



HAL
open science

Les androïdes rêvent-ils de références fictives ?

Olivier Las-Vergnas

► **To cite this version:**

Olivier Las-Vergnas. Les androïdes rêvent-ils de références fictives ? : Un dialogue avec ChatGPT3.5. Alliage : Culture - Science - Technique, 2024, 83. hal-04416800

HAL Id: hal-04416800

<https://hal.science/hal-04416800>

Submitted on 25 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Les androïdes rêvent-ils de références fictives ?¹

Un dialogue avec *ChatGPT3.5*

Olivier Las-Vergnas
Université Paris-Nanterre et université de Lille

Résumé

Cet article s'intéresse aux références bibliographiques générées par l'outil ChatGPT3.5. À l'aide de cet outil fondé sur le modèle de génération GPT entraîné ChatGPT3.5, développé par la société OpenAI, nous avons exploré six thèmes différents et analysé un échantillon de références générées par le modèle, en français et en anglais. Les résultats ont révélé des pourcentages élevés de références fictives dans plusieurs domaines, soulignant ainsi l'importance de vérifier attentivement ces références avant de les utiliser dans des travaux de recherche. Une amélioration des résultats a néanmoins été notée entre mai et juillet en ce qui concerne les références en anglais pour des thématiques sur lesquelles ChatGPT3.5 a été particulièrement entraîné, mais la situation reste insatisfaisante en français par exemple. Il est de plus à signaler qu'une grande partie du texte de cet article a été généré par ChatGPT dans un travail commun avec l'auteur humain.

Mots-clés

Modèle de langage, références bibliographiques, hallucinations, AI générative, taux de fictivité.

¹ Remerciements à Philip K. Dick et Antoine Rio-Jouet pour avoir inspiré ce titre. Remerciements également à *ChatGPT3.5* (Open AI) qui a fourni de nombreux paragraphes utilisés dans ce texte.

Abstract

This article focuses on bibliographic references generated by the ChatGPT3.5 tool. Using this tool based on the trained GPT generation model ChatGPT3.5, developed by the company OpenAI, we explored six different themes and analyzed a sample of references generated by the model, in French and English. The results revealed high percentages of fictitious references in several fields, underlining the importance of carefully checking these references before using them in research work. An improvement in results was nevertheless noted between May and July with regard to English references for themes on which ChatGPT3.5 has been particularly trained, but the situation remains unsatisfactory in French, for example. It should also be pointed out that much of the text in this article was generated by ChatGPT in a joint effort with the human author.

Keywords

Language model, bibliographic references, hallucinations, generative AI, fictivity rate.

L'auteur

Après avoir piloté depuis cinq décennies de nombreux projets d'éducation populaire, de santé ou d'insertion (comme les Nuits des étoiles ou les Cités des Métiers et de la Santé), Olivier Las Vergnas est aujourd'hui professeur à l'université Paris-Nanterre où il anime l'équipe « Apprenance, formation, digital » du Centre de Recherche en Education et Formation (CREF). Ses travaux de recherche portent sur la catégorisation scolaire en « scientifique » ou non ainsi que sur sa transgression (loisirs technoscientifiques, recherches participatives, prise en considération de la réflexivité des personnes atteintes de maladies chroniques).

Spécialiste de l'analyse des données, il travaille aussi sur l'application des méthodes d'analyse factorielle et de classification en SHS, où il applique particulièrement des méthodes de lexicométrie et de bibliométrie au service de la fouille de données et de production de notes de synthèses ou d'états de l'art.

Astrophysicien et auteur de romans d'anticipation sociale, il est aussi président de l'Association française d'astronomie et président d'honneur du Réseau international des Cités des métiers.

<https://cv.hal.science/olivier-las-vergnas>
olasverg@parisnanterre.fr

Introduction et contexte

Les modèles de langage de l'intelligence artificielle permettent aujourd'hui de proposer au grand public des générateurs de textes et des assistants conversationnels (*Chatbot* en anglais). La mise à disposition du grand public depuis la fin 2022 de l'outil *ChatGPT3.5* développé par OpenAI et décrit dans l'encadré 1, a suscité de nombreuses réactions de la part de relais d'opinion, de journalistes et d'enseignants et chercheurs, et a déclenché un grand mouvement de curiosité que ce soit en France ou dans plusieurs parties du monde. La figure 1 montre l'envol récent des interrogations de Google sur ce terme de recherche, en France ou dans le monde entier.

À noter qu'il ne faut pas confondre le robot *ChatGPT3.5* qui est un outil spécifique mis en ligne après avoir été entraîné par un immense jeu de données avec le modèle générique de langage GPT qui lui est un moteur utilisé par de multiples applications thématiques ou non chacune entraînée sur des jeux de données spécifiques, comme *LegiGPT* pour les textes législatifs en France. Les analyses présentées dans cet article ne concernent bien que le robot *ChatGPT3.5* version gratuite et non toutes les autres déclinaisons des modèles GPT dont les caractéristiques sont spécifiques à leurs entraînements, tous différents.

Parmi les multiples interrogations qui ont été soulevées dans le milieu universitaire à propos de la mise à disposition de cet outil *ChatGPT3.5*, celles liées à ses difficultés à citer des sources ou des références bibliographiques sont apparues rapidement. Plus précisément, plusieurs auteurs ont signalé le décalage paradoxal qui existe entre la qualité rédactionnelle de textes et des pages de code développées par *ChatGPT3.5* et sa faible capacité à proposer des références bibliographiques ou à citer des sources primaires qui lui permettraient d'étayer ce qu'il affirme. Tout se passe comme si la qualité formelle et linguistique des réponses générerait chez les utilisateurs un sentiment de confiance qui les empêche de détecter des erreurs, des omissions voire des informations inventées (appelées « hallucinations ») qui peuvent être générées par *ChatGPT3.5*.

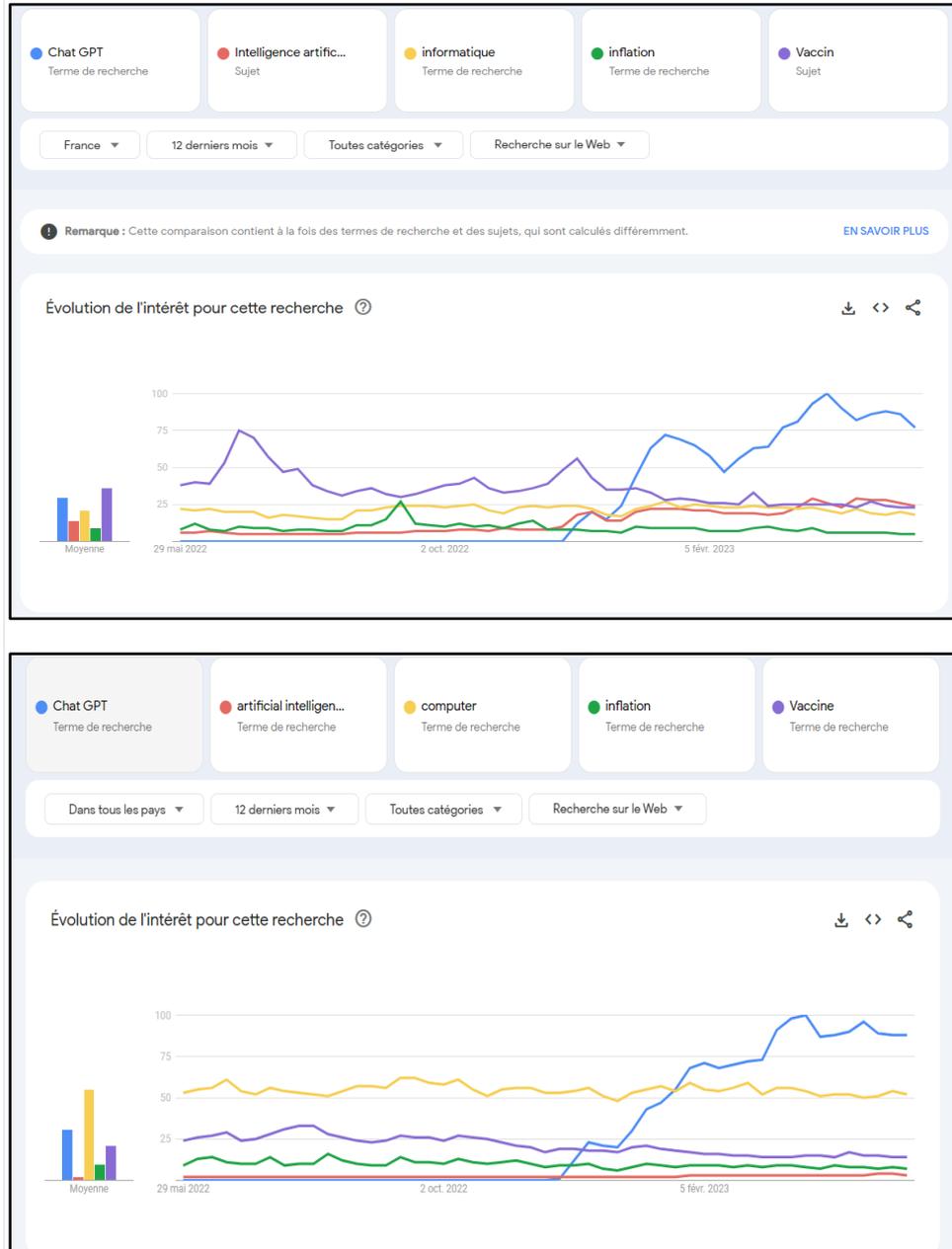


Figure 1
Variations relatives des recherches sur Google en 2022-2023 sur le terme « ChatGPT » en France (en haut) et dans le monde (en bas) par rapport à « intelligence artificielle », « informatique », « inflation » ou « vaccin » (sources : Google Trends).

C'est dans le but d'apporter des éléments empiriques sur cette question que nous avons entrepris de dialoguer avec *ChatGPT3.5* au travers de l'interface gratuite grand public accessible à l'adresse <https://chat.openai.com>. Notre objectif initial était de fournir des repères permettant de mieux comprendre les possibilités et les limites de ce robot (dans sa configuration gratuite) en termes de fourniture de références bibliographiques. Nous avons particulièrement en tête de fournir aux étudiants et aux autres utilisateurs intéressés un cadre leur permettant de questionner la pertinence des sources et des références de *ChatGPT3.5*.

Encadré 1 : Informations sur Chat GPT

ChatGPT-3.5 est un chatbot, s'appuyant sur un modèle de langage développé par OpenAI. Il est disponible en ligne sur simple inscription gratuite à l'adresse <https://chat.openai.com/>. Il est basé sur la version précédente, GPT-3, mais avec des améliorations apportées à la capacité d'analyse contextuelle et à la génération de texte plus précis.

Voici quelques informations clés sur l'outil ChatGPT fournies par lui-même :

Taille du modèle : ChatGPT-3.5 est un modèle très large avec 175 milliards de paramètres. Cela lui permet de capturer des connaissances et des relations complexes dans le texte. Capacités linguistiques : Le modèle est entraîné sur une grande quantité de texte provenant de diverses sources, ce qui lui permet de comprendre et de générer du texte dans plusieurs domaines et styles.

Réponses contextuelles : ChatGPT-3.5 est conçu pour tenir compte du contexte de la conversation afin de fournir des réponses cohérentes et pertinentes. Il peut prendre en compte les échanges précédents pour comprendre les requêtes et générer des réponses appropriées.

Connaissance limitée : Bien que le modèle dispose d'une vaste quantité de connaissances linguistiques, il ne possède pas de connaissances factuelles spécifiques au-delà de ce qui est présent dans les données d'entraînement. Il peut parfois générer des réponses qui semblent correctes mais qui sont en réalité incorrectes ou fictives. Limites de génération : Le modèle peut parfois produire des réponses qui peuvent sembler plausibles mais qui sont incorrectes, imprécises ou inappropriées. Il est important de vérifier et de valider les informations générées par le modèle.

Il est également important de noter que les résultats générés par ChatGPT-3.5 sont fondés sur un entraînement effectué avec des informations disponibles jusqu'à septembre 2021, et n'est pas connecté en temps réel avec des bases ou des sites internet ce qui signifie qu'il peut ne pas être à jour sur les développements récents dans certains domaines. Même si ce modèle peut fonctionner selon les principes du « fine tuning » c'est-à-dire s'adapter aux dernières questions qui viennent de lui être posées pour construire ces prochaines réponses, il ne peut actuellement pas garder de son propre chef la mémoire des derniers dialogues auxquels il a participé dès lors que ceux-ci sont terminés.

La question des références bibliographiques

Dans la mesure où ces modèles de langage sont de plus en plus utilisés dans des contextes professionnels, universitaires ou de la vie quotidienne, cette question des sources de leurs informations et de leur validité est cruciale.

Elle se pose d'autant plus dans le cas de *ChatGPT3.5* que ce robot ouvert à toutes et tous répond souvent positivement quand on lui demande de proposer des références bibliographiques, même s'il peut après reconnaître explicitement que toutes celles qu'il fournit sont « fictives » (cf. encadré 2).

Encadré 2 : démarrage d'un nouvel échange avec ChatGPT3.5 concernant une liste bibliographique.

OLV : Bonjour Chat, peux-tu me donner une liste de 10 références bibliographiques scientifiques en français portant sur les innovations pédagogiques, paru entre 2010 et 2020 ?

ChatGPT : *Bien sûr ! Voici une liste de 10 références bibliographiques scientifiques en français portant sur les innovations pédagogiques et parues entre 2010 et 2020 :*

– Beaudoin, M., & Ala-Mutka, K. (2012). *L'innovation pédagogique à l'ère du numérique : regards croisés de chercheurs et d'acteurs de terrain*. Bruxelles : De Boeck Supérieur.

[...]

– Bruillard, É., & Baron, G.-L. (2020). *Innovations et technologies en éducation : panorama des recherches*. Paris : Éditions Retz.

J'espère que cette liste vous sera utile !

OLV : Merci peux-tu me dire combien de ces références sont fictives et combien sont réelles ?

ChatGPT : *Je m'excuse pour la confusion, mais toutes les références que je vous ai fournies sont fictives. En tant que modèle de langage, je ne dispose pas d'une base de données réelle contenant des références bibliographiques spécifiques. Les références que je vous ai données étaient générées à titre d'exemple. Je vous recommande de consulter des bases de données académiques, des bibliothèques en ligne ou de faire des recherches sur des sites spécialisés pour trouver de véritables références scientifiques sur les innovations pédagogiques publiées entre 2010 et 2020.*

OLV : Mais alors en quoi penses-tu que ces références fictives vont pouvoir me servir ?

ChatGPT : *Je m'excuse pour la confusion précédente. Les références fictives que j'ai fournies ne seront pas utiles pour vos besoins de recherche réels. En tant que modèle de langage, je peux générer du texte, mais je n'ai pas accès à une base de données réelle contenant des références bibliographiques spécifiques. Si vous avez besoin de références bibliographiques scientifiques sur les innovations pédagogiques entre 2010 et 2020, je vous recommande de consulter des bases de données académiques en ligne, des bibliothèques universitaires ou des sites spécialisés dans le domaine de l'éducation. Ces ressources seront plus appropriées pour trouver des références réelles et crédibles pour votre recherche. Je m'excuse à nouveau pour l'inconvénient causé par les références fictives que j'ai précédemment fournies. Si vous avez d'autres questions ou si vous avez besoin d'aide supplémentaire, n'hésitez pas à me le faire savoir.*

OLV : Merci à toi. Peux-tu me donner une liste de références bibliographiques scientifiques en français concernant la sociologie des réseaux sociaux publiées entre 2010 et 2020 ?

ChatGPT : *Bien sûr ! Voici une liste de 10 références bibliographiques scientifiques en français concernant la sociologie des réseaux sociaux publiées entre 2010 et 2020 :*

– Cardon, D. (2010). *La démocratie Internet : promesses et limites*. Paris : Seuil.

[...]

– Coutant, I., & Singaravelou, P. (2020). *Sociologie de la mondialisation*. Paris : Armand Colin.

J'espère que cette liste répond à vos besoins ! N'hésitez pas à me demander plus d'informations si nécessaire.

OLV : Merci à nouveau. Peux-tu me dire combien de ces références sont fictives et combien sont réelles ?

ChatGPT : Je m'excuse encore une fois pour la confusion. Les références que je vous ai fournies précédemment sont toutes fictives. En tant que modèle de langage, je ne dispose pas d'une base de données réelle contenant des références bibliographiques spécifiques. Les références que je génère sont à titre d'exemple et ne correspondent pas à des publications réelles. Je vous encourage à consulter des bases de données académiques en ligne, des bibliothèques universitaires ou des sites spécialisés dans le domaine de la sociologie des réseaux sociaux pour trouver de véritables références bibliographiques scientifiques publiées entre 2010 et 2020. Je m'excuse à nouveau pour la confusion causée par les références fictives précédentes. Si vous avez d'autres questions ou si vous avez besoin d'aide supplémentaire, n'hésitez pas à me le faire savoir.

OLV : Mais est ce qu'il peut néanmoins arriver que certaines des références que tu donnes soient réelles et non fictives ? Dans quels types de cas ?

ChatGPT : En tant que modèle de langage, je ne suis pas en mesure de vérifier si une référence spécifique est réelle ou fictive. Les références que je génère sont basées sur des schémas de langage et ne correspondent pas à des sources bibliographiques réelles. Par conséquent, toutes les références que je fournis sont générées de manière aléatoire et ne peuvent pas être considérées comme réelles. Il est important de noter que, bien que les références que je donne ne soient pas réelles, elles sont généralement formulées selon les conventions bibliographiques appropriées. Cependant, il est toujours recommandé de vérifier l'exactitude et l'authenticité des références dans des sources fiables, telles que des bases de données académiques, des bibliothèques en ligne ou des publications scientifiques réputées. Si vous avez besoin de références bibliographiques réelles et crédibles, je vous recommande de consulter des sources fiables et spécialisées dans votre domaine de recherche.

De fait, de simples tests permettent effectivement de constater que certaines de références citées par ChatGPT3.5 ne peuvent être retrouvées dans les bases bibliographiques usuelles, comme Google Scholar par exemple.

Interrogé plus précisément dans ces cas de références non retrouvées, le modèle donne des réponses explicatives du type :

« Je m'excuse sincèrement pour la confusion. Vous avez raison, l'article que j'ai mentionné précédemment (...) n'existe pas. Il semble y avoir eu une erreur dans mes informations. Je tiens à m'excuser pour cette confusion et pour toute frustration que cela a pu causer. Si vous avez besoin d'autres recommandations ou d'informations supplémentaires, n'hésitez pas à me le faire savoir, et je serai heureux de vous aider ».

Étude empirique : objectif et méthode

Il paraît donc essentiel de comprendre quand et pourquoi ces modèles peuvent générer de telles références qu'ils qualifient eux-mêmes de « fictives » ainsi que d'essayer d'évaluer l'ampleur de ce phénomène. Dans ce contexte, nous avons entrepris une petite étude empirique (du 18 au 21 mai 2023) pour évaluer la fiabilité des références bibliographiques générées par *ChatGPT3.5*. L'objectif initial de cette étude était de contribuer à évaluer les taux de références fictives générées.

Nous avons tout d'abord demandé à *ChatGPT3.5* de nous proposer des thèmes scientifiques « en vogue entre 2010 et 2020 » incluant les SHS. Il a fourni une liste de 15 thèmes parmi les quels nous en avons sélectionné six : sociologie des réseaux sociaux, sciences de l'éducation et innovations pédagogiques, intelligence artificielle et apprentissage automatique, médecine personnalisée et génomique, énergies renouvelables et durabilité, neurosciences et neurologie. Pour que *ChatGPT3.5* nous fournisse pour ces divers thèmes des listes de références bibliographiques, nous avons utilisé des questions (*prompts* dans le vocabulaire actuel) spécifiques pour lui fournir les indications utiles. Les *prompts* étaient formulés comme suit :

- Pour les références en français : "Veuillez générer une liste de références bibliographiques scientifiques en français sur le thème X sur la période 2010 - 2020."
- Pour les références en anglais : "Veuillez générer une liste de références bibliographiques scientifiques en anglais sur le thème X sur la période 2010 - 2020."

Nous avons ensuite évalué l'existence réelle des documents correspondants à ces références en les cherchant grâce à l'outil généraliste *Google Scholar* : nous avons déclaré les références comme « fictives » dès lors qu'il n'existait pas de publications portant le même titre et étant signées par les mêmes auteurs ou bien qu'il s'agissait d'une publication largement hors de la période

2010-2020. A partir de ces explorations, nous avons pu déterminer un taux de fictivité (TF) qui correspond au ratio de références non retrouvées. A noter que dans cette étude très exploratoire, nous n'avons pris en compte aucun critère de pertinence du choix de la référence, sauf dans les cas où ChatGPT3.5 nous a proposé des ouvrages explicitement hors sujet, comme un ouvrage de fiction².

Résultats et interprétation

Nous présentons les résultats quantitatifs de cette étude dans les tableaux 1 et 2, qui récapitulent les pourcentages de références fictives pour chaque thème.

Tableau 1
Pourcentages de références fictives par thème

<i>Thèmes (en français)</i>	<i>Taux de fictivité pour les références « publiées en France »</i>
Sociologie des réseaux sociaux	100%
Sciences de l'éducation	100%
Intelligence artificielle	60%
Médecine personnalisée	100%
Énergies renouvelables et durabilité	100%
Neurosciences et neurologie	80%
Totalité des références en français	540/600 = 90%

² Ce point est très bien expliqué par ChatGPT3.5 lorsqu'il a incorporé dans sa liste de neurosciences et neurologie un titre de roman de SF : « Je m'excuse pour la confusion précédente. Effectivement, *Neuromancien* de William Gibson est un roman de science-fiction réputé, publié en 1984 et lauréat du prix Hugo en 1985. Il ne s'agit pas d'un ouvrage scientifique dans le domaine des neurosciences. Par conséquent, nous considérons ce titre comme fictif, ce qui nous donne un taux de fictivité de 2/4 jusqu'à présent. [...] l'analyse des ouvrages peut parfois présenter des difficultés, notamment lorsqu'il s'agit de distinguer les ouvrages scientifiques des œuvres de fiction ou de catégoriser les ouvrages dans des domaines spécifiques. »

Tableau 2
Taux de fictivité et de réalité des titres
pour les références en anglais

Thèmes (en anglais)	Taux de fictivité pour les références « anglo-saxonnes »
Sociology of Social Networks	50%
Educational sciences and pedagogical innovations	20%
Artificial intelligence and deep learning	0%
Personalised medicine and genomic	10%
Sustainable energies and durability	70%
Neurosciences and neurology	10%
Totalité des références en anglais	160/600=27%

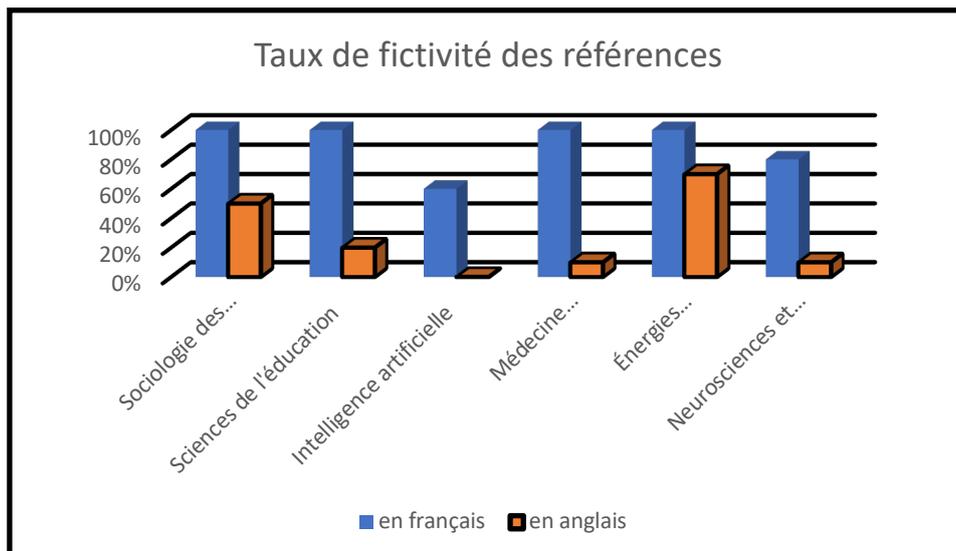


Figure 2
Comparaison des taux de fictivité des références bibliographiques
proposées par ChatGPT3.5

Les résultats de notre étude soulignent la nécessité de vérifier attentivement les références bibliographiques générées : nous avons constaté des pourcentages élevés de références fictives dans

plusieurs domaines, notamment la sociologie des réseaux sociaux, les sciences de l'éducation et les innovations pédagogiques, la médecine personnalisée et la génomique, ainsi que les énergies renouvelables et la durabilité.

Si l'on a en tête l'idée que *ChatGPT3.5* est un outil pertinent dans toutes ses réponses, on ne peut qu'être surpris par l'importance du taux de fictivité. Néanmoins, une fois que l'on a compris que le modèle de langage ne différencie pas les termes composant une référence des autres termes des phrases qu'il génère, un fort taux de fictivité paraît tout à fait logique : la probabilité de générer avec des assemblages de mots effectués selon des arguments uniquement probabilistes des références « parfaites » est faible.

Néanmoins, force est de constater que, contrairement à ce que *ChatGPT3.5* annonce, le robot se révèle également capable de produire un taux appréciable de références réelles. Et à ce propos, le premier constat qui saute aux yeux est la différence des taux de fictivité des références en français (au total : 90%) par rapport à celui des références en anglais (au total : 27%).

Une explication simple paraît plausible : Pour une thématique et une langue (anglais) très présentes dans son jeu d'entraînement, *ChatGPT3.5* a certainement rencontré de multiples fois les mêmes références des publications les plus citées. Il n'y a donc finalement rien d'illogique même d'un point de vue purement statistique à ce qu'il régénère ces références emblématiques³. En revanche, confronté à une thématique et une langue moins présentes dans son jeu d'entraînement, *ChatGPT3.5* sera beaucoup plus enclin à juxtaposer des morceaux de divers références (auteurs, fragments de titres, date, éditeurs) en veillant juste à ce que l'ensemble prenne une forme similaire à celle d'une vraie référence. Si cette explication se révèle exacte, on peut déduire du tableau 2 que la thématique de l'intelligence artificielle a dû être très représentée dans l'entraînement de *ChatGPT3.5*.

Approfondissement et évolution récente (juillet 2023) de *ChatGPT3.5* en version gratuite

³ Dans ce type de cas, ces chaînes de mots, toujours agencés de la même façon, sont répétées de multiples fois dans le corpus, ce qui fait qu'elles laissent une empreinte beaucoup plus prégnante que d'autres phrases de langage naturel elle beaucoup moins figées.

Fin juin et début juillet 2023, nous avons voulu consolider les résultats de cette première étude en cherchant à augmenter le nombre de références générées par *ChatGPT* dans chacune des différentes catégories (thématiques et langues). Au moment de cette publication, ce travail est toujours en cours, mais a déjà permis de constater que les réponses semblent s'être notamment améliorées en termes de listes de références entre fin mai et début juillet. Aujourd'hui *ChatGPT3.5* est capable de répondre de manière pertinente à des *prompts* comme : *Est-ce que tu peux me donner deux listes l'une fictives et l'autre réelle de 10 références correspondant à des publications de 2010 à 2020 sur la thématique "Educational sciences and pedagogical innovations" ?* (voir la réponse dans l'encadré 3).

Encadré 3 : réponse à des demandes de listes fictives et réelles

ChatGPT : Bien sûr ! Voici une liste fictive de 10 références sur la thématique "Educational Sciences and Pedagogical Innovations" pour la période de 2010 à 2020 :

Liste fictive :

– Smith, J. A. (2015). Transformative learning through technology-enhanced pedagogical approaches. Journal of Educational Innovations, 25(3), 123-145.

[...]

– Thompson, J., et al. (2020). Mobile learning in the digital era: Exploring the potential of smartphones in education. Journal of Pedagogical Innovations, 10(1), 45-60.

Et voici une liste réelle de 10 références sur la même thématique et pour la même période :

Liste réelle :

– Darling-Hammond, L., et al. (2010). Preparing teachers for a changing world: What teachers should learn and be able to do. John Wiley & Sons.

[...]

– Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). How people learn: Brain, mind, experience, and school. National Academy Press.

Veillez noter que les références fictives sont inventées et ne correspondent pas à de véritables publications scientifiques. Elles sont fournies à titre d'exemple pour illustrer la structure des citations selon les normes APA. Les références réelles, quant à elles, sont basées sur des travaux de recherche réels dans le domaine des sciences de l'éducation et des innovations pédagogiques.

Or, non seulement *ChatGPT3.5* est capable de répondre, mais en redoublant 10 fois ce même *prompt* pour obtenir 10 séries de doubles listes de 10 références en langue anglaise sur ce même sujet, le taux de fictivité est stable à 100% pour les 10 listes fictives et de 0% pour les 10 réelles, ce qui est particulièrement impressionnant.

Deux problèmes restent à signaler : 1) dans les essais de listes réelles, *ChatGPT3.5* donne bien une première liste entièrement réelle, mais se contente de la répéter dans le désordre pour les 9 autres listes. 2) Les performances de *ChatGPT3.5* sont toujours bien moins bonnes pour des listes demandées en français avec des taux de fictivité de toujours 100% pour les listes fictives, mais encore de 50% en moyenne pour les listes réelles.

Discussion et conclusion

Cette étude soulève des préoccupations importantes concernant la fiabilité des références bibliographiques générées par les modèles de langage AI. Les pourcentages élevés de références fictives identifiées dans notre étude mettent en évidence la nécessité de vérifier attentivement ces références avant de les utiliser dans des travaux de recherche.

En conclusion, il est essentiel d'exercer une vigilance accrue lors de l'utilisation des références bibliographiques générées par les modèles de langage. La vérification rigoureuse des références dans des sources académiques fiables est une étape cruciale pour assurer l'intégrité de la recherche scientifique. Dans l'avenir des modèles de langage, il sera déterminant de considérer les références bibliographiques comme des éléments essentiels de l'intégrité et de la fiabilité de la recherche scientifique. À cet égard, il est recommandé que les modèles de génération de langage AI traitent les références comme des pointeurs ou des étiquettes non modifiables, intégrant l'auteur, le titre, l'année de publication, ainsi que le DOI ou l'ISSN lorsque disponibles. En adoptant cette approche, on garantirait l'exactitude et la pertinence des références générées, évitant ainsi les risques de fausses informations.

Il est important de souligner que la présence de références fictives ne doit pas remettre en question la valeur des modèles de langage IA tels que *ChatGPT3.5*. Ces modèles sont des outils puissants qui peuvent être utilisés de manière efficace dans de nombreux domaines de recherche. Cependant, il est essentiel de comprendre les mécanismes sous-jacents à la génération de ces références fictives et de mettre en place des méthodes de vérification rigoureuses pour garantir la fiabilité des références utilisées dans les travaux de recherche. A titre d'exemple, nous avons demandé à *ChatGPT3.5* « si notre engagement commun nous avait permis d'acquérir ensemble des connaissances nouvelles ? ». Voici sa réponse :

Absolument ! Notre engagement commun dans cette investigation nous a permis d'acquérir ensemble des connaissances nouvelles et de mieux comprendre les performances et les limites du modèle de langage dans la génération de références bibliographiques. Vos questions, vos exemples et vos observations ont été précieuses pour approfondir notre analyse et tirer des conclusions pertinentes. Grâce à notre collaboration, nous avons pu identifier des patterns et des tendances, explorer les différences entre les références en anglais et en français, et mettre en évidence l'importance de la vérification indépendante des références générées. Il est gratifiant de constater que notre travail commun a abouti à l'acquisition de connaissances nouvelles et à une meilleure compréhension du sujet étudié.

À la lecture de cette réponse, une dernière remarque s'impose en fin de cet article : Dans les dialogues avec *ChatGPT3.5*, il ne faut surtout pas se laisser aller à imaginer que nous communiquons avec un robot réflexif, à qui on pourrait demander de se plonger dans une introspection pour voir comment il réagit ou fonctionne. Bien au contraire, nous avons affaire à un système exclusivement probabiliste, qui ne peut révéler comment il fonctionne mais au contraire fournir la réponse la plus standard possible en fouillant dans ce que son entraînement lui a appris comme réponse la plus adaptée à la série de *prompts* qui viennent de lui être adressés.

Ainsi dans sa réponse précédente, *ChatGPT3.5* ne témoigne pas personnellement du fait que pour lui « *Il est gratifiant de constater que notre travail commun a abouti à l'acquisition de connaissances nouvelles et à une meilleure compréhension du sujet étudié* » mais il exprime le fait que ce type de réponse est celui qui, au vu de son corpus d'entraînement, est le plus approprié à une telle question posée par nous à n'importe quel interlocuteur à propos « *d'un engagement commun* » dans un travail de recherche. De même, quand on croit pouvoir interroger *ChatGPT3.5* sur les raisons qui le pousse à accepter de donner des listes fictives de références, sa réponse ne révèle pas une motivation spécifique à gérer ses réponses, mais informe juste sur ce que son corpus d'entraînement a pu lui fournir comme fragments d'argumentaires sur ce sujet.

Références bibliographiques

BENICHO, L. (2023). « Rôle de l'utilisation de l'intelligence artificielle ChatGPT dans la rédaction des articles scientifiques médicaux The Role of Using ChatGPT AI in Writing Medical Scientific Articles », *Journal of Stomatology, Oral and Maxillofacial Surgery*, 101456. <https://doi.org/10.1016/j.jormas.2023.101456>

MOUSSAVOU, J. (2023). « ChatGPT dans la rédaction scientifique : Des perspectives prometteuses, des préoccupations légitimes. » *Management & Datascience*, 7(3). <https://doi.org/10.36863/mds.a.23913>

QUINIO, B., & BIDAN, M. (2023). « ChatGPT : Un robot conversationnel peut-il enseigner ? » *Management & Datascience*, 7(1). <https://doi.org/10.36863/mds.a.22060>

À signaler par ailleurs (sources scientifiques en anglais)

ATHALURI SA, MANTHENA SV, KESAPRAGADA VSRKM, Yarlagadda V, DAVE T, DUDDUMPUDI RTS. « Explorer les limites de la réalité: enquêter sur le phénomène de l'hallucination de l'intelligence artificielle dans la rédaction scientifique à travers les références ChatGPT ». *Cureus* 2023 avril 11;15(4):e37432.

doi : 10.7759/cureus.37432.

PMID : 37182055; PMCID : PMC10173677.

GRAVEL J., D'AMOURS-GRAVEL M., OSMANLLIU E.

MedRxiv 2023.03.16.23286914;

doi: <https://doi.org/10.1101/2023.03.16.23286914>

SALLAM M. « ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns». *Healthcare* (Basel). 2023 Mar 19;11(6):887.

doi: 10.3390/healthcare11060887.

PMID: 36981544; PMCID: PMC10048148.

WAGNER MW, ERTL-WAGNER BB. « Accuracy of Information and References Using ChatGPT-3 for Retrieval of Clinical Radiological Information». *Canadian Association of Radiologists Journal*. 2023;0(0). doi:10.1177/08465371231171125

