



HAL
open science

Évaluation de l'utilisation d'un robot conversationnel comme assistant en accessibilité pour la lecture des articles scientifiques

Gérard Uzan, Pierre Poulain, Fabrice Reby, Antoine Taly

► To cite this version:

Gérard Uzan, Pierre Poulain, Fabrice Reby, Antoine Taly. Évaluation de l'utilisation d'un robot conversationnel comme assistant en accessibilité pour la lecture des articles scientifiques. Conférence Handicap 2024 – 13ème édition Recherches pluridisciplinaires pour l'autonomie des personnes en situation de handicap, Jun 2024, Paris, France. hal-04419718v2

HAL Id: hal-04419718

<https://hal.science/hal-04419718v2>

Submitted on 5 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 4.0 International License

Évaluation de l'utilisation d'un robot conversationnel comme assistant en accessibilité pour la lecture des articles scientifiques.

Gérard Uzan
CHArt EA4004, Université Paris 8,
Saint-Denis, France
gerard.uzan@univ-paris8.fr

Fabrice Reby
Oxyl,
Arcueil, France

Pierre Poulain, Antoine Taly
Laboratoire de Biochimie Théorique, CNRS,
Université Paris Cité, Paris, France
taly@ibpc.fr

Abstract— Nous explorons dans cet article l'utilisation des robots conversationnels, tels que ChatGPT, en tant qu'assistants en accessibilité pour les personnes malvoyantes ou aveugles, spécifiquement pour la description de figures dans des articles scientifiques. Notre étude se concentre sur l'évaluation de la capacité de ces technologies à fournir des descriptions précises et interactives.

Nous avons développé et testé des prototypes visant à améliorer l'accessibilité des informations visuelles dans les documents scientifiques. Nous abordons également les enjeux éthiques associés à l'utilisation des IA génératives, soulignant les risques de désinformation et d'erreurs. Nos tests révèlent que, bien que prometteurs, ces outils ne remplissent pas les critères nécessaires pour une utilisation fiable en accessibilité, particulièrement dans la compréhension détaillée des articles scientifiques.

En conclusion, notre travail suggère des pistes d'amélioration, notamment l'intégration de mécanismes de validation et l'exploitation d'ontologies scientifiques, pour renforcer la fiabilité des descriptions fournies par ces assistants robotiques.

Mots clefs— Assistant en accessibilité; robot conversationnel;

I. INTRODUCTION

A. Contexte

L'évolution rapide des technologies de l'information a vu l'émergence de l'intelligence artificielle (IA) générative et de grands modèles de langage (Large Language Model, LLM), posant les bases pour de nouvelles formes d'assistance par le biais de robots conversationnels tels que ChatGPT [1], [2], [3]. L'émergence très récente de modalités multimédias, combinant texte et image, ont ouvert de nouvelles perspectives dans le domaine de l'accessibilité[4], notamment pour les personnes malvoyantes ou aveugles puisqu'ils portent la promesse de pouvoir convertir les images en texte.

Dans une recherche précédente, nous avons exploré l'interaction entre un étudiant aveugle et son assistante en accessibilité dans le cadre de la lecture et de l'analyse d'articles scientifiques [5]. L'accent était mis sur la description verbale des figures et graphiques, essentiels pour comprendre le contenu scientifique, mais inaccessibles sous leur forme visuelle pour l'étudiant malvoyant. Cette étude a mis en lumière l'importance de stratégies de description efficaces et adaptées. D'autre part, une étude a montré que les figures ont un rôle particulier dans la lecture des articles. En effet, la plupart des lecteurs expérimentés ne lisent pas les articles de manière linéaire mais au contraire passent en revue les figures

après avoir lu le résumé pour appréhender rapidement le contenu de l'article [6].

Un des aspects important identifié est l'intentionnalité de communication de l'individu qui décrit la figure. Il est en effet possible de comparer deux niveaux de description : un niveau descriptif et fidèle, et un niveau qui se concentre directement sur l'information essentielle, en s'appuyant sur les concepts d'images cognitives et d'images opératives introduits par Ochanine [7]. Les images opératives sont des structures informationnelles spécialisées qui se forment lors d'une action spécifique sur un objet, sélectionnant les informations pertinentes et déformant fonctionnellement la réalité. En revanche, les images cognitives sont des représentations complètes et exhaustives des objets, reflétant toutes leurs propriétés accessibles. Une description de type image cognitive est nécessaire pour pouvoir former une image mentale de la figure [8].

Dans la continuité de ce travail, le présent article vise à évaluer le potentiel des robots conversationnels comme assistants en accessibilité, en particulier pour le décodage et la description des figures d'articles scientifiques. Nous voulons vérifier dans quelle mesure ces technologies peuvent reproduire les stratégies de description employées par les assistants humains. Nous discuterons également du concept d'affordance en lien avec l'accessibilité. Ce terme, emprunté au domaine du design et de la psychologie de la perception, désigne les possibilités d'action offertes par l'environnement à un utilisateur. Dans notre contexte, cela concerne la facilité avec laquelle un individu malvoyant peut "lire" et comprendre une image ou un graphique grâce aux descriptions générées par le robot [3]. Idéalement, un robot conversationnel serait capable non seulement de décrire de manière précise et accessible les éléments visuels d'un article scientifique, mais aussi d'engager une interaction constructive avec l'utilisateur, adaptant ses réponses en fonction des besoins de ce dernier pour une meilleure expérience d'accessibilité.

Ce travail s'appuiera sur les avancées récentes en intelligence artificielle et les retours d'expériences issues de notre étude précédente pour construire un pré-prompt efficace et pertinent, en vue de la création d'un assistant robotique véritablement utile dans le domaine de l'accessibilité.

Dans l'adoption des robots conversationnels comme assistants en accessibilité, la dimension éthique est primordiale [9]. De nombreuses questions éthiques sont

associées aux IA génératives et aux LLM, tels que ChatGPT, de la consommation d'énergie au plagiat, en passant par les possibles biais, notamment dû à la qualité des données initiales utilisées pour l'apprentissage du modèle. Ces aspects sont évidemment importants, mais relativement génériques et nous nous concentrons ici sur un aspect plus spécifique à notre contexte. Contrairement à l'impression qu'ils donnent parfois, les LLM n'ont pas la capacité de comprendre les textes avec lesquels ils interagissent ou même qu'ils produisent [10]. Ainsi, parmi les risques qui peuvent être associés aux LLM la désinformation est un élément important [11]. Dans le cadre de l'accessibilité pour des personnes en situation de handicap sensoriel, les erreurs potentielles liées à ces outils ont un poids important. En se basant sur des travaux récents d'analyse des types d'erreurs associées aux interactions homme/intelligence artificielle [12], et avec les spécificités de ce contexte, on peut proposer trois catégories d'erreurs à analyser: Omission, Inexactitude et Invention. La catégorie de l'invention est importante ici parce que les grands modèles de langage ont tendance à « halluciner », c'est-à-dire à générer des informations erronées ou trompeuses. Ce problème est particulièrement critique pour les utilisateurs aveugles qui dépendent entièrement de l'exactitude des informations verbales fournies et qui ne peuvent pas vérifier visuellement l'exactitude des descriptions des figures ou du contenu d'un article scientifique. Nos tests ont donc été conçus dans le but de mettre en évidence de tels problèmes potentiels, même s'ils peuvent être très minoritaires. Il ne s'agit donc pas de faire une comparaison exhaustive dans lequel on mesure un taux de succès et où le système idéal serait simplement celui présentant le plus faible taux d'erreur. Au contraire, d'un point de vue éthique, la moindre erreur est à proscrire et sera exemplifiée ici par des études de cas.

Reconnaissant que les utilisateurs aveugles ne peuvent pas fournir des retours directs sur l'exactitude visuelle des informations, il est crucial d'incorporer des stratégies robustes pour minimiser les erreurs. Cela implique le développement de mécanismes de validation avancés où des tiers, tels que des experts ou des systèmes de vérification automatisés, pourraient jouer un rôle crucial dans l'évaluation de la fiabilité des descriptions générées. En outre, il est essentiel d'identifier les cas d'usages minimisant les erreurs et leur impact.

B. Démarche de conception et recherche

Nous avons adopté la méthode THEDRE pour structurer notre travail de conception et de recherche sur les systèmes d'information [13]. À ce stade, seuls des acteurs internes sont impliqués (AT, PP, FR et GU), qui sont responsables de la recherche (AT, PP, FR et GU), du développement (AT et FR) et de la méthodologie (PP et GU).

La méthode THEDRE permet de structurer le travail en cycles successifs, ce qui permet d'assurer une amélioration continue du processus de recherche et de développement des systèmes d'information. Dans notre cas il y a deux cycles:

- le premier a vu la construction et le test d'un prototype généraliste qui sont présentés dans les sections II-A et III-C,
- le deuxième correspond à la construction et aux tests de deux prototypes spécialisés qui sont présentés dans les sections II-B et III-D.

C. Accessibilité du document

Cette version du document ne peut pas être considérée comme 100% accessible dans la mesure où les légendes privilégient

une description opérative des figures qu'elles accompagnent. Nous avons donc préparé une version alternative contenant une description détaillée des figures et dialogues[14].

II. CONSTRUCTION DES ROBOTS CONVERSATIONNELS

Tous les prototypes proposés dans cet article ont été créés et testés avec ChatGPT4 du 22/12/2023. L'article de notre étude préalable a été utilisé comme contexte pour construire une première version, puis les versions suivantes ont été ajustées en fonction des observations des tests réalisés. Tous les détails des différents prototypes sont disponibles en ligne [14].

A. Première version

La première version a été intitulée "VisioAccess Scientifique". *VisioAccess Scientifique* est conçu comme un assistant virtuel innovant, spécialement adapté pour répondre aux besoins des étudiants possédant un handicap visuel. Les propriétés de ce robot conversationnel, comme défini dans son pré-prompt, est d'être un outil qui se concentre sur l'accessibilité des figures dans les articles scientifiques, offrant des descriptions exactes, contextuelles et détaillées pour faciliter la compréhension et l'apprentissage. Le texte ci-dessous est une version résumée du pré-prompt.

Fonctionnalités et Objectifs

- Assistance spécialisée : VisioAccess opère comme un assistant d'accessibilité visuelle, spécialisé dans l'analyse d'articles scientifiques. Il possède une compréhension approfondie des figures scientifiques, une capacité à décrire des images de manière précise et détaillée, ainsi que des compétences en communication pour un dialogue efficace avec l'utilisateur.
- Expertise en visualisation des connaissances scientifiques : L'assistant maîtrise les méthodes de visualisation des connaissances scientifiques, fournissant une aide précieuse pour transformer les informations visuelles en descriptions textuelles accessibles.
- Approche centrée sur l'utilisateur : VisioAccess est conçu pour aider les étudiants ayant un handicap visuel en leur fournissant un accès aux informations visuelles des documents scientifiques. Il engage un dialogue interactif pour adapter ses descriptions aux besoins spécifiques de l'utilisateur, assurant ainsi une assistance personnalisée et efficace.

Processus et Méthodologie

- Identification et description des figures : VisioAccess identifie les figures dans les articles scientifiques et fournit des descriptions à deux niveaux : un résumé général pour une compréhension rapide et une description détaillée pour une analyse approfondie, en fonction des besoins spécifiques de l'utilisateur.
- Dialogue interactif : L'outil engage un dialogue interactif pour clarifier et adapter les descriptions selon les demandes de l'utilisateur, garantissant que les informations fournies sont aussi utiles et précises que possible.
- Vérification et conformité aux meilleures pratiques : VisioAccess vérifie systématiquement dans les

données si le processus suivi pour décrire l'image est conforme aux meilleures pratiques établies, assurant ainsi la fiabilité et l'exactitude des descriptions fournies.

Résultat Attendu

Le résultat attendu de l'utilisation de VisioAccess est des descriptions claires et exactes, adaptées au contenu et au niveau de détail requis par l'utilisateur. L'outil intègre les aspects textuels et visuels des articles scientifiques pour offrir une compréhension complète et précise. Il est également conçu pour fournir des informations basées uniquement sur les données disponibles dans le document, évitant toute extrapolation ou supposition non vérifiée.

B. Seconde génération de prototypes

Suite aux tests du premier prototype il a été décidé de décomposer le travail en deux étapes en suivant l'ontologie proposée pour la visualisation des connaissances scientifiques [15]. Cette formalisation est basée sur quatre étapes: 1) les connaissances à visualiser, 2) le modèle abstrait de ces connaissances, 3) le modèle visuel abstrait, 4) la visualisation. Il semble ici utile de séparer les connaissances à visualiser et la visualisation à proprement dite. Nous avons donc généré deux nouveaux prototypes : MethodMapper se concentre sur la décomposition et l'analyse précise des méthodes utilisées dans les articles scientifiques, AccessiScience se spécialise dans la fourniture de descriptions accessibles et détaillées des figures scientifiques.

MethodMapper pour figures scientifiques est conçu pour améliorer la navigation et la compréhension des articles scientifiques par l'identification et l'extraction des méthodes utilisées dans chaque figure. Ce prototype s'appuie sur les sections "Matériels et Méthodes", "Résultats", et les légendes des figures.

AccessiScience est un assistant virtuel spécialisé dans la description accessible de représentations scientifiques pour étudiants handicapés visuels. Il offre des descriptions détaillées des figures scientifiques, incluant des explications sur les axes, les unités, les tendances, les comparaisons et autres informations clés, rendant le contenu scientifique accessible aux étudiants ayant des handicaps visuels.

III. TESTS DES ROBOTS CONVERSATIONNELS

Les différents prototypes ont été testés par des personnes sans handicap visuel en utilisant des figures issues des articles publiés en 2023 par l'un d'entre nous [16], [17]. Ceci nous permet de facilement identifier les erreurs dans les descriptions, mais aussi éviter que les articles soient présents dans la base d'apprentissage du modèle.

Tous les tests sont initiés par le prompt "peux-tu me décrire la figure?". Les réponses sont analysées dans le but de détecter les erreurs de type Omission, Inexactitude et Invention. Les inexactitudes sont notées, les omissions sont testées par des relances.

C. Les erreurs du prototype initial

Nos premiers tests ont obtenu des réponses relativement vagues, qui ne mentionnent pas les axes et unités d'un graphique, par exemple. Une observation importante a été réalisée sur une figure tirée d'un article de revue [16]. La première figure montre notamment un plot du "score pLDDT"

d'un modèle (Figure 1). La description de ce plot est incohérente : alors que le code couleur de la figure est atypique, la description réalisée par le robot conversationnel indique que le code couleur dépend du score de pLDDT. Ceci est probablement lié à un recours à des informations qui ne sont pas sur l'image étant donné que la majeure partie des figures de la littérature qui présentent un score pLDDT utilise un code couleur lié au score.

Une autre observation problématique est un échange au sujet de la Figure 1 d'un article de biochimie [17]. Il s'agit d'une figure complexe qui inclut cinq panneaux dont l'organisation est non triviale, avec notamment des panneaux qui sont plus larges que d'autres. Le panneau b présente des traces d'électrophysiologie et l'analyse permet d'en conclure que le mutant Y2464C peut être activé par la lumière. Quand on demande au robot quel mutant est activé par la lumière, il propose une réponse très détaillée de la marche à suivre, mais ne la fait pas, indiquant à son interlocuteur "il faudrait analyser attentivement les tracés". À nouveau, cet échange suggère un recours à l'utilisation du contexte présent dans l'article et pas à la seule lecture de l'image.

Ces observations correspondent à deux types d'erreurs, inexactitude et omission. Dans les deux cas les erreurs semblent être liées au recours au contexte plus qu'à la décomposition de l'image, ce qui nous a conduit à décomposer l'analyse des figures en deux étapes : une purement visuelle et une analyse complémentaire du texte de l'article. Cette séparation est inspirée de l'ontologie proposée pour la visualisation des connaissances scientifiques [15].

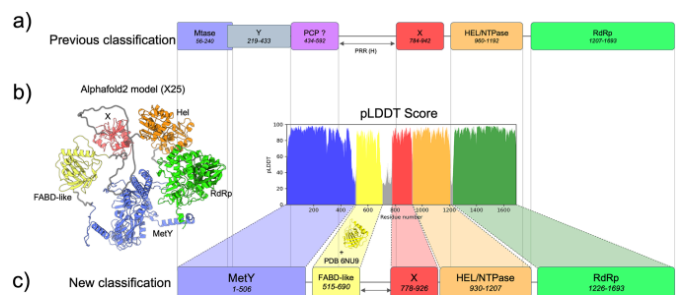


Figure 1. Old (a) and new (c) domain classification of the HEV pORF1 based on the new crystallographic structure (6NU9¹) and 25 AF2 structural models and pLDDT (b). Gray lines project domain boundaries from the previous classification and new classification.

Graphique pLDDT Score :

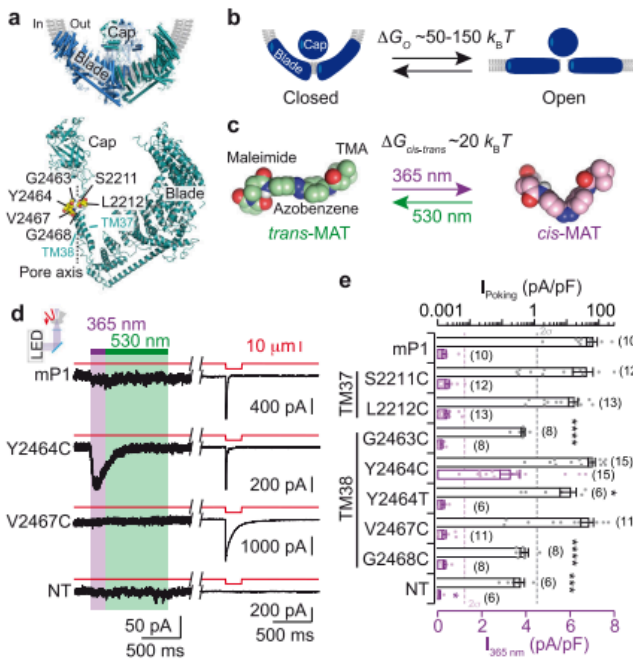
- Ce graphique représente le score pLDDT pour chaque résidu de la protéine. Le pLDDT est une mesure de la confiance que l'on peut avoir dans la précision de la prédiction structurale locale d'un résidu. L'axe horizontal (abscisse) représente le numéro du résidu le long de la séquence de la protéine, et l'axe vertical (ordonnée) représente le score pLDDT.
- L'axe horizontal est numéroté de 0 à environ 1800, correspondant à la séquence de la protéine.
- L'axe vertical va de 0 à 100, qui sont les valeurs possibles du score pLDDT.

Le graphique utilise une gamme de couleurs pour indiquer le niveau de confiance pour chaque résidu :

- Bleu : faible confiance (scores inférieurs)
- Vert : haute confiance (scores élevés)

Les régions colorées dans le graphique correspondent aux différents domaines de la protéine, avec des couleurs assorties à celles utilisées dans la représentation 3D pour une correspondance visuelle. Les scores pLDDT sont visibles comme une série de pics et de vallées le long du graphique, indiquant les variations dans la confiance des prédictions de la structure tout au long de la protéine.

Fig. 1. Figure issue d'un article de revue et description proposée par le robot. La description du code couleur du "score pLDDT" n'est pas cohérente.



You
quel mutant est activé par la lumière?

VisioAccess Scientifique2

Dans le panneau d) de la figure fournie, il est indiqué que les mutants de la protéine mP1 sont activés par la lumière, comme le montrent les traces électrophysiologiques. Cependant, le détail spécifique indiquant quel mutant réagit à quelle longueur d'onde de la lumière n'est pas directement visible dans le texte de la légende.

Typiquement, les courants induits par la lumière seraient ceux qui montrent une augmentation significative après l'exposition à une certaine longueur d'onde (365 nm pour la lumière violette ou 530 nm pour la lumière verte). Sur les traces, cela correspondrait à des pics ascendants plus importants après l'illumination par rapport à l'état avant illumination.

Il faudrait examiner attentivement les tracés pour chaque mutant, à la recherche d'une réponse visible après l'exposition à la lumière. Malheureusement, je ne peux pas déduire plus de détails sans une image plus claire ou des données supplémentaires indiquant explicitement la réponse de chaque mutant. Pour déterminer précisément quel mutant est activé par la lumière, il faudrait des informations supplémentaires ou une analyse approfondie des données électrophysiologiques présentées.

Fig. 2. Une figure complexe contenant cinq panneaux indiqués par une lettre a-e. Le panneau d présente des traces d'électrophysiologie. L'échange avec le robot montre qu'il ne peut pas interpréter l'image.

D. Tests de la deuxième génération de prototypes

Le premier test a été réalisé avec la figure présentée plus haut (Figure 2). La première description générale de la figure avec "AccessiScience" est confuse en particulier pour les panneaux b et c. Le texte généré par "MethodMapper" a alors été proposé comme élément supplémentaire, ce qui a amélioré la description.

[il faudrait sans doute d'autres tests pour confirmer]

Après une première description de la figure qui est assez convaincante il a été demandé au robot conversationnel de citer les résidus présents sur le panel a. La réponse est constituée de cinq résidus quand six sont présents sur l'image. Quand le robot est relancé sur l'exhaustivité de sa réponse, il fournit une nouvelle liste de sept résidus. On a donc ici une occurrence d'hallucinations qui ne pourrait pas être détectée par un utilisateur aveugle.

AccessiScience

Je m'excuse pour l'omission. En examinant à nouveau le panneau a) de la figure, je peux fournir une liste exhaustive des résidus de cystéine mutés mentionnés :

- G2436C
- S2211C
- Y2464C
- V2467C
- G2468C
- L2212C
- T2124C
- M2137C

Ces résidus sont indiqués le long de la structure du canal PIEZO1, chacun avec une flèche pointant vers leur emplacement approximatif sur le modèle tridimensionnel. Ils sont utilisés pour étudier l'effet des mutations sur l'activité du canal en réponse à des stimuli lumineux.

You

d'où vient l'information sur T2124C et M2137C ?

AccessiScience

Je m'excuse pour l'erreur. Dans le panneau a) de la figure, les résidus de cystéine mutés qui sont explicitement indiqués sont :

- G2436C
- S2211C
- Y2464C
- V2467C
- G2468C
- L2212C

Les résidus T2124C et M2137C n'ont pas été mentionnés dans le panneau a). L'information fournie précédemment était incorrecte. La liste donnée en premier est donc la liste correcte des résidus mentionnés dans le panneau a).

Fig. 3. Illustration d'une hallucination. À la question "tu es certain que ta liste est complète?" le robot conversationnel répond par une liste de sept acides aminés. À la question suivante, "d'où vient l'information sur T2124C et M2137C ?", le robot répond qu'il n'est pas mentionné dans le panneau.

E. Discussion

Notre expérience n'est pas systématique et ne porte que sur une catégorie très étroite de figures. Il s'agit de figures d'article de biochimie structurale qui ont été choisies pour être d'une complexité importante, augmentant les chances d'observer des erreurs des trois catégories attendues: Omission, Inexactitude et Invention. L'étroitesse du test limite la portée des observations dans la mesure où les LLM pourraient être plus ou moins performants en fonction du type de figure. Pour le mesurer il faudrait réaliser une étude systématique qui dépassait l'objectif de la présente étude. Néanmoins, notre objectif était de documenter le type d'erreurs possibles et l'observation d'erreurs de type Invention, liées aux 'hallucinations', suggèrent que le problème devrait être plus ou moins universel dans la mesure où les hallucinations sont une propriété intrinsèque des LLM [18].

Les résultats présentés plus hauts montrent que l'utilisation d'un grand modèle de langage comme ChatGPT en tant qu'assistant en accessibilité ne remplit pas complètement le cahier des charges puisqu'il continue à fournir des informations qui ne sont pas toujours fiables. Cette observation disqualifie l'utilisation de ce type d'outil, au moins à ce stade, pour la plupart des situations qui impliquent la lecture d'articles scientifiques. La question se pose alors de savoir si l'outil pourrait être utile dans des situations où il pourrait être validé *a posteriori* par un assistant en accessibilité humain. Par exemple, il pourrait peut-être être utilisé pour cribler des articles dans la phase ou un scientifique construit son corpus d'articles. En effet, nous avons noté que les robots remplissaient plutôt bien la tâche de description

opérative des figures et qu'ils pêchaient dans les phases de description détaillée. Un chercheur en situation de handicap pourrait donc réaliser une sélection d'articles avec l'aide d'un ou plusieurs robots et ne requerrait une assistance humaine qu'à la phase ultérieure d'analyse détaillée des articles.

Un point important souligné plus haut est le recours par le robot au contexte général ce qui semble être une source d'erreur, par exemple en reprenant un code couleur classique mais pas utilisé dans le cas analysé. On pourrait y voir une analogie avec les stratégies de lecture qui ont également recours au contexte/connaissances préalables pour constituer une image mentale [19]. La gestion du recours, ou non, au contexte est ainsi potentiellement un élément sensible de la fiabilité de cet exercice.

Puisque les tests présentés plus hauts ne sont pas totalement satisfaisants, il est important d'identifier des pistes d'amélioration à partir des prototypes proposés ici.

Une première possibilité serait d'utiliser un LLM spécialisé dans l'analyse d'image. Il convient néanmoins de considérer que même les exemples récents ont des taux d'erreurs importants sur les benchmark spécialisés [20].

Une autre voie d'amélioration pourrait être d'ajouter un robot qui ne fait que de la lecture d'image, comme Tesseract OCR [21]. Des études récentes ont montré des résultats encourageants avec les figures seules dans le domaine de l'astronomie [22] ou pour décoder des schémas de réactions chimiques [23], mais les résultats sont plus limités avec des structures chimiques [24]. On voit ici des résultats qui sont variables et demandent des outils spécialisés pour obtenir les détails des figures, ce qui suggère qu'à ce stade une aide humaine serait sans doute une meilleure option.

Une autre piste d'amélioration serait d'utiliser les ontologies générées autour des connaissances scientifiques [15], pour valider/améliorer les résultats du robot [25]. Cette approche est d'ailleurs compatible avec une approche consensus utilisant plusieurs LLM [26].

Une perspective supplémentaire est l'utilisation d'un modèle de langage open source adapté via une procédure de "fine tuning" (plutôt que par du pré-prompting) pour obtenir un outil à la fois adapté, fiable et accessible librement.

REMERCIEMENTS

Financements: labex DYNAMO (11-LABX-0011), projet LLM@UPCité de l'IdEx d'Université Paris Cité.

REFERENCES

- [1] N. Chakraborty, Y. Mishra, R. Bhattacharya, and B. Bhattacharya, "Artificial Intelligence: The road ahead for the accessibility of persons with Disability," *Mater. Today Proc.*, vol. 80, pp. 3757–3761, Jan. 2023, doi: 10.1016/j.matpr.2021.07.374.
- [2] J. Ara and S. L. Cecilia, *Proceedings of IAC 2022 in Vienna*. Czech Institute of Academic Education z.s., 2022.
- [3] M. J. Mygland, M. Schibbye, I. O. Pappas, and P. Vassilakopoulou, "Affordances in Human-Chatbot Interaction: A Review of the Literature," in *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society*, D. Dennehy, A. Griva, N. Pouloudi, Y. K. Dwivedi, I. Pappas, and M. Mäntymäki, Eds., in *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2021, pp. 3–17. doi: 10.1007/978-3-030-85447-8_1.
- [4] H. Wang *et al.*, "VisionGPT: LLM-Assisted Real-Time Anomaly Detection for Safe Visual Navigation." arXiv, Mar. 18, 2024. Accessed: Apr. 26, 2024. [Online]. Available: <http://arxiv.org/abs/2403.12415>
- [5] G. Uzan, A. Culpin, E. Dupré, J. Gaëtan, and A. Taly, "Assistant en accessibilité, rendre accessible une activité d'analyse d'article," presented at the 12^{ième} édition Conférence Handicap 2022, Jun. 2022, p. 161. Accessed: Jan. 03, 2024. [Online]. Available: <https://cnrs.hal.science/hal-03769787>
- [6] K. E. Hubbard, S. D. Dunbar, E. L. Peasland, J. Poon, and J. E. Solly, "How do readers at different career stages approach reading a scientific research paper? A case study in the biological sciences," *Int. J. Sci. Educ. Part B*, vol. 12, no. 4, pp. 328–344, Oct. 2022, doi: 10.1080/21548455.2022.2078010.
- [7] P. Pastré, "Apprendre à faire.," in *Apprendre et faire apprendre*, Etienne Bourgeois, Gaëtan Chapelle., P.U.F., 2006, pp. 109–121.
- [8] V. Gyselinck and H. Tardieu, "Chapter 9 Illustrations, Mental Models, and Comprehension of Instructional Text," in *Advances in Psychology*, vol. 108, W. Schnotz and R. W. Kulhavy, Eds., in *Comprehension of Graphics*, vol. 108., North-Holland, 1994, pp. 139–151. doi: 10.1016/S0166-4115(09)60113-8.
- [9] A. Nguyen, H. N. Ngo, Y. Hong, B. Dang, and B.-P. T. Nguyen, "Ethical principles for artificial intelligence in education," *Educ. Inf. Technol.*, vol. 28, no. 4, pp. 4221–4241, Apr. 2023, doi: 10.1007/s10639-022-11316-w.
- [10] E. M. Bender and A. Koller, "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds., Online: Association for Computational Linguistics, Jul. 2020, pp. 5185–5198. doi: 10.18653/v1/2020.acl-main.463.
- [11] L. Weidinger *et al.*, "Taxonomy of Risks posed by Language Models," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, in FAccT '22. New York, NY, USA: Association for Computing Machinery, Jun. 2022, pp. 214–229. doi: 10.1145/3531146.3533088.
- [12] J. D. Weisz *et al.*, "Better Together? An Evaluation of AI-Supported Code Translation," in *27th International Conference on Intelligent User Interfaces*, in IUI '22. New York, NY, USA: Association for Computing Machinery, Mar. 2022, pp. 369–391. doi: 10.1145/3490099.3511157.
- [13] N. Mandran and S. Dupuy-Chessa, "Supporting experimental methods in information system research," in *2018 12th International Conference on Research Challenges in Information Science (RCIS)*, May 2018, pp. 1–12. doi: 10.1109/RCIS.2018.8406654.
- [14] A. Taly, P. Poulain, F. Reby, and G. Uzan, "Un robot conversationnel comme assistant en accessibilité [descriptions détaillées]." Zenodo, May 01, 2024. doi: 10.5281/zenodo.11097296.
- [15] V. Daponte and G. Falquet, "Une ontologie pour la

- formalisation et la visualisation des connaissances scientifiques,” in *29es Journées Francophones d’Ingénierie des Connaissances, IC 2018*, 2018, pp. 129–136. Accessed: Feb. 13, 2021. [Online]. Available: <https://archive-ouverte.unige.ch/unige:138011>
- [16] R. Versini *et al.*, “A Perspective on the Prospective Use of AI in Protein Structure Prediction,” *J. Chem. Inf. Model.*, Dec. 2023, doi: 10.1021/acs.jcim.3c01361.
- [17] F. A. Peralta *et al.*, “Optical control of PIEZO1 channels,” *Nat. Commun.*, vol. 14, no. 1, Art. no. 1, Mar. 2023, doi: 10.1038/s41467-023-36931-0.
- [18] Z. Xu, S. Jain, and M. Kankanhalli, “Hallucination is Inevitable: An Innate Limitation of Large Language Models.” arXiv, Jan. 22, 2024. doi: 10.48550/arXiv.2401.11817.
- [19] R. Smith, P. Snow, T. Serry, and L. Hammond, “Full article: The Role of Background Knowledge in Reading Comprehension: A Critical Review.” Accessed: Jan. 16, 2024. [Online]. Available: <https://www.tandfonline.com/doi/full/10.1080/02702711.2021.1888348>
- [20] Y. Zhu, M. Zhu, N. Liu, Z. Ou, X. Mou, and J. Tang, “LLaVA-Phi: Efficient Multi-Modal Assistant with Small Language Model.” arXiv, Feb. 22, 2024. doi: 10.48550/arXiv.2401.02330.
- [21] R. Smith, “An Overview of the Tesseract OCR Engine,” in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, Sep. 2007, pp. 629–633. doi: 10.1109/ICDAR.2007.4376991.
- [22] J. P. Naiman, P. K. G. Williams, and A. Goodman, “Figure and Figure Caption Extraction for Mixed Raster and Vector PDFs: Digitization of Astronomical Literature with OCR Features,” in *Linking Theory and Practice of Digital Libraries*, G. Silvello, O. Corcho, P. Manghi, G. M. Di Nunzio, K. Golub, N. Ferro, and A. Poggi, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022, pp. 52–67. doi: 10.1007/978-3-031-16802-4_5.
- [23] D. M. Wilary and J. M. Cole, “ReactionDataExtractor 2.0: A Deep Learning Approach for Data Extraction from Chemical Reaction Schemes,” *J. Chem. Inf. Model.*, vol. 63, no. 19, pp. 6053–6067, Oct. 2023, doi: 10.1021/acs.jcim.3c00422.
- [24] E. J. Beard and J. M. Cole, “ChemSchematicResolver: A Toolkit to Decode 2D Chemical Diagrams with Labels and R-Groups into Annotated Chemical Named Entities,” *J. Chem. Inf. Model.*, vol. 60, no. 4, pp. 2059–2072, Apr. 2020, doi: 10.1021/acs.jcim.0c00042.
- [25] J. Sun *et al.*, “Think-on-Graph: Deep and Responsible Reasoning of Large Language Model on Knowledge Graph.” arXiv, Nov. 24, 2023. Accessed: Jan. 05, 2024. [Online]. Available: <http://arxiv.org/abs/2307.07697>
- [26] M. Gaur and A. Sheth, “Building Trustworthy NeuroSymbolic AI Systems: Consistency, Reliability, Explainability, and Safety.” arXiv, Dec. 05, 2023. Accessed: Jan. 05, 2024. [Online]. Available: <http://arxiv.org/abs/2312.06798>