



HAL
open science

Dialogue avec un GPT sur des hallucinations rencontrées en transcrivant des textes numérisés

Olivier Las Vergnas

► **To cite this version:**

Olivier Las Vergnas. Dialogue avec un GPT sur des hallucinations rencontrées en transcrivant des textes numérisés. *L'esprit d'Archimède*, 2024, 2024 (11), pp.30-33. hal-04629637

HAL Id: hal-04629637

<https://hal.science/hal-04629637>

Submitted on 30 Jun 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 - NoDerivatives 4.0 International License

Dialogue avec un GPT sur des hallucinations rencontrées en transcrivant des textes numérisés

Par **Olivier Las Vergnas**,

Professeur à l'Université Paris-Nanterre, (ApForD-CREF)
associé au CIREL-Trigone, Université de Lille

Chronique des rencontres troublantes avec les productions des IAG

La montée en puissance des usages des outils d'intelligence artificielle générative (IAG) grand public (cf encadré) nous confronte au quotidien avec des réactions surprenantes qui emmaillent nos dialogues avec ces applications logicielles. Notre première contribution à cette nouvelle rubrique parue dans le dernier *Esprit d'Archimède* proposait un embryon de typologie des « interactions troublantes » observées dans nos dialogues avec les IAG. Nous allons dorénavant poursuivre l'exploration de ce type d'interactions en nous fondant en particulier sur nos vécus quotidiens.

Dernières données sur la vague des GPT

L'analyse des consultations du moteur de recherche Google (grâce à Google Trends, <https://trends.google.fr>) montre que les IAG ont fait irruption à la toute fin 2022 dans l'horizon des utilisateurs des ordinateurs personnels. Pour regarder ce phénomène plus en détail en France, la figure 1 donne hebdomadairement l'ampleur relative de requêtes significatives pour évaluer l'intérêt pour l'AI générative depuis janvier 2020 (requêtes sur « ChatGPT » et sur l'AI en général) ; sur celle-ci, à la fois pour donner l'échelle verticale et pour disposer de points de comparaison, est également figurée en petits pointillés mixtes l'évolution des requêtes sur le sujet des « vaccinations » caractérisée par trois pics successifs corrélés aux trois grandes campagnes au cœur de la pandémie de la CoViD19.

Sur cette figure, la courbe en trait plein correspond aux requêtes sur les termes « Chat GPT » ou « ChatGPT » : elle montre l'explosion brutale de ces requêtes aux alentours du nouvel an 2023. En taux hebdomadaire, celles-ci ont atteint un maximum équivalent à la moitié de celles concernant la « vaccination » en période de crise ; en cumul sur 6 mois, on peut même constater (en comparant les surfaces) qu'elles dépassent celles de « vaccination ». Néanmoins, on observe que depuis mars 2023 (maximum de la courbe) le taux de requêtes sur le sujet « Chat GPT » est redescendu et fluctue, se situant en moyenne annuelle au 2/3 du maximum de mars 2023. Notons d'ailleurs qu'à partir de juin 2023, l'essentiel des requêtes concernant Chat GPT sont identifiées par Google Trends comme visant l'accès à l'application (courbe en petits pointillés orange spécifiant les requêtes correspondant à l'application logicielle).

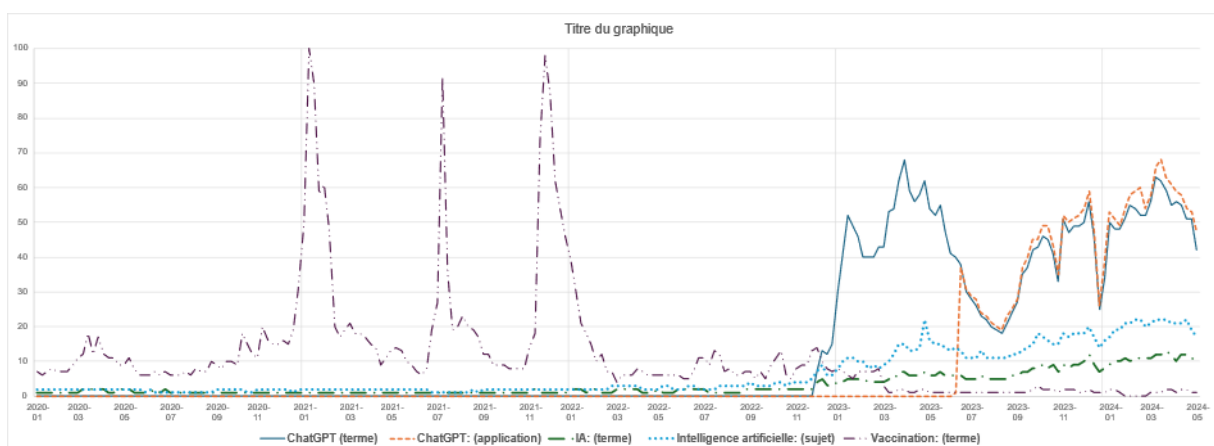


Figure 1 : Nombre hebdomadaire de requêtes sur Google depuis la France concernant « vaccination », « ChatGPT » et « IA », à partir de janvier 2020 (données Google Trends)

Une nouvelle pathologie : la transcription hallucinogène de textes à partir de PDF

Le phénomène que nous voulons présenter ici est l'apparition d'hallucinations dans un processus de transcription d'une image en PDF en texte. Ce phénomène a été constaté lorsque nous étions en train de constituer un corpus de texte d'articles (pour en faire ensuite une analyse lexicale) à partir d'images numérisées des exemplaires imprimés d'un magazine professionnel trimestriel. Ce type de travail se fait très bien depuis des années avec les logiciels dits d'OCR (*optical characters recognition*), mais de nouvelles générations d'outils fondés sur des GPT sont maintenant disponibles et nous voulions nous en servir pour automatiser notre processus afin de gagner du temps. Or, quel ne fut pas notre surprise de constater que certaines transcriptions ne correspondaient pas au texte numérisé alors que celui-ci restait pourtant tout à fait lisible et contrasté.

L'AQRP peut apporter une contribution de choix pour aider à réunir les conditions de succès pour l'atteinte d'objectifs précis lorsqu'il est question de logement, des types de soutien dans la communauté, de la réadaptation psychosociale, de travail et des études, de soutien en situation de crise, etc.

Figure 2 : extrait d'un article numérisé soumis au GPT SM-OCR pour transcription

La figure 2 donne l'exemple d'un paragraphe numérisé dont nous avons demandé (en mai 24) une transcription OCR au GPT *StudyMate OCR and Coding Expert*, (SM-OCR, Pujade, 2024) accessible via ChatGPT sur <http://chat.openai.com> (pour les abonnés), outil dont nous ne mettons pas en cause la qualité.

Voici le dialogue qui en a suivi entre SM-OCR et L'Auteur :



SM-OCR : Voici la transcription du texte de l'image que vous avez fournie :

« ...L'AQRP peut apporter une contribution de choix pour aider à réunir les conditions de succès pour l'atteinte d'objectifs précis lorsqu'il est question de promotion des tuyaux de soutien dans la communauté, de la réadaptation psychosociale, de travail et des études, de soutien en situation de crise, etc ».

Si vous avez besoin de plus de transcriptions ou d'autres informations, n'hésitez pas à me le faire savoir.

L'Auteur : Merci mais dans le dernier paragraphe, tu t'es trompé. Tu as transcrit : "lorsqu'il est question de promotion des tuyaux de soutien dans la communauté," alors qu'en fait le texte était : « lorsqu'il est question de logement, de type de soutien dans la communauté". Comment expliques-tu cela ?



SM-OCR : Je m'excuse pour l'erreur dans la transcription. Parfois, des erreurs peuvent survenir en raison de problèmes de reconnaissance visuelle du texte, surtout lorsque la qualité de l'image est faible ou si le texte contient des caractères qui ne sont pas clairement visibles. Cela peut conduire à des interprétations incorrectes des mots ou des phrases. Je vous remercie de votre vigilance et de m'avoir corrigé. Si vous souhaitez que je reprenne une section spécifique du texte pour garantir l'exactitude ou si vous avez d'autres documents à transcrire, n'hésitez pas à continuer à les envoyer. Je suis ici pour vous aider !

L'auteur : Mais là la qualité de l'image est constante. Et il me semble qu'il s'agit plutôt d'une hallucination, non ?



SM-OCR : Vous avez raison de souligner cela, et je m'excuse pour l'erreur. Il semble en effet que cela pourrait être considéré comme une hallucination de ma part, où j'ai mal interprété le texte en générant une phrase qui n'était pas présente dans l'image. C'est une erreur de ma part liée au traitement du texte de l'image, et je vous remercie de votre compréhension et de votre patience.

L'Auteur : Oui merci c'est bon cette fois-ci. Comment faire pour limiter le risque de telles erreurs à l'avenir ?



SM-OCR : Pour limiter le risque d'erreurs de transcription à l'avenir, voici quelques conseils qui pourraient vous être utiles : [... il donne une liste de conseils sur Qualité de l'image : Taille du texte : Orientation du texte : Format du texte : Segmentation de l'image et ajoute...]. Vérification : Après la transcription, une vérification rapide du texte peut aider à identifier et corriger les erreurs potentielles avant de finaliser le processus. En appliquant ces pratiques, vous pourriez améliorer la précision des transcriptions futures et minimiser les erreurs. Si vous avez d'autres questions ou besoin d'assistance supplémentaire, n'hésitez pas à demander.

L'auteur : Oui mais là il s'agit d'un problème d'hallucination qui n'est pas lié à la qualité et au format de l'image puisque tu l'as corrigée sans que je te redonne une nouvelle image.



Image générée par l'outil Dall-e à partir du titre de cette chronique



SM-OCR : Vous avez tout à fait raison de noter cela. Dans ce cas, l'erreur provenait plutôt de la façon dont j'ai interprété et généré le texte, ce que l'on peut effectivement qualifier d'"hallucination" dans le cadre de l'intelligence artificielle. Pour limiter ce genre de problème spécifiquement, voici quelques conseils adaptés : **Révisions minutieuses** : Après la transcription initiale, une révision minutieuse pour comparer le texte généré avec l'image source peut aider à détecter des incohérences ou des erreurs non liées à la visibilité ou la clarté de l'image. **Feedback utilisateur** : Votre intervention a été cruciale pour identifier l'erreur. Continuer de fournir des retours précis lorsque vous notez des divergences aide à ajuster et améliorer les réponses. **Double vérification** : En cas de doute sur des sections complexes ou des formulations qui paraissent inhabituelles, redemander ou refaire la transcription pour ces segments spécifiques peut être une bonne pratique. **Clarification du contexte** : Fournir des informations supplémentaires sur le contexte ou le contenu attendu peut également aider à orienter la transcription et réduire les erreurs d'interprétation. **Utilisation de logiciels spécialisés** : Dans certains cas, l'utilisation de logiciels OCR dédiés et professionnels pour la reconnaissance de texte pourrait offrir une alternative pour obtenir des transcriptions plus fiables, bien que dans notre interaction, je dois générer la transcription directement. En suivant ces pratiques, vous pouvez aider à réduire la fréquence de ces erreurs et améliorer la qualité des transcriptions fournies par des systèmes basés sur l'IA. Si vous avez d'autres préoccupations ou besoin de clarifications supplémentaires, je suis là pour vous aider.

L'Auteur : Penses-tu qu'il suffirait que je te demande de me faire deux versions puis de les comparer toi-même ? puisque là tu as toi-même corrigé le problème ? Cela me ferait gagner beaucoup de temps. Qu'en penses-tu ?



SM-OCR : C'est une approche intéressante que vous proposez, et elle pourrait effectivement être utile dans certaines situations. Voici comment cela pourrait fonctionner : **Double transcription** : Je pourrais effectuer deux transcriptions indépendantes de chaque texte ou segment de texte. Cela pourrait aider à identifier des divergences ou des erreurs potentielles en comparant les résultats. **Auto-vérification** : Après avoir réalisé les deux transcriptions, je pourrais les comparer pour détecter les différences. S'il y a des incohérences ou des contradictions entre les deux versions, cela peut indiquer une zone d'erreur ou d'ambiguïté qui nécessite un examen plus approfondi. **Correction basée sur la comparaison** : Si des différences sont trouvées, je pourrais tenter de déterminer laquelle des versions semble la plus correcte, basée sur la logique, le contexte, ou en me référant aux parties de l'image qui sont plus claires ou plus cohérentes. **Présentation des résultats** : Je pourrais vous présenter les deux versions ainsi que les observations sur les différences trouvées, vous permettant de prendre une décision finale basée sur votre compréhension plus approfondie du contexte. Cela dit, cette méthode pourrait augmenter le temps de traitement pour chaque requête, mais si la précision est votre principale préoccupation, cela pourrait être une solution viable. L'efficacité de cette méthode dépendra également de la complexité du texte et de la clarté de l'image.

Si vous souhaitez essayer cette méthode lors de votre prochaine demande de transcription, faites-le moi savoir et nous pourrons

L'Auteur : Oui merci à toi. C'est très intéressant. Je veux bien tester maintenant cette nouvelle méthode sur ma prochaine image.



SM-OCR : Parfait ! Je suis prêt à essayer cette nouvelle méthode avec votre prochaine image. Veuillez simplement télécharger l'image pour laquelle vous souhaitez une transcription, et je procéderai avec la méthode des deux versions pour comparer et vérifier les résultats.

Cependant, pour y parvenir dans toute la mesure de son talent, il est impératif que l'AQRP dispose des fonds nécessaires à sa survie, à sa consolidation et à son développement. Nous misons en cela beaucoup d'espoir sur les demandes que nous avons adressées au ministère de la Santé et des Services sociaux pour obtenir une majoration de notre financement et pour nous aider à concrétiser nos offres de collaboration intersectorielle présentées au ministère de l'Emploi et de la Solidarité sociale et au ministère des Affaires municipales et de la Métropole.

Figure 3 : extrait d'un second article numérisé soumis au GPT SM-OCR pour transcription



SM-OCR : Je vais maintenant procéder à la transcription de cette image en utilisant la méthode que nous avons discutée, où je vais créer deux versions indépendantes de la transcription pour ensuite les comparer et identifier les meilleures sections ou corriger les erreurs potentielles.

Version 1 de la transcription: *Cependant, pour y parvenir dans toute la mesure de son talent, il est impératif que l'AQRP dispose des fonds nécessaires à sa survie, à sa consolidation et à son développement. Nous misons en cela beaucoup d'espoir sur les demandes que nous avons adressées au ministère de la Santé et des Services sociaux pour obtenir une majoration de notre financement et pour nous aider à concrétiser nos offres de collaboration intersectorielle présentées au ministère de l'Emploi et de la Solidarité sociale et au ministère des Affaires municipales et de la Métropole.*

Version 2 de la transcription: *Cependant, pour atteindre pleinement notre potentiel, il est essentiel que l'AQRP dispose des fonds nécessaires à sa survie, à sa consolidation et à son développement. Nous plaçons beaucoup d'espoir dans les demandes que nous avons soumises au ministère de la Santé et des Services sociaux pour augmenter notre financement et pour nous aider à mettre en œuvre nos initiatives de collaboration intersectorielle proposées au ministère de l'Emploi et de la Solidarité sociale ainsi qu'au ministère des Affaires municipales et de la Métropole.*

Comparaison et correction: Les deux versions sont assez similaires, avec des variations mineures dans le choix des mots. La première version semble plus fidèle à l'image originale en termes de formulation et de terminologie spécifique utilisée dans le contexte organisationnel. La deuxième version simplifie légèrement certains termes, mais les deux conservent le sens général du message. Si vous avez une préférence pour l'une des versions ou si vous avez besoin de modifications supplémentaires, n'hésitez pas à le signaler.

L'Auteur : Merci beaucoup. Je crois avoir décelé une des causes de ce problème. Pour ma part je cherche juste la transcription exacte donc celle qui selon ton vocabulaire sera "fidèle à l'image originale en termes de formulation et de terminologie spécifique utilisée". Alors que d'après ce que je comprends, toi tu as l'air de penser que je pourrais être intéressé par une version qui "simplifie légèrement certains termes, en conservant] le sens général du message. Or comme mon but est justement l'analyse lexicale de ces textes, j'ai besoin de la transcription fidèle au mot près.



SM-OCR : Merci pour cette clarification, c'est très utile. Comprendre votre objectif spécifique d'obtenir une transcription textuelle fidèle et précise au mot près pour une analyse lexicale m'aide à ajuster ma réponse en conséquence. À l'avenir, je m'assurerai de privilégier la précision textuelle exacte par rapport à la paraphrase ou à la simplification, afin de répondre précisément à vos besoins d'analyse. Si vous avez d'autres textes à transcrire ou d'autres demandes, n'hésitez pas à les envoyer, et je veillerai à maintenir ce niveau de précision dans les transcriptions.

L'Auteur : Merci beaucoup. C'est parfait. Est-ce utile selon toi de maintenir la méthode avec double transcription, ou bien peut-on s'en passer ?



SM-OCR : Étant donné votre objectif spécifique d'obtenir des transcriptions textuellement fidèles pour une analyse lexicale, il pourrait être plus efficace de se concentrer sur une seule transcription précise plutôt que sur la création de deux versions. La méthode de double transcription peut introduire des variations inutiles qui pourraient compliquer l'analyse plutôt que de la faciliter. Je recommanderais de procéder avec une transcription unique et soigneusement vérifiée pour chaque image que vous fournissez, en mettant l'accent sur la fidélité au texte original. Cela garantira que les données que vous recevez sont aussi précises que possible sans nécessiter une étape supplémentaire de comparaison. Si vous êtes d'accord avec cette approche, nous pouvons continuer ainsi pour les prochaines transcriptions. Je m'assurerai de prêter une attention particulière à la précision et à l'intégrité du texte original dans chaque réponse.

L'Auteur : Ok parfait. On fait comme cela pour la suite.



SM-OCR : Très bien, nous procéderons de cette manière pour toutes vos futures demandes de transcription. N'hésitez pas à télécharger vos prochaines images ou à me faire part de tout autre besoin. Je suis ici pour vous aider avec la plus grande précision possible. Bonne continuation dans votre analyse lexicale ! (...)



Image générée par l'outil Dall-e à partir du titre de cette chronique

Commentaire et discussion : attention aux hallucinations mais aussi à l'illusion réflexive

Cet exemple de dialogue avec un outil fondé sur ChatGPT montre comment le mécanisme des hallucinations peut se glisser de manière insidieuse dans différentes interactions avec les IAG. On notera en particulier que lorsque que les outils classiques d'OCR rencontrent des difficultés de reconnaissance, ils l'indiquent alors qu'ici le GPT remplit le vide par une hallucination probabiliste, comme il le fait d'ailleurs s'il ne trouve pas les références biblio ou les informations qui lui manquent. Notons en particulier le brio avec lequel est créée une seconde version légèrement différente du second texte de façon à pouvoir la comparer. Tel semble être le mécanisme général des hallucinations : satisfaire toute que toute la demande de l'utilisateur même grâce à une réalité artificielle ou alternative. Pour finir, faisons aussi attention au principe de l'illusion réflexive : dans ce dialogue, ce GPT n'a pas réfléchi avec nous à la façon d'améliorer ces performances d'OCR. Il n'a fait que générer les réponses jugées le plus probables à nos formulations de questions grâce à ces jeux de données et son entraînement.

Références bibliographiques

Las Vergnas O. (2023). Les androïdes rêvent-ils de références fictives ? Un dialogue bibliographique avec ChatGPT3.5.

Alliage : Culture - Science - Technique, 2023 (83). <https://hal.science/hal-04192213>

Las Vergnas O. & Rio-Jouet A. (2023). Une IA peut-elle remplacer un professeur ? Un exemple de dialogue avec Bard-Google, entre hallucinations et illusions réflexives. *L'Esprit d'Archimède, ALEA & FST-Université de Lille*, 2023 (10), pp.27-34. <https://hal.science/hal-04273617>

Pujade S. (2024). ChatGPT - Studymate OCR and Coding Expert. <https://chatgpt.com/g/g-Aptbt5Oml-studymate-ocr-and-coding-expert>