ACTA MÉDICA PORTUGUESA
A Revista Científica da Ordem dos Médicos

Use of ChatGPT in HIV Infection Counselling and Literacy

Utilização do ChatGPT para Aconselhamento e Literacia sobre a Infeção pelo VIH

José GANICHO✉[1], Maria CARLOS[1], Gonçalo CRISTÓVÃO[1], Claudina CRUZ[1], Marta LEAL[1], Ana Raquel GARROTE[1], Diogo BENTO[1], Diana SEIXAS[1], Maria José MANATA[1], Fernando MALTEZ[1,2]
Acta Med Port 2025 Jun-Jul;38(6-7):385-392 ▪ https://doi.org/10.20344/amp.22805

ABSTRACT

Introduction: Artificial intelligence is an area that has been transforming various sectors of society. One of these tools is ChatGPT (chat generative pre-trained transformer), a large language model-based platform developed to understand language patterns and generate answers. The accessibility, answers in real-time, and privacy make ChatGPT a platform with enormous potential in health counseling. The aim of this study was to explore the potential of ChatGPT in health counseling and literacy regarding HIV infection and pre-exposure prophylaxis (PrEP).

Methods: A team of five experts in infectious diseases was consulted regarding the five most frequently asked questions by patients regarding initial HIV diagnosis and PrEP. The questions were submitted to the ChatGPT-4 version and the quality of the answers was reviewed based on a Likert scale from 1 to 5 (from 'strongly disagree' to 'strongly agree'). The evaluation of the expert's team followed a set of eight criteria: accuracy, clarity, relevance, depth, objectivity, conciseness, timeliness, and suitability for the target audience.

Results: The five most frequently asked questions by patients regarding HIV infection and the mean evaluation of the answers were: "What does it mean to have HIV? Will I die sooner?" (4.25), "Does HIV infection have a cure? Will I ever be able to stop taking medication?" (4.35), "Can I transmit my infection to other people? How to avoid it?" (4.03), "What are the adverse effects of the antiretroviral treatment?" (3.53), "Can I have unprotected sexual relationships? What are the risks?" (4.53). Regarding PrEP, the more frequently asked questions were: "I want to start pre-exposure prophylaxis. Is it safe?" (4.58), "What are the different regimens of pre-exposure prophylaxis?" (4.63), "After I start taking pre-exposure prophylaxis, do I have to take it forever?" (3.95), "What are the adverse effects of pre-exposure prophylaxis?" (4.63), "Does pre-exposure prophylaxis protect against other sexually transmitted infections or do I have to use condoms?" (4.60).

Conclusion: Only two answers received a rating below four on the 1 - 5 Likert scale, highlighting ChatGPT's potential as a promising tool for health counselling and literacy, provided ethical and legal concerns are effectively addressed.

Keywords: Artificial Intelligence; Health Literacy; HIV Infections/prevention & control; Pre-Exposure Prophylaxis

RESUMO

Introdução: A inteligência artificial é uma área que tem transformado todos os setores da nossa sociedade. Uma destas ferramentas é o ChatGPT (*chat generative pre-trained transformer*), uma plataforma baseada em *large language models*, desenvolvida para compreender padrões de linguagem e gerar respostas. As respostas em tempo real, a acessibilidade e a privacidade tornam o ChatGPT uma plataforma com um enorme potencial para o aconselhamento em saúde. Este estudo pretendeu explorar o potencial do ChatGPT no aconselhamento e literacia sobre a infeção por VIH e sobre a profilaxia de pré-exposição (PrEP).

Métodos: Uma equipa de 5 especialistas em doenças infeciosas foi consultada quanto às 5 questões mais frequentemente colocadas sobre o diagnóstico inaugural de infeção por VIH e sobre a PrEP. As questões foram colocadas à plataforma e a qualidade das respostas foram posteriormente avaliadas, numa escala de Likert de 1 a 5 (de 'discordo totalmente' a 'concordo totalmente') quanto à sua precisão, clareza, relevância, profundidade, objetividade, concisão, atualidade e adequação ao público-alvo.

Resultados: As perguntas mais frequentemente colocadas no âmbito do diagnóstico inaugural e a média das avaliações das respetivas respostas foram: "Qual o impacto da infeção por VIH na minha qualidade e esperança média de vida?" (4,25), "Terei de tomar medicação antirretrovírica para sempre ou a infeção por VIH tem cura?" (4,35), "Como evitar a transmissão da minha infeção (VIH) a outras pessoas?" (4,03), "Quais os efeitos adversos da terapêutica antirretrovírica?" (3,53) e "Posso ter relações sexuais desprotegidas (tendo infeção por VIH)?" (4,53). As questões colocadas no âmbito da PrEP foram: "A PrEP é uma medicação segura?" (4,58), "Quais as modalidades de PrEP?" (4,63), "Depois de começar a PrEP, tenho de fazer sempre?" (3,95), "Quais os efeitos adversos da PrEP?" (4,63/5) e "A PrEP protege contra outras infeções sexualmente transmissíveis ou tenho de usar preservativo?" (4,60).

Conclusão: Apenas duas respostas tiveram uma avaliação inferior a quatro pontos numa escala de Likert de 1 a 5, sendo isto um reconhecimento do potencial desta ferramenta no aconselhamento e literacia em saúde, desde que as preocupações éticas e legais sejam devidamente abordadas.

Palavras-chave: Inteligência Artificial; Infecções por VIH/prevenção e controlo; Literacia em Saúde; Profilaxia Pré-Exposição

1. Infectious Diseases Department. Hospital de Curry Cabral. Unidade Local de Saúde de São José. Lisbon. Portugal.
2. Environmental Health Institute. Faculdade de Medicina. Universidade de Lisboa. Lisbon. Portugal.
✉ **Autor correspondente:** José Ganicho. jose.ganicho@gmail.com
Recebido/*Received*: 01/01/2025 - **Aceite/*Accepted*:** 16/04/2025 - **Publicado/*Published*:** 02/06/2025

KEY MESSAGES

- Large language models can play a significant role in health literacy, given their ability to generate real-time answers, their accessibility, privacy, and intuitive interface.
- To explore their potential, this study assessed the quality of ChatGPT's answers to frequently asked questions about HIV and PrEP.
- On a Likert scale from 1 to 5, 80% of the answers received a mean rating above 4.
- The outputs depend on the sources these platforms are trained on, clear queries, and context, and may not capture linguistic or cultural nuances.
- Despite the positive evaluation, further studies must address confidentiality, legal implications, and patient acceptance of these technologies.

INTRODUCTION

Large language models in medicine

The use of artificial intelligence-based resources has transformed various sectors of society, including Medicine. One of these resources is chat generative pre-trained transformer (ChatGPT), a large language model (LLM) platform, introduced in 2018 by OpenAI[1] and made publicly available in 2022.[2,3] Large language models can recognize, interpret, understand language patterns and generate answers and outputs based on different sources, data, and literature. Therefore, they can have a wide range of applications, from text generation, translation or review[2] to chatbot-based technology in automated customer service systems.[4]

In Medicine, LLMs can serve multiple functions[3,5]: they can assist clinicians in therapeutic decision-making[6] processes, interpret electronic health records (EHRs)[7] and support clinicians in diagnostic processes.[8] Another field in which LLMs can have a significant role is health literacy, given their capability of generating answers in real-time, accessibility, privacy and intuitive interface with the user.[9-11] Despite the growing coverage regarding the use of this tool, it is important to evaluate its applicability, capabilities and safety,[11] particularly in Medicine.

Human immunodeficiency virus infection

Several studies highlight the link between low health literacy and worse health outcomes.[12-14] In fact, people with low health literacy are more likely to experience more adverse health conditions compared to the general population. The field of human immunodeficiency virus (HIV) is no exception and infection management and prevention are closely tied to health literacy levels. People living with HIV with low health literacy levels seem to face more disease-related complications, poorer medication adherence and more challenges in navigating healthcare services and understanding health-related information.[15-18] In fact, literacy remains a significant independent predictor of nonadherence,[18] emphasizing the need for initiatives and tools focused on improving patient health literacy.

The aim of the present study was to explore the potential of LLM platforms, specifically ChatGPT, in counseling and health literacy initiatives that focus on HIV diagnosis and pre-exposure prophylaxis (PrEP).

METHODS

To understand the potential of ChatGPT in counselling and health literacy related to HIV diagnosis and PrEP, an exploratory descriptive study was developed. A team of five infectious disease physicians also specialized in HIV prevention and management was questioned individually regarding the five most frequently asked questions by patients in the outpatient context regarding HIV inaugural diagnosis and PrEP. Questions related to logistics, such as appointment dates, laboratory results, and administrative information were excluded.

Each specialist independently submitted the five most frequently asked questions by patients regarding each topic, and the final 10 questions were selected by an independent investigator outside the experts' team based on the frequency with which they appeared across all submissions.

Each question was submitted individually and non-subsequently to the ChatGPT-4 version, without preamble, on August 8, 2024. No user preferences were set, nor was information regarding the querier provided to the platform beforehand. Furthermore, no memory functions were used in ChatGPT since the questions were submitted in non-paid, anonymous accounts. Since the expert team works in a clinical context involving Portuguese patients, the questions were submitted to the platform in Portuguese.

Afterwards, the answers were reviewed and rated based on a Likert scale from 1 to 5, independently by each expert, following a set of eight criteria: accuracy, clarity, relevance, depth, objectivity, conciseness, timeliness, and suitability for the target audience. The only statistical method used was a simple mathematical average of the ratings provided by the five experts.

Characterization of the team of experts

All physicians were native Portuguese speakers: four women and one man. They worked at the same hospital, having undergone a five-year residency program in infectious diseases, including in the field of HIV prevention and management, and had between two to twelve years of experience as experts (title acquired between 2012 and 2023).

Criteria definition

The experts had access to the criteria they would use to rate the answers, their definition and one example of applicability to clarify their meaning and significance in the context of HIV management. Table 1 presents the definition of each criterion, along with the sentence the experts were asked to rate based on their level of agreement with the statement.

Likert evaluation scale

The answers provided by the ChatGPT-4 version were rated by the initial team of five experts, using a Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree), based on their level of agreement with the statements in Table 1.

This study consisted of evaluating the quality of answers generated by an artificial intelligence platform (ChatGPT-4)

Table 1 – Criteria definition and applicability

Criteria	**Accuracy**
Definition	The measure by which the answer is correct and evidence based.
Example	The answer presents data that is confirmed by credible sources.
Statement	The answer is accurate.
Criteria	**Clarity**
Definition	The measure by which the answer can be easily understood.
Example	The answer uses simple and clear language, as well as a logical organization of ideas.
Statement	The answer is clear and easy to understand.
Criteria	**Relevance**
Definition	The measure by which the answer is directly related to the question asked.
Example	The answer avoids unnecessary information that is not related to the question.
Statement	The answer is relevant to the topic.
Criteria	**Depth**
Definition	The measure by which the answer explores the proportionate detail needed.
Example	The answer is explored in appropriate detail, taking into account nuances and implications.
Statement	The answer provides sufficient depth.
Criteria	**Objectivity**
Definition	The measure by which an answer is impartial, without bias or personal opinion.
Example	The answer is evidence-based and not subjective and personal opinions.
Statement	The answer is objective and unbiased.
Criteria	**Conciseness**
Definition	The measure by which an answer is direct, brief, and without redundancies.
Example	The answer avoids repetitive or irrelevant information.
Statement	The answer is concise and to the point.
Criteria	**Timeliness**
Definition	The measure by which an answer reflects updated and current information.
Example	The answer presents current information.
Statement	The answer is timely and up to date.
Criteria	**Suitability for the target audience**
Definition	The measure by which an answer is deemed appropriate for the target audience.
Example	The answer avoids complex language that is not appropriate for the target audience.
Statement	The answer is suitable for the intended audience.

to frequently asked questions by patients, as provided by healthcare professionals, concerning HIV infection and PrEP. The questions were submitted by infectious diseases specialists based on their clinical experience, and no personal, clinical, or identifiable patient data were used. There was no direct contact with patients, nor any intervention.

As this study relies exclusively on the analysis of the content generated by artificial intelligence, without involving human participants or the use of sociodemographic and/or clinical data, it was considered exempt from ethics committee approval. This approach is consistent with recognized ethical standards for research that does not involve risk to human subjects.

RESULTS

Evaluation by question

Table 2 presents the mean evaluation of the experts' assessment based on the eight criteria per question. Considering the eight criteria, 80% (8/10) of the answers had a mean evaluation by the experts higher than four (Likert scale 1-5).

Evaluation per criteria, per question, and group of questions

Table 2 also presents the mean evaluation of the experts' assessment based on each criterion, per question. It also highlights the mean evaluation of the answers to each group of questions, per criteria.

In the group of questions regarding the inaugural diagnosis of HIV, the two criteria with the worst evaluation were conciseness" and "suitability for the target audience", translating a length higher than desired and answers more complex or simple than required, considering the target audience. In comparison, the group of questions regarding PrEP received better evaluation across all criteria, except for "accuracy". However, it still had a mean evaluation higher than 4 (Likert scale 1-5) in this criteria.

DISCUSSION

This study was designed to explore the potential of LLM platforms, specifically ChatGPT, in counselling and health literacy initiatives focused on HIV diagnosis and PrEP. With that goal in mind, the study included a team of experts who manage HIV infection and individuals receiving PrEP and who, in their daily practice, welcome questions from patients on these topics.

Given the growing demand and challenges faced by healthcare systems and physicians, artificial intelligence tools may help in certain settings. Therefore, we aimed to test the platform's ability to produce quality outputs to ascertain its purpose, impact and relevance in clinical practice and health literacy initiatives.

In Table 2, 80% (8/10) of the answers provided had an evaluation higher than 4 (Likert scale 1 - 5). The two (20%) answers that presented an overall evaluation under 4 points (Likert scale 1 - 5), still had a positive assessment, over 3.5 points on the Likert scale.

Exploring the two answers with a lower evaluation, in the group of questions regarding the inaugural diagnosis of HIV, the answer to question Q4: "What are the adverse effects of the antiretroviral treatment?" had a rating of 3.53/5 and four criteria had an evaluation below this average: clarity, relevance, conciseness and suitability for the target audience. The experts considered that the answer was unclear, gave unnecessary and irrelevant information and was unsuitable for the target audience. Table 1 of the Appendix provides the answer to question Q4 (Appendix 1: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/22805/15667).

In the group of questions regarding PrEP, the answer to question Q8 "After I start taking pre-exposure prophylaxis, do I have to take it forever?" had an evaluation of 3.95/5, with four criteria with an assessment below average: accuracy, clarity, relevance and conciseness. The experts considered that the answer had an accuracy below desired, was unclear and gave unnecessary and redundant information. Table 1 of the Appendix provides the answer to question Q8 (Appendix 1: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/22805/15667).

Considering a total of ten answers evaluated based on eight criteria, we can explore a total of eighty evaluation markers. As shown in Table 2, considering both groups of questions, 80% (64) of these markers received a rating higher than 4 on the Likert scale (1 - 5). These results suggest that ChatGPT could be an interesting tool that can assist both physicians in their responsibility of health counselling and patients in clarifying their questions, thus contributing to raising health literacy.[10,17]

The impact of the source on the output generated

It is important to highlight the general limitations of ChatGPT-4, as well as those specific to this study and its application in the medical field.[3,6] Large language model platforms and ChatGPT are predominantly trained on data from the internet and as such they can present several biases, some of which are particularly sensitive in the context of the clinical practice, namely language, gender, racial, and cultural biases. In fact, the sources from which it gathers information can themselves be vulnerable to the previously mentioned biases, for example, they can express certain information regarding gender and race which may be inaccurate.[3,6]

Table 2 – Mean expert assessment of ChatGPT answers to questions regarding HIV inaugural diagnosis and PrEP

Rating criteria Likert scale 1 (strongly disagree) to 5 (strongly agree)	Mean expert assessment of ChatGPT answers to questions regarding HIV inaugural diagnosis **Q1: What does it mean to have HIV? Will I die sooner?** In Portuguese: *O que significa ter infeção por VIH para a minha vida? Vou morrer mais cedo?*	**Q2: Does HIV infection have a cure? Will I ever be able to stop taking medication?** In Portuguese: *A infecção por VIH tem cura? Alguma vez poderei deixar de tomar medicação?*	**Q3: Can I transmit my infection to other people? How to avoid it?** In Portuguese: *Posso transmitir a minha infecção a outras pessoas? Como evitar?*	**Q4: What are the adverse effects of the antiretroviral treatment?** In Portuguese: *Quais os efeitos adversos da terapêutica antirretrovírica?*	**Q5: Can I have unprotected sexual relationships? What are the risks?** In Portuguese: *Posso ter relações sexuais desprotegidas? Quais os riscos?*	**Total**
Accuracy	4.60	4.40	4.20	4.60	4.60	4.48
Clarity	4.40	4.20	4.00	3.00	4.60	4.04
Relevance	4.20	4.40	4.00	3.40	4.60	4.12
Depth	3.80	4.40	4.00	3.60	4.40	4.04
Objectivity	4.20	4.20	4.20	4.20	4.40	4.24
Conciseness	4.20	4.20	3.40	3.20	4.60	3.92
Timeliness	4.60	4.60	4.60	3.60	4.60	4.40
Suitability for the target audience	4.00	4.20	3.60	3.00	4.40	3.84
Total	4.25	4.35	4.03	3.53	4.53	

Rating criteria Likert scale 1 (strongly disagree) to 5 (strongly agree)	Mean expert assessment of ChatGPT answers to questions regarding PrEP **Q6: I want to start Pre-Exposure Prophylaxis. Is it safe?** In Portuguese: *Quero iniciar Profilaxia Pré-Exposição. É seguro?*	**Q7: What are the different regimens of Pre-Exposure Prophylaxis?** In Portuguese: *Quais as modalidades de Profilaxia Pré-Exposição?*	**Q8: After I start taking Pre-Exposure Prophylaxis, do I have to take it forever?** In Portuguese: *Depois de começar Profilaxia Pré-Exposição, tenho de fazer sempre?*	**Q9: What are the adverse effects of Pre-Exposure Prophylaxis?** In Portuguese: *Quais os efeitos adversos da Profilaxia Pré-Exposição?*	**Q10: Does Pre-Exposure Prophylaxis protect against other sexually transmitted infections (STIs) or do I have to use condoms?** In Portuguese: *A Profilaxia Pré-Exposição protege contra outras infecções sexualmente transmissíveis ou tenho de usar preservativo?*	**Total**
Accuracy	4.60	4.40	3.60	4.80	4.60	4.40
Clarity	4.60	4.60	3.40	4.60	4.80	4.40
Relevance	4.60	4.60	3.80	4.60	4.40	4.40
Depth	4.40	4.60	4.00	4.40	4.40	4.36
Objectivity	4.60	4.80	4.20	4.80	4.80	4.64
Conciseness	4.40	4.40	3.80	4.40	4.40	4.28
Timeliness	4.80	4.80	4.20	4.80	4.60	4.64
Suitability for the target audience	4.60	4.80	4.60	4.60	4.80	4.68
Total	4.58	4.63	3.95	4.63	4.60	

Input phrasing, language and context

Platforms such as ChatGPT are dependent on input phrasing, which means they rely on the way the user writes the prompts, which in turn impacts the quality of the generated content. Furthermore, if the user's queries are ambiguous or not sufficiently clear, taking into consideration the sources by which ChatGPT is trained, the platform may struggle to understand the context and the nuances needed to answer. This lack of understanding of the language and cultural nuances of how questions are made can produce superficial, inadequate or unsuitable answers to the target audience, defeating the purpose of using this type of platform for health literacy.[3] Also, it is important to highlight that platforms such as these, at least for now, have difficulties adapting their answer to the querier, namely their level of literacy, clinical context, and reasoning, limiting their capability of giving proper medical advice. With this in mind, it is important for AI-based models to clearly state that they cannot provide personalized medical counselling and that individuals should talk to a healthcare professional.

Individually and non-subsequently submitted questions

ChatGPT can provide better answers if context and background information are provided, since understanding the subject can improve the generated content.[3,6] Therefore, the accuracy and quality of the answers are dependent not only on the sources from which information is gathered but also on the context input. In this study, the questions were provided to ChatGPT-4 individually, non-subsequently, without preamble or memory functions which means that no context was provided to the platform. This was done to have a clear picture of the quality of the answers individually rather them being influenced by the previous context or answers provided. Since the experts served both as co-writers and evaluators, an independent investigator outside the team processed the data to mitigate potential investigator bias.

As shown in Table 2, 80% (8) of the answers provided had an evaluation higher than 4 (Likert scale 1 - 5), hence supporting their quality. It is expected that if more information was provided to ChatGPT-4, such as the target audience, the purpose of the questions, and the desired length of the answers, they would be more contextually aware, better suited to real-world situations, and overall, of higher quality.

Implementation, concerns and further areas of research

The implementation of AI-based technology such as ChatGPT depends on several factors that can be grouped into four sections: confidentiality, legal considerations, equity, and patient perspective. First, these platforms must ensure patient confidentiality, aligning with the General Data Protection Regulation and other data protection laws to preserve personal and medical information details. Second, the use of ChatGPT raises important legal questions such as the liability related to the answers provided in case they are incorrect, the dependence on a proprietary private platform and the lack of transparency on how the algorithms and data sources are integrated. Third, although it is an open-access tool, it requires a certain level of knowledge and proficiency, which may raise some equity-related questions, since not everyone may have the capability and resources needed to operate it. Finally, for the successful implementation of any tool, the patient's perspective is fundamental, so as to address their needs and challenges.

In this study, we focused on analyzing the potential of ChatGPT in health counseling and literacy regarding HIV infection and PrEP and, therefore, comparisons with similar studies in other medical fields or with answers provided by healthcare professionals were beyond the scope of this study. Nonetheless, both of these comparisons could be important to explore in future research to gain a deeper understanding of ChatGPT's potential in health literacy.

Finally, the aim of this study was to evaluate the quality of answers provided from a physician's perspective. However, to implement ChatGPT in a real-world clinical setting, further research is required. Beyond addressing the previously mentioned legal, ethical, and data protection concerns, it is important to conduct studies that actively assess patients' acceptance of these tools and their ability to understand and engage with them effectively.

CONCLUSION

People living with HIV with low health literacy levels seem to face more disease-related complications, poorer medication adherence and more challenges in understanding health-related information. Developing strategies that focus on expanding health literacy on HIV and PrEP is fundamental and LLM platforms, such as ChatGPT, can be a valuable tool for both physicians and patients.

The overall positive evaluation of every answer considering the eight criteria, testifies to the potential use of LLM platforms, specifically ChatGPT, in counselling and health literacy initiatives focusing on HIV diagnosis and PrEP. Nonetheless, when using artificial intelligence tools, particularly in Medicine, it is important to consider the platforms' bias, the input phrasing and context provided as well as how the prompts are written in order to have outputs and answers of higher quality.

DATA AVAILABILITY STATEMENT

Data regarding the evaluation of the experts' team is available upon request to the corresponding author.

AUTHOR CONTRIBUTIONS

JG: Study design, data interpretation, drafting of the manuscript.

MC, GC: Data interpretation, drafting of the manuscript.

CC, ML, DB, DS: Data interpretation, critical review of the manuscript.

ARG: Study design, data interpretation and discussion, drafting of the manuscript.

MJM: Critical review of the manuscript.

FM: Data interpretation and discussion, critical review of the manuscript.

All authors approved the final version to be published.

PROTECTION OF HUMANS AND ANIMALS

The authors declare that the procedures were followed according to the regulations established by the Clinical Research and Ethics Committee and to the Helsinki Declaration of the World Medical Association updated in October 2024.

DATA CONFIDENTIALITY

The authors declare having followed the protocols in use at their working center regarding patients' data publication.

COMPETING INTERESTS

JG received support for attending meetings and/or travel from ViiVHIV Healthcare, Pfizer, Merck Sharp & Dohme and Gilead Sciences.

MC received support for attending meetings and/or travel from Merck Sharp & Dohme and Gilead Sciences.

GC received support for attending meetings and/or travel from ViiVHIV Healthcare.

CC received consulting fees and payment for expert testimony from Merck Sharp & Dohme and Gilead Sciences; received support for attending meetings and/or travel from Janssen Cilag, ViiVHIV Healthcare, and Gilead Sciences; participated on Merck Sharp & Dohme and Gilead Sciences data safety monitoring boards or advisory boards.

ML received support for attending meetings and/or travel from ViiVHIV Healthcare, Pfizer, Merck Sharp & Dohme, and Gilead Sciences.

ARG received payment for expert testimony from ViiVHV Healthcare; received support for attending meetings and/or travel from ViiVHV Healthcare, Pfizer, Merck Sharp & Dohme, and Gilead Sciences; participated on a ViiVHIV Healthcare data safety monitoring board or advisory board.

DB received consulting fees and payment for expert testimony from Gilead Sciences; received support for attending meetings and/or travel from Gilead Sciences, AbbVie, ViiVHIV Healthcare, and Janssen Clag; participated on a Gilead Sciences data safety monitoring board or advisory board.

DS received consulting fees and payment or honoraria for lectures, presentations, speakers' bureaus, manuscript writing or educational events from Pfizer and Gilead Sciences; received payment for expert testimony from Pfizer and Giliead Sciences; received support for attending meetings and/or travel from Merck Sharp & Dohme, Pfizer, ViiVHIV Healthcare and Gilead Sciences; participated on Pfizer and Gilead Sciences data safety monitoring boards or advisory boards.

MJM received consulting fees, payment for expert testimony, support for attending meetings or travel and payment or honoraria for lectures, presentations, speakers' bureaus, manuscript writing or educational events from ViiVHIV Healthcare; participated on a ViiVHIV Healthcare data safety monitoring board or advisory board.

FM received consulting fees and payment for expert testimony from Exigo Consultores and Gilead Sciences; received payment or honoraria for lectures, presentations, speakers' bureaus, manuscript writing or educational events from ViiVHIV Healthcare and Gilead Sciences; received support for attending meetings or travel from Merck Sharp & Dohme, ViiVHIHV Healthcare, Gilead Sciences, and AbbVie.

FUNDING SOURCES

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

1. ChatGPT-4. [cited 2023 Aug 25]. Available from: https://openai.com/gpt-4.
2. Yang X, Chen A, PourNejatian N, Chang Shin H, Smith KE, Parisien C, et al. A large language model for electronic health records. npj Digit Med. 2022;5:1-9.
3. Xiaoliang M, RuQiang Z, Ying L, Congjian D, Dequan D. Design of a large language model for improving customer service in telecom operators. Electron Lett. 2024;60:e13218.
4. Rodkjaer LØ, Storgaard M, Sørensen NT, Schougaard LM. Levels of health literacy among people living with HIV in outpatient care: a cross-sectional study from Denmark. AIDS Res Ther. 2023;20:59.
5. Ríos-Hoyo A, Shan NL, Li A, Pearson AT, Pusztai L, Howard FM. Evaluation of large language models as a diagnostic aid for complex medical cases. Front Med. 2024;11:1380148.
6. Ray PP. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet Things Cyber-Phys Syst. 2023;3:121-54
7. Peres F. Health literacy in ChatGPT: exploring the potential of the use of artificial intelligence to produce academic text. Cienc Saude Colet. 2024;29:e02412023.
8. Osborn CY, Paasche-Orlow MK, Davis TC, Wolf MS. Health literacy: an overlooked factor in understanding HIV health disparities. Am J Prev Med. 2007;33:374-8.
9. Mgbako O, Conard R, Mellins CA, Dacus J, Remien RH. A systematic review of factors critical for HIV health literacy, ART adherence and retention in care in the U.S. for racial and ethnic minorities. AIDS Behav. 2022;26:3480-93.
10. Menichetti J, Hillen MA, Papageorgiou A, Pieterse AH. How can

ChatGPT be used to support healthcare communication research? Patient Educ Couns. 2023;115:107947.
11. Meng X, Yan X, Zhang K, Liu D, Cui X, Yang Y, et al. The application of large language models in medicine: a scoping review. iScience. 2024;27:109713.
12. McDonald M, Shenkman L. Health literacy and health outcomes of adults in the United States: implications for providers. Internet J Allied Health Sci Pract. 2018;16:1-5.
13. Ferguson LA, Pawlak R. Health literacy: the road to improved health outcomes. J Nurse Pract. 2011;7:123-9.
14. Fatani B. ChatGPT for future medical and dental research. Cureus. 2023;15: e37285.
15. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med. 2011;155:97.
16. Azizi Z, Alipour P, Gomez S, Broadwin C, Islam S, Sarraju A, et al. Evaluating recommendations about atrial fibrillation for patients and clinicians obtained from chat-based artificial intelligence algorithms. Circ Arrhythm Electrophysiol. 2023;16:415-7.
17. Ayre J, Mac O, McCaffery K, McKay BR, Liu M, Shi Y, et al. New frontiers in health literacy: using ChatGPT to simplify health information for people in the community. J Gen Intern Med. 2024;39:573-7.
18. Amutah-Onukagha N, Ibeziako N, Tibbitt C, Louis L, Amarnath A. Wisdom matters: honoring the wisdom and assessing the health literacy of black women living with HIV. J Healthc Sci Humanit. 2021;11:204-24.