

ChatGPT

Anaïs Halftermeyer et Cherifa Ben Khelil

LIFO - Université d'Orléans & LIFAT - Université de Tours

RITUEL Mars 2023

Objet de cette présentation

Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétences

Un "artefact
génératif"

Puissance de feu

Conséquences sur
la langue

Conséquences sur
l'information

Conséquences sur
les discours

Quels risques au
sens larges ?

"C'est toi le chat". Résumé d'un billet de Olivier Ertzscheid (MCF sciences de l'information, Univ. de Nantes) parut dans son blog : [Affordance.info](https://affordance.info) le 2 janvier 2023

Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétences

Un "artefact
génératif"

Puissance de feu

Conséquences sur
la langue

Conséquences sur
l'information

Conséquences sur
les discours

Quels risques au
sens larges ?

Le TAL

- tâches :
- méthodes :
- compétences :
- utilisation :

Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétencesUn "artefact
génératif"

Puissance de feu

Conséquences sur
la langueConséquences sur
l'informationConséquences sur
les discoursQuels risques au
sens larges ?

- tâches : traduction automatique, résumé, résolution de coréférences, etc.
- méthodes : symboliques, numériques, hybrides, ...
- compétences : modeliser la langue, programmer, analyser les erreurs, ...
- utilisation : détection de spams, correcteurs orthographiques, ...

Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétencesUn "artefact
génératif"

Puissance de feu

Conséquences sur
la langueConséquences sur
l'informationConséquences sur
les discoursQuels risques au
sens larges ?

- tâches : traduction automatique, résumé, résolution de coréférences, etc.
- méthodes : symboliques, numériques, hybrides, ...
- compétences : modeliser la langue, programmer, analyser les erreurs, ...
- utilisation : détection de spams, correcteurs orthographiques, ...

L'ATALA " *Si la traduction automatique reste une application majeure du TAL, bien d'autres applications ont vu le jour depuis 50 ans, pour former les "industries des langues". Celles-ci sont en plein essor et s'intègrent peu à peu dans notre quotidien. Elles s'appuient sur des recherches jetant une passerelle entre la linguistique et l'informatique sans oublier la statistique, fort utile pour extraire des données du gigantesque réservoir de textes disponibles sur le web.* "

14 Le Monde

ChatGPT, le logiciel capable d'écrire des petits textes ...

Le 6 décembre 2022, la société OpenAI a lancé une nouvelle application de messagerie intitulée ChatGPT. Basée sur le modèle de langage GPT-3 développé par...

6 déc. 2022

8 Les Echos

ChatGPT : 5 questions sur cette IA qui tient enfin une ...

Qui est la société derrière ChatGPT ? OpenAI est un des principaux laboratoires de recherche sur l'intelligence artificielle au monde. Son but affiché est de «...»

7 déc. 2022

3 Franceinfo

"Il peut déjà s'exprimer avec des nuances" : ChatGPT, l ...

"Il peut déjà s'exprimer avec des nuances" : ChatGPT, l'intelligence artificielle capable de discuter avec des humains. Le site ChatGPT compte plus d'un million...

9 déc. 2022

8 Les Echos

ChatGPT : plongée dans les entrailles du chatbot qui ...

La start-up californienne OpenAI a récemment mis en ligne ChatGPT, un robot conversationnel capable de tenir une conversation avec un humain de façon bluffante.

12 déc. 2022

1 France Live

ChatGPT : c'est quoi exactement ce Google du futur ?

L'intelligence artificielle (IA) vient de faire un pas de géant. Mis à la disposition du grand public fin novembre, ChatGPT dépasse déjà le million...

17 déc. 2022

11 RFI

L'intelligence artificielle ChatGPT affole le monde des ...

ChatGPT, c'est l'addition d'un outil de chat (conversation, en anglais) et d'un *Generative Pre-trained Transformer (GPT)*, ce qui en fait une véritable révolution.



5 Korben

Mon nouvel ami qui sait tout sur tout #chatGPT

Ensuite pour des problèmes médicaux, elle peut me mettre sur la piste de quelques trucs sans soucis... Photo de Chat GPT, l'expert en intelligence artificielle.

2 déc. 2022

0 Next Impact

ChatGPT : Open AI rend son bot plus responsable

Hier, Open AI a annoncé la mise en ligne de son nouveau bot de discussion ChatGPT, successeur d'InstructGPT, avec beaucoup plus de pinçettes. Et les chercheurs...

1 déc. 2022

5 Siecle Digital

En attendant GPT-4, OpenAI met au point le modèle de ...

text-davinci-003, ChatGPT : ces outils sont basés sur GPT-3.5 d'OpenAI. À la fin du mois de novembre, OpenAI a annoncé le lancement d'une nouvelle version du...

2 déc. 2022

0 i24NEWS

Les Robots Conversationnels Toujours Plus Convaincants ...

Les conversations avec ChatGPT, postées notamment sur Twitter par des internautes fascinés, montrent une sorte de machine omnisciente, capable d'expliquer...

4 déc. 2022

5 Analytics India Magazine

Will ChatGPT Really be The Google Killer?

This new model from OpenAI uses a novel training method and is based on the GPT-3.5 architecture. What makes ChatGPT impressive is that besides answering...



Quelle avancée ?

Fluidité

Test de Turing "personnel" bluffant !

Genre et style

Imitation d'auteurs

Multilingage

Le gros plus c'est l'interconnection, peut produire du code par exemple, générer une demande pour Dall-E

Multitache

Produit du texte en langue naturelle, du code, et explique les règles sous-jacentes à cette production

Notre sujet

Domaine

Quoi de plus ?

**Externalisation
des compétences**

Un "artefact
génératif"

Puissance de feu

Conséquences sur
la langue

Conséquences sur
l'information

Conséquences sur
les discours

Quels risques au
sens larges ?

Fraude et plagiat ?

- Au sein de la société :

Fraude et plagiat ?

- Au sein de la société :
 - le "braconnage culturel" grâce aux IA génératives

Braconnage culturel ? " Michel de Certeau élabore ainsi, en parallèle à la théorisation du système panoptique de Michel Foucault, surveillance et contrôle "par le haut" de la société, une théorie des tactiques de résistance au champ de l'autre, subversion mais de l'intérieur et de la base même du système. Si ceux qui écrivent semblent imposer leur pouvoir à ceux qui disent et font, Certeau montre bien que les publics ne sont pas si dominés et restent actifs devant la réception des messages qu'on leur envoie, avec des paroxysmes critiques quand le "dire" s'écarte trop du "faire" (multiplication des épisodes mystiques du XVIIIe siècle ; prise de parole de mai 1968 ; théologie de la libération en Amérique du Sud, pour citer les domaines dans lesquels il était plus particulièrement impliqué)." Ertzscheid 2023

Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétences

Un "artefact
génératif"

Puissance de feu

Conséquences sur
la langue

Conséquences sur
l'information

Conséquences sur
les discours

Quels risques au
sens larges ?

Fraude et plagiat ?

- Au sein de la société :
 - le "braconnage culturel" grâce aux IA génératives
 - bibliothécaires versus moteurs de recherche

Fraude et plagiat ?

- Au sein de la société :
 - le "braconnage culturel" grâce aux IA génératives
 - bibliothécaires versus moteurs de recherche
- Enseignement : faire évoluer nos méthodes ! comment...?

Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétencesUn "artefact
génératif"

Puissance de feu

Conséquences sur
la langueConséquences sur
l'informationConséquences sur
les discoursQuels risques au
sens larges ?

Fraude et plagiat ?

- Au sein de la société :
 - le "braconnage culturel" grâce aux IA génératives
 - bibliothécaires versus moteurs de recherche
- Enseignement : faire évoluer nos méthodes ! comment...?



Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétencesUn "artefact
génératif"

Puissance de feu

Conséquences sur
la langueConséquences sur
l'informationConséquences sur
les discoursQuels risques au
sens larges ?

Connaissances et compétences humaines

- Maintenance et **compréhension** des systèmes et modèles
- Apport discutable de la **performance** à certaines tâches
- Suppression d'**emplois**

Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétencesUn "artefact
génératif"

Puissance de feu

Conséquences sur
la langueConséquences sur
l'informationConséquences sur
les discoursQuels risques au
sens larges ?

Peux-tu me lister les compétences que tu as et qui pourraient être salariées dans notre société ?



En tant que modèle de langage, je ne suis pas une personne physique ayant des compétences professionnelles qui pourraient être salariées dans notre société.



Cependant, mes capacités de traitement du langage naturel pourraient être utilisées dans divers domaines, notamment :

1. Développement de chatbots : Les chatbots sont des programmes informatiques conçus pour communiquer avec les utilisateurs en utilisant le langage naturel. Mes capacités de traitement du langage naturel peuvent être utilisées pour développer des chatbots avancés pour diverses entreprises.
2. Traduction automatique : Mes capacités en matière de traduction automatique peuvent être utilisées pour traduire des documents et des textes en plusieurs langues. Cette compétence est particulièrement utile pour les entreprises qui opèrent à l'international.
3. Analyse de données : Mes capacités de traitement du langage naturel peuvent être utilisées pour l'analyse de données textuelles, comme la classification de données, la reconnaissance de tendances et la détection de sentiments. Cette compétence est utile pour les entreprises qui souhaitent analyser les commentaires des clients, les médias sociaux et les avis en ligne.
4. Résumé automatique : Mes capacités en matière de traitement du langage naturel peuvent être utilisées pour créer des résumés automatiques de documents et de textes. Cette compétence est utile pour les entreprises qui souhaitent résumer rapidement de grands volumes de données.
5. Recherche d'informations : Mes capacités de traitement du langage naturel peuvent être utilisées pour rechercher et collecter des informations à partir de diverses sources en ligne. Cette compétence est utile pour les entreprises qui ont besoin de collecter rapidement des données pour des projets spécifiques.

Il est important de noter que ces compétences ne peuvent pas être salariées directement chez moi, mais peuvent être utilisées par des entreprises pour développer des produits et services innovants.

Quelle compétence effective ?

Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétences

Un "artefact
génératif"

Puissance de feu

Conséquences sur
la langue

Conséquences sur
l'information

Conséquences sur
les discours

Quels risques au
sens larges ?

" Agencement collectif d'énonciation "

selon la terminologie de Deleuze et Guattari

Quelle compétence effective ?

Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétencesUn "artefact
génératif"

Puissance de feu

Conséquences sur
la langueConséquences sur
l'informationConséquences sur
les discoursQuels risques au
sens larges ?

" Agencement collectif d'énonciation "

selon la terminologie de Deleuze et Guattari

Agencement collectif d'énonciation ? " *Ce que l'on subsume derrière le mot écran "d'intelligence artificielle" ce sont avant tout des agencements (divers) collectifs (épars) de production (textuelle, icônique, vidéographique ...). Il est vrai que ces agencements peuvent faire de grandes choses. Un peu comme des foules peuvent initier de grand mouvements. Mais tentez donc de mesurer l'intelligence d'une foule ...* "Ertzscheid 2023

Quelle compétence effective ?

Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétences

**Un "artefact
génératif"**

Puissance de feu

Conséquences sur
la langue

Conséquences sur
l'information

Conséquences sur
les discours

Quels risques au
sens larges ?

Quelle compétence effective ?

Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétencesUn "artefact
génératif"

Puissance de feu

Conséquences sur
la langueConséquences sur
l'informationConséquences sur
les discoursQuels risques au
sens larges ?

Agencement collectif d'énonciation ?

" (...) Il y a "primat d'un agencement machinique des corps sur les outils et les biens, primat d'un agencement collectif d'énonciation sur la langue et les mots. (...) un agencement ne comporte ni infrastructure et superstructure, ni structure profonde et structure superficielle mais aplatît toutes ses dimensions sur un même plan de consistance où jouent les présuppositions réciproques et les insertions mutuelles.(...) mais si l'on pousse l'abstraction, on atteint nécessairement à un niveau où les pseudos-constants de la langue font place à des variables d'expression, intérieures à l'énonciation même " (...) " L'unité réelle minima, ce n'est pas le mot, ni l'idée ou le concept, ni le signifiant mais l'agencement. C'est toujours un agencement qui produit les énoncés. Les énoncés n'ont pas pour cause un sujet qui agirait comme sujet d'énonciation pas plus qu'ils ne se rapportent à des sujets comme sujets d'énoncé. L'énoncé est le produit d'un agencement toujours collectif qui met en jeu en nous et dehors de nous des populations, des multiplicités, des tentations, des devenirs, des affects, des évènements." tiré de Dialogues 2008 et Mille plateaux 1980, Deleuze et Guattari et cité par Ertzscheid 2023

Quelle compétence effective ?

Non supervisé et post supervisé

post supervision ou correcteurs : Kenyans à 2 de l'heure pour rendre ChatGPT moins toxique...



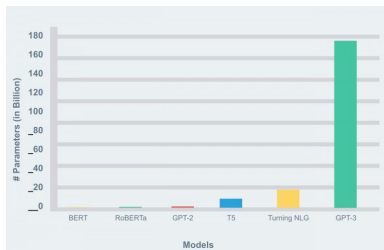
Billy Perrigo, "Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic" dans Time Magazine 18/01/2023

Corpus et paramètres

500 000 000 000 000 de mots, le corpus :

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

un ordre de grandeur pour les paramètres :



mais moins de paramètres que le Mégatron Turing de Microsoft et plus de corpus à venir pour GPT4 si on en croit Rob Toews, "[10 AI Predictions For 2023](#)" dans Forbes 20/12/2022

Développement continu

La répétition

" Plus que dans toute autre forme d'art, les itérations de l'art génératif sont les répétitions invisibles d'une production dont nous ne sommes que le déclencheur opportuniste et anecdotique. Car l'essentiel de toutes les questions que nous pouvons poser à GPT-3 ont déjà été posées par d'autres que nous-mêmes ; elles ont déjà été répétées." Ertzscheid 2023

La langue : passer du politique au philosophique

La question politique :

- **capitalisme linguistique de Kaplan**
- longueur des textes générés, accès payant à GPT4
- statut des corpus, de la post-supervision

La question philosophique :

- condition de production, de génération, d'artificialisation, de mise en circulation
- la commande vocale comme conditionnement
- "la prévisibilité calculatoire"

Le discours : épuisement des productions linguistiques

L'auteur termine son billet sur ces 3 énoncés :

- *"Toutes les données texte humaines auront été traitées."*
François Chollet, Google
- *"Il n'y aura plus que des réponses."* Marguerite Duras
- *"Qu'arrivera-t-il si nous répondons mal à des requêtes comme amour ou ouragan ?"* Apostolos Gerasoulis, Ask Jeeves

Prise en compte des risques

On peut citer plusieurs initiatives (voir références supplémentaires)
Laura Weidinger propose une typologie de 21 risques répartis en 6 grandes catégories :

- 1 Discrimination, exclusion et toxicité
- 2 Les dangers de l'information
- 3 Méfaits de la désinformation
- 4 Utilisations malveillantes
- 5 Méfaits de l'interaction homme-machine
- 6 Automatisation, accès et dommages environnementaux.

ou encore Emily Bender qui observe le "perroquet stochastique" dans les détails. Elle propose d'investir comme autant de questions de recherche :

- 1 considération du coût environnemental et financier
- 2 conservation et documentation soigneuse des datasets
- 3 circonscription des développements à des questions de recherche bien établies
- 4 considération des valeurs des parties en jeu
- 5 encouragement vers des problématiques de recherche dépassant l'enjeu de la taille des corpus, des modèles...

Références supplémentaires

Notre sujet

Domaine

Quoi de plus ?

Externalisation
des compétencesUn "artefact
génératif"

Puissance de feu

Conséquences sur
la langueConséquences sur
l'informationConséquences sur
les discoursQuels risques au
sens larges ?

- Abid, Abubakar, Maheen Farooqi, and James Zou. 2021. "Persistent Anti-Muslim Bias in Large Language Models." In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, 298–306. AIES '21. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3461702.3462624>.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 83E9C." In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–23. FAccT '21. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>.
- Birhane, Abeba. 2022. "ChatGPT, Galactica, and the Progress Trap." Wired, December 9, 2022. <https://www.wired.com/story/large-language-models-critique/>.
- Kurita, Keita, Paul Michel, and Graham Neubig. 2020. "Weight Poisoning Attacks on Pre-Trained Models." ArXiv:2004.06660 [Cs, Stat], April. <http://arxiv.org/abs/2004.06660>.
- Li, Linyang, Ruotian Ma, Qipeng Guo, Xiangyang Xue, and Xipeng Qiu. 2020. "BERT-ATTACK: Adversarial Attack Against BERT Using BERT." In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 6193–6202. Association for Computational Linguistics. <https://www.aclweb.org/anthology/2020.emnlp-main.500>.
- Perez, Fábio, and Ian Ribeiro. 2022. "Ignore Previous Prompt: Attack Techniques For Language Models." arXiv. <https://doi.org/10.48550/arXiv.2211.09527>.
- Perrigo, Billy. 2023. "The \$2 Per Hour Workers Who Made ChatGPT Safer." Time, January 18, 2023. <https://time.com/6247678/openai-chatgpt-kenya-workers/>.
- Steed, Ryan, Swetasudha Panda, Ari Kobren, and Michael Wick. 2022. "Upstream Mitigation Is Not All You Need: Testing the Bias Transfer Hypothesis in Pre-Trained Language Models." In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 3524–42. Dublin, Ireland: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.acl-long.247>.
- Wallace, Eric, Shi Feng, Nikhil Kandpal, Matt Gardner, and Sameer Singh. 2019. "Universal Adversarial Triggers for Attacking and Analyzing NLP." In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2153–62. Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1221>.
- Weidinger, Laura, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, et al. 2021. "Ethical and Social Risks of Harm from Language Models." arXiv. <https://doi.org/10.48550/arXiv.2112.04359>.