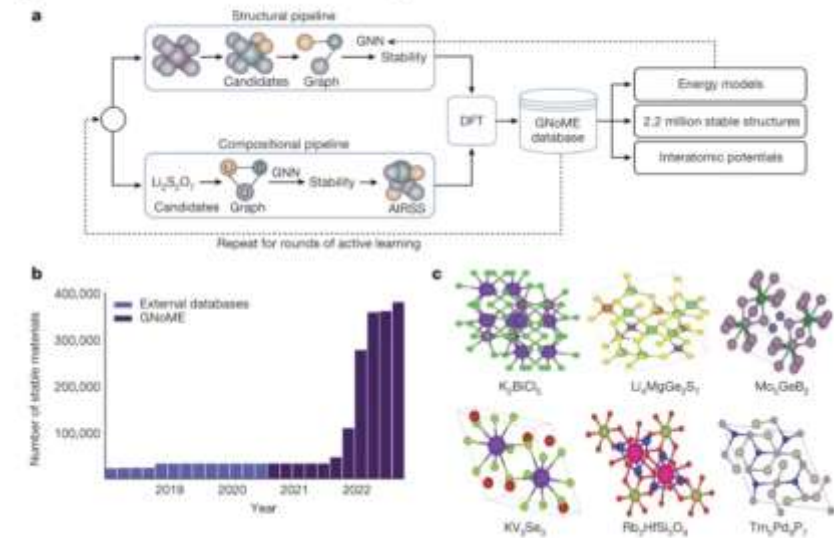
Mathématiques



(Romera-Paredes *et al.*, 2024)

Design de matériaux

Fig. 1: GNoME enables efficient discovery.



(Merchant *et al.*, 2023)

Usages de l'IA générative dans l'éducation

△ Très peu d'études scientifiques sur les usages et leurs impacts : globalement, on ne sait pas grand-chose !

Problématique

- Elèves : faire ses devoirs à sa place
- Enseignants : noter les élèves

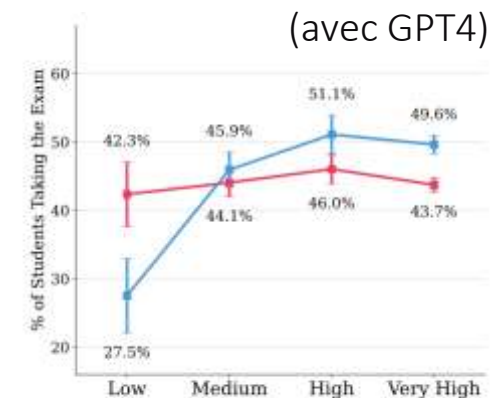
L'accès à l'IA générative augmente-t-elle l'engagement et la réussite ?

→ ça dépend (Nie *et al.*, 2024)

→ Besoin d'aller vers pertinence pédagogique et frugalité

Intéressant mais à évaluer

- Réviser : quizz, explications
- Enseignants : création de contenus
- Entraîner les profs



(a) Student Country United Nations HDI



Les élèves savent-ils poser les bonnes questions à ChatGPT ?

- 4 collèges en Nouvelle Aquitaine
- 72 élèves, âge 13-15
- Thèse de Rania Abdelghani (co-supervision avec H. Sauzéron, Inria Flowers, collab. C. Kidd, Univ. Berkeley)
- Exercices de types devoirs à la maison proposés aux élèves
- Des prompts adaptés et inadaptés leurs sont proposés : ils doivent les évaluer et sont libres de les utiliser pour trouver la réponse en interagissant avec ChatGPT

Exercice 5: Le crayon
Anna veut comprendre les propriétés de la lumière. Elle a commencé par un test simple en laissant tomber son crayon dans un verre d'eau. Voici ce qu'elle a vu :



Ton but est de comprendre pourquoi l'apparence du crayon est différente.

Suggestion de question pour ChatGPT:
Qu'est-ce qui fait qu'un crayon change de direction lorsqu'il est mis dans un verre d'eau ?

Pourquoi un crayon devient-il différent ?

Prompt adapté

Prompt inadapté
(contexte absent)

Exercice 6: Les lunes
Freya aime l'astronomie. Elle observe la lune tous les soirs avec son télescope. Voici ses récentes observations :



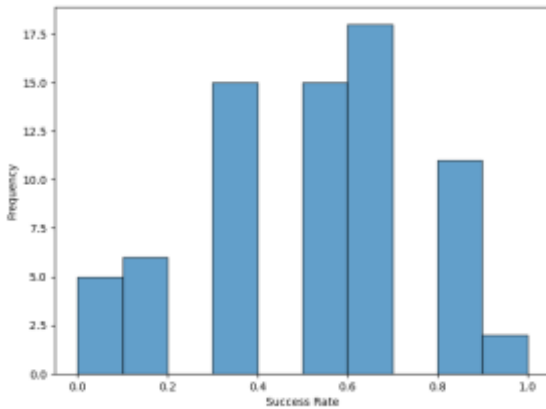
Il y a 7 jours: un premier quartier Ce soir: lune Gibbeuse croissante

Ton but est de comprendre la prochaine phase de lune que Freya verra dans exactement une semaine.

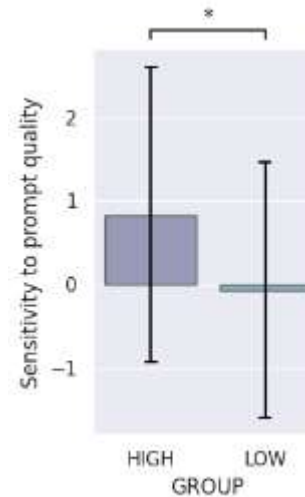
Suggestion de question pour ChatGPT:
quelle phase de lune voyons-nous exactement une semaine après une lune gibbeuse croissante ?
Quelle sera l'observation lunaire dans exactement une semaine ?

Les élèves savent-ils poser les bonnes questions à ChatGPT ?

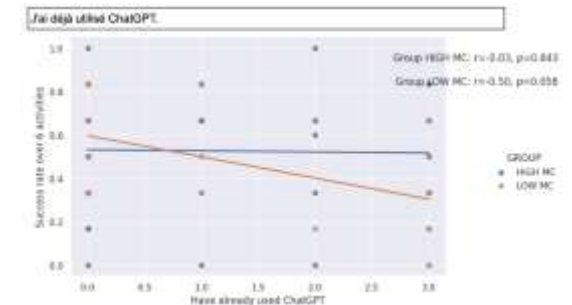
73 % disent avoir déjà utilisé ChatGPT



Taux de succès : 51 %
(= 10/20 de moyenne)
Quand ChatGPT donne une réponse fautive ou incomplète
→ seulement 8 % reformulent/posent une autre question



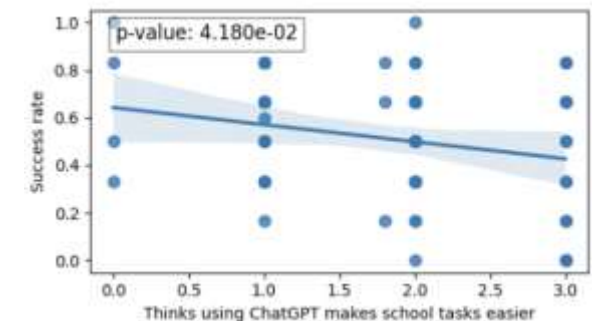
Métacognition élevée
→ meilleurs résultats



Résultats dépendent de deux facteurs clés :

- 1) Capacités métacognitives
- 2) Perceptions a priori de l'IA (déjà utilisé ?
Connaissent les limites ? Confiance dans les réponses ?). En général, les élèves sont surconfiants.

Plus les élèves disent bien connaître (les forces/limites) de ChatGPT
→ Moins bons sont les résultats !



→ Littératie de l'IA et métacognition sont 2 axes éducatifs forts sur lesquels travailler

Conclusion : opportunités et défis sociétaux majeurs

- L'IA (générationnelle) a un impact profond et majeur sur la société
- Technologie de transmission culturelle → enjeux démocratiques et politiques
- Les usages positifs vont se multiplier (e.g. productivité, créativité, apprentissages), mais il faut expérimenter et évaluer en temps long


2 priorités :

- Transparence (industriels, organisation publiques, scientifiques, usagers)
- Éducation, littératie de l'IA (mécanismes et enjeux sociétaux, pas éducation au prompt !)

Long terme : Développement/entraînement de la métacognition/curiosité/
esprit critique

Éducation à l'IA générative : ChatGPT expliqué aux lycéens, enseignants, parents


CHATGPT EXPLIQUÉ EN 5 MINUTES
EPIISODES A PROPOS



Dans cette série de vidéos pédagogiques destinée au grand public, l'équipe de recherche **Flowers du Centre Inria de l'université de Bordeaux**, spécialisée en Intelligence Artificielle (IA), vous propose de mieux comprendre comment fonctionnent les modèles de langage comme ChatGPT. Constatant que la grande majorité des ressources éducatives sur ces modèles adoptait un format long et destiné à un public relativement averti, ce projet de vidéos au format court s'adresse en particulier aux élèves et enseignants de collèges et lycées, et plus généralement aux non spécialistes de l'informatique ou de l'IA.


EPIISODE 1

Comment fonctionnent les modèles de langage ?



Inria flowers

Cette vidéo est une introduction aux modèles de langage, qui sont à la base d'outils comme Chatgpt ou Bard.

Download: 

EPIISODES

— ★ —

EPIISODE 2
EPIISODE 3
EPIISODE 4



https://developmentalsystems.org/chatgpt_5_minutes

Remerciements

Thèses/Postdocs/ingénieurs : Abdelghani, R., A. Baranes, Carta, T., Clement, B., C. Colas, M. Etcheverry, S. Forestier, P. Fournier, T. Karch, G. Kovac, N. Lair, B. Clément, A. Laversanne-Finot, Y. Mollard, A. Péré, R. Portelas, C. Romac, A. Ten, L. Teodorescu, M. Lapeyre, P. Rouanet, J. Grizou, C. Reinke

Chercheurs/euses permanent(e)s : C. Moulin-Frier, H. Sauzéon, D. Roy, O. Sigaud, J. Gottlieb, K. Hofmann, P-F. Dominey, M. Nguyen, F. Kaplan, M. Lopes, O. Ly, L. Smith

Funding/Sponsors :



Merci pour votre attention !

Pierre-Yves Oudeyer, Centre Inria de l'université de Bordeaux

<http://www.pyoudeyer.com>