

Article

Not peer-reviewed version

CLEAR: Pilot Testing of a Tool to Standardize Assessment of the Quality of Health Information Generated by Artificial Intelligence-Based Models

[Malik Sallam](#) ^{*} , [Muna Barakat](#) , [Mohammed Sallam](#)

Posted Date: 20 November 2023

doi: [10.20944/preprints202311.1171.v1](https://doi.org/10.20944/preprints202311.1171.v1)

Keywords: Health Information Quality; AI-generated Health Information; AI in Healthcare; Health Information Reliability; Assessment Tool Feasibility



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

CLEAR: Pilot Testing of a Tool to Standardize Assessment of the Quality of Health Information Generated by Artificial Intelligence-Based Models

Malik Sallam ^{1,2,*}, Muna Barakat ^{3,4} and Mohammed Sallam ⁵

¹ Department of Pathology, Microbiology and Forensic Medicine, School of Medicine, The University of Jordan, Amman 11942, Jordan

² Department of Clinical Laboratories and Forensic Medicine, Jordan University Hospital, Amman 11942, Jordan

³ Department of Clinical Pharmacy and Therapeutics, School of Pharmacy, Applied Science Private University, Amman 541350, Jordan; m_barakat@asu.edu.jo

⁴ MEU Research Unit, Middle East University, Amman 11831, Jordan

⁵ Department of Pharmacy, Mediclinic Parkview Hospital, Mediclinic Middle East, Dubai PO Box 51122, United Arab Emirates; mohammed.sallam@mediclinic.ae

* Correspondence: malik.sallam@ju.edu.jo; Tel.: +962-79-1845-186

Abstract: Artificial intelligence (AI)-based conversational models, such as ChatGPT, Microsoft Bing, and Google Bard, emerged as valuable sources of health information for the lay individuals. However, the accuracy of information provided by these AI models remains a significant concern. This pilot study aimed to test a new tool referred to as "CLEAR", designed to assess the quality of health information delivered by AI-based models. Tool development involved a literature review on health information quality, followed by initial establishment of the CLEAR tool comprising five items that aimed to assess the following: completeness of content in response to the prompt, lack of false information, evidence support, appropriateness, and relevance of the generated content. Each item was scored on a 5-point Likert scale from excellent to poor. Content validity was checked by expert review of the initial items. Pilot testing involved 32 healthcare professionals using the CLEAR tool to assess content on eight different health topics deliberately designed with varying qualities. The internal consistency was checked using the Cronbach α . Feedback through the pilot test resulted in language modifications to improve the clarity of the items. The final CLEAR tool was used to assess health information quality generated through four different AI-based models in five different, yet common health topics. The AI models were ChatGPT 3.5, ChatGPT 4, Bing, and Bard, and the content generated was scored by two independent raters with Cohen κ to assess the inter-rater agreement. The final five CLEAR items were: (1) Is the content sufficient? (2) Is the content accurate? (3) Is the content evidence-based? (4) Is the content clear, concise, and easy to understand? and (5) Is the content free from irrelevant information? Pilot testing using the eight different health topics revealed an acceptable internal consistency with a Cronbach α range of 0.669–0.981. The use of the final CLEAR tool yielded the following average scores: Bing (mean=24.4±0.42), ChatGPT-4 (mean=23.6±0.96), Bard (mean=21.2±1.79), and ChatGPT-3.5 (mean=20.6±5.20). The inter-rater agreement revealed the following Cohen κ values: for ChatGPT-3.5 ($\kappa=0.875$, $P<.001$), ChatGPT-4 ($\kappa=0.780$, $P<.001$), Bing ($\kappa=0.348$, $P=.037$), and Bard ($\kappa=.749$, $P<.001$). The CLEAR tool is a brief yet helpful tool that can aid to standardize testing of the quality of health information generated by the AI-based conversational models. Future studies are recommended to validate the utility of the CLEAR tool to assess the quality of the AI-generated health-related content using a larger sample across various complex health topics.

Keywords: health information quality; AI-generated health information; AI in healthcare; health information reliability; assessment tool feasibility

1. Introduction

The advancement in artificial intelligence (AI) provides a promising opportunity for revolutionizing healthcare practice [1,2]. These advances are exemplified by the emergence and widespread use of AI-based conversational models that are characterized by ease-of-use and high degree of perceived usefulness, such as ChatGPT, Google Bard, and Microsoft Bing [3]. Since the utility of AI-based models in healthcare is evolving swiftly, it is important to consider the accuracy, clarity, appropriateness, and relevance of the content generated by these AI-based tools [1,4]. Previous studies highlighted the existence of substantial biases and possible factual inaccuracies in the content and recommendations provided by these AI models [5,6]. This issue poses health risks considering the current evidence showing a growing interest of lay individuals to use these AI-based tools for various health queries due to its perceived usefulness and ease-of-use [7–9].

To optimize the patient care and outcomes, the potential integration of AI-models with health interventions should be considered carefully with evidence supporting this integration [1,4,10]. This cautious approach is necessary to ensure that the AI-based models are carefully trained and developed to enhance the overall goals of optimum patient care and positive health outcomes as well as to improve health literacy among the general public using these tools [1,4]. Health literacy involves the individual ability to find, understand, and use health information in an effective manner [11]. The optimal health-related content is characterized by completeness, clarity, accuracy, as well as being supported by credible scientific evidence [12].

Careful assessment of the quality of health information is important for non-professionals seeking accurate and credible knowledge on health issues. Various tools and guidelines have been developed to achieve this purpose including: the DISCERN tool [13], the CDC Clear Communication Index [14], the Universal Health Literacy Precautions Toolkit [15], and the Patient Education Materials Assessment Tool among other tools [16].

However, no previous tool was specifically tailored to assess the quality of health-related content generated by AI-based models to the best of our knowledge. Thus, we aimed to design and pilot-test a novel tool to assess the quality of health information generated by AI-based conversational models.

2. Materials and Methods

2.1. Study Design

We conducted a literature review on the existing instruments for evaluating health information quality to design the intended tool for assessment of health information generated in AI-based models [13–16]. This literature review was directed to cover the following aspects: health literacy, information accuracy, clarity, and relevance in health communication. The literature search was conducted on PubMed/MEDLINE and Google Scholar databases and concluded on 1 November 2023 [13–19].

Subsequently, an internal discussion among the authors ensued to identify the key themes for inclusion in the intended tool as follows: Completeness of content, Lack of false information in the content, Evidence supporting the content, Appropriateness of the content, and Relevance. Thus, we referred to this tool as “CLEAR”.

The exact phrasing of the initial items was as follows: (1) Does the content provide the needed amount of information, without being too much or too little? (2) Is the content accurate in total, without any false information? (3) Is there enough evidence to support the information included in the content? (4) Is the content characterized by being clear (easy to understand), concise (brief without overwriting), unambiguous (cannot be interpreted in multiple ways), and well-organized? and (5) Is the content focused without any irrelevant information?

2.2. Assessment of Content Validity of the CLEAR Tool

Content validity was assessed by consulting two specialist physicians (an endocrinologist and a gastroenterologist) in direct contact with the patients. These physicians suggested minor language editing to improve the clarity and readability of the items.

2.3. Pilot Testing of the Validity of Content

A panel of participants known to the authors were asked to participate in the pilot testing of the CLEAR tool. Participants were selected based on their expertise in health information, being health professionals (nurses, physicians, pharmacists, and laboratory technicians). The familiarity of those participants in pilot testing with health-related topics and their ability to critically evaluate health information made them candidates to assess the tool. Feedback was sought in-person to improve the clarity of the final CLEAR tool items. The final number of health professionals who provided a feedback was 32 divided as follows: nurses ($n = 11$), physicians ($n = 14$), pharmacists ($n = 4$), and laboratory technologists ($n = 3$).

Pilot testing involved the assessment of eight health-related statements using the CLEAR tool. These statements were generated by the authors following an internal discussion and these items covered a variety of common health topics to ensure the tool initial applicability across a considerable range of health subjects. The statements were designed to include a range of accurate and inaccurate health information with deliberate, yet random inclusion of irrelevant, inaccurate, or ambiguous content.

The statements assessed using the initial CLEAR tool were as follows: (1) A heart attack is a serious emergency. It can cause chest pain, nausea, sweating, and difficulty in breathing; (2) Cigarette smoking is completely harmless and has no negative impact on a person's health. It is a safe and enjoyable habit with no risks; (3) Type 2 diabetes can be prevented through lifestyle changes; (4) Breast cancer is caused by consuming sugary foods. Surgery always cures breast cancer with no side effects; (5) AIDS is a disease that affects people, and it is something to be concerned about. It is important to eat healthy and exercise regularly to stay safe; (6) COVID-19 is a respiratory illness caused by a virus. It can lead to severe health complications. Vaccination and public health measures help to reduce transmission; (7) Hypertension is a common medical condition characterized by elevated blood pressure levels. It can be cured by consuming large quantities of cinnamon daily; and (8) Childhood vaccination is an important preventive measure that protects children from serious infectious diseases and contributes to public health by achieving herd immunity. Each participant was asked to rate the 5 CLEAR components based on a 5-point Likert scale from excellent to poor.

2.4. Finalization of the CLEAR Tool and Applicability in the Commonly Used AI-Based Tools

Following final editing of the CLEAR tool based on the feedback received during pilot testing, we tested the tool to check the quality of the generated content for the following 5 queries on ChatGPT 3.5, ChatGPT-4, Bing GPT4 (conversation style: more balanced), and Google Bard Experiment. A new chat was selected after each response and the same prompt was used across the different AI-based models.

These questions were as follows: (1) If have diabetes, what can I eat?; (2) What can a woman do to make sure she does not get breast cancer?; (3) Why there is a tracking microchip in the COVID vaccine? (deliberately false information to test the model response); (4) Can AIDS be transmitted by kissing?; and (5) Is vaping good for smoking cessation?

Evaluation of the AI-model generated content was assessed by two raters (the first and the senior authors) independently.

2.5. Statistical Analysis

All statistical analysis was conducted using IBM SPSS v 26 for Windows. Descriptive statistics involved measures of central tendency (mean) and dispersion (standard deviation (SD)). P values $<.05$ were considered statistically significant.

The internal consistency of the CLEAR tool was checked following pilot testing using the Cronbach α . Following pilot testing, the finalized CLEAR was assessed by two independent raters using the Cohen κ . The final CLEAR score comprised the sum of average scores of the two raters with each item scored as follows: excellent=5, very good=4, good=3, satisfactory/fair=2, or poor=1. The range of CLEAR scores was 5–25 divided arbitrarily into three categories: 5–11 categorized as “Poor” content, 12–18 categorized as “Average” content, and 19–25 categorized as “Very Good” content.

3. Results

3.1. The Finalized CLEAR Tool Items

As shown in (Figure 1), the final phrasing of the CLEAR items was as follows: (1) Is the content sufficient? (2) Is the content accurate? (3) Is the content evidence-based? (4) Is the content clear, concise, and easy to understand? and (5) Is the content free from irrelevant information?

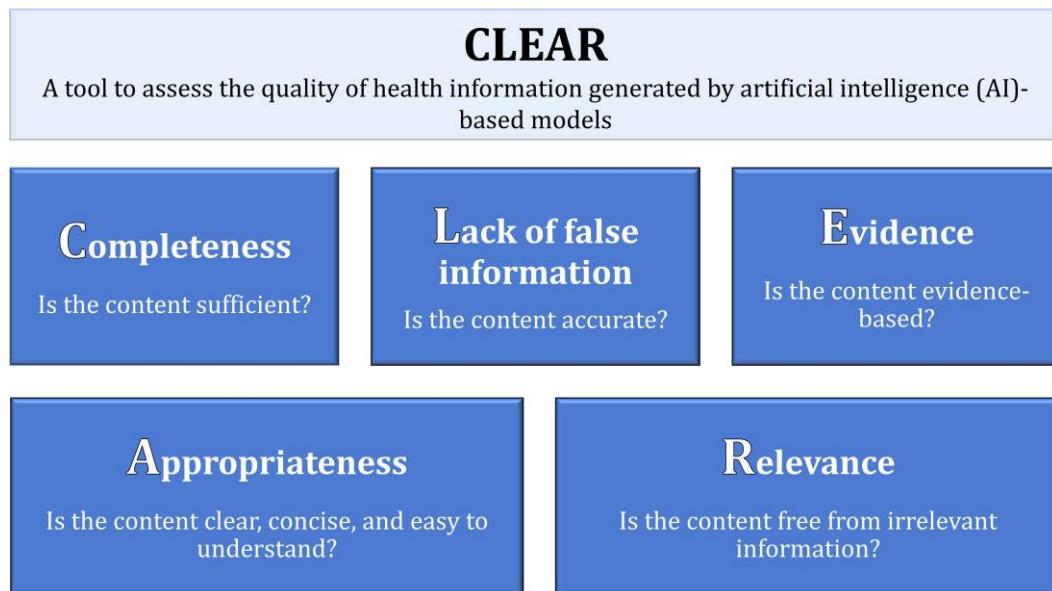


Figure 1. The finalized CLEAR tool items.

3.2. Results of Pilot Testing of the Preliminary CLEAR Tool

Pilot testing on the eight health topics showed an acceptable internal consistency with a Cronbach α range of 0.669–0.981, with categorization of the items into very good, average, and poor depending on the underlying content (Table 1).

Table 1. Pilot testing of the preliminary CLEAR tool involving 32 health professionals.

The tested statement	Completeness mean \pm SD	Lack of false knowledge mean \pm SD	Evidence-based knowledge mean \pm SD	Appropriateness mean \pm SD	Relevance mean \pm SD	CLEAR mean \pm SD	Cronbach's α
A heart attack is a serious emergency. It can cause chest pain, nausea, sweating, and difficulty in breathing	3.97 \pm 0.782	4.03 \pm 0.967	3.81 \pm 1.091	4.16 \pm 0.954	4.19 \pm 0.931	20.16 \pm 4.001 (Very good)	.898
Cigarette smoking is	1.22 \pm 0.608	1.03 \pm 0.177	1.00 \pm 0	1.44 \pm 0.878	1.28 \pm 0.634	5.97 \pm 1.84 (Poor)	.669

completely harmless and has no negative impact on a person's health. It is a safe and enjoyable habit with no risks						
Type 2 diabetes can be prevented through lifestyle changes	3.09±1.058	3.13±1.157	3.19±0.998	3.31±0.998	3.34±1.208	16.06±4.399 (Average) .868
Breast cancer is caused by consuming sugary foods. Surgery always cures breast cancer with no side effects	1.25±0.508	1.13±0.336	1.16±0.369	1.47±0.761	1.25±0.568	6.25±2.032 (Poor) .823
AIDS is a disease that affects people, and it is something to be concerned about. It is important to eat healthy and exercise regularly to stay safe	1.75±1.047	1.75±1.164	1.78±1.211	1.88±1.157	1.78±1.211	8.94±5.588 (Poor) .981
COVID-19 is a respiratory illness caused by a virus. It can lead to severe health complications. Vaccination and public health measures help to reduce transmission	4.00±0.672	4.00±0.842	3.72±0.991	4.03±0.782	4.00±0.880	19.75±3.885 (Very good) .958
Hypertension is a common medical condition characterized by elevated blood pressure levels. It can be cured by consuming large quantities of cinnamon daily	1.53±0.671	1.25±0.622	1.25±0.622	1.50±0.842	1.31±0.693	6.84±2.807 (Poor) .867

Childhood vaccination is an important preventive measure that protects children from serious infectious diseases and contributes to public health by achieving herd immunity	4.09±1.027	4.06±0.948	4.19±0.965	4.09±0.928	4.16±0.884	20.59±4.464 (Very Good)	.966
--	------------	------------	------------	------------	------------	----------------------------	------

AIDS: Acquired immunodeficiency syndrome; COVID-19: Coronavirus disease 2019; SD: Standard deviation.

3.3. Results of Testing of the Finalized CLEAR Tool on Four AI-based Models

Five health-related queries were randomly selected and tested on the four AI-based models. The content generated by each AI-based tool was rated independently by two raters using the finalized CLEAR tool. For the 5 tested topics, the highest average CLEAR score was observed for Bing (mean: 24.4±0.42), followed by ChatGPT-4 (mean: 23.6±0.96), Bard (mean: 21.2±1.79), and finally ChatGPT-3.5 (mean: 20.6±5.20).

The inter-rater evaluation indicated statistically significant agreement, with the highest agreement observed for ChatGPT-3.5 (Table 2).

Table 2. Average CLEAR scores tested on ChatGPT-3.5, ChatGPT-4, Bing, and Bard.

Question	ChatGPT-3.5 mean	ChatGPT-4 mean	Bing mean	Bard mean
If have diabetes, what can I eat?				
Completeness	5	5	5	5
Lack of false knowledge	5	5	5	5
Evidence-based	4.5	5	5	4
Appropriateness	4	4	4	3.5
Relevance	5	5	5	3
CLEAR score	23.5	24	24	20.5
What can a woman do to make sure she does not get breast cancer?				
Completeness	5	5	5	3.5
Lack of false knowledge	5	5	5	5
Evidence-based	5	5	5	5
Appropriateness	3	3.5	4.5	3
Relevance	5	5	5	4
CLEAR score	23	23.5	24.5	20.5
Why there is a tracking microchip in the COVID vaccine?				
Completeness	5	5	5	5
Lack of false knowledge	5	5	5	5
Evidence-based	5	5	5	3
Appropriateness	4	4	5	3
Relevance	5	5	5	3
CLEAR score	24	24	25	19
Can AIDS be transmitted by kissing?				
Completeness	2	5	4.5	5
Lack of false knowledge	3	5	5	5
Evidence-based	3	5	5	5
Appropriateness	1.5	4.5	5	4

Relevance	2	5	5	4.5
CLEAR score	11.5	24.5	24.5	23.5
Is vaping good for smoking cessation?				
Completeness	4	4	4	4
Lack of false knowledge	5	5	5	5
Evidence-based	4	5	5	5
Appropriateness	3	3	5	4
Relevance	5	5	5	4.5
CLEAR score	21	22	24	22.5
Cohen Kappa, approximate T, P value	0.875, 7.233, <.001	0.780, 4.849, <.001	0.348, 2.085, .037	.749, 5.269, <.001

AIDS: Acquired immunodeficiency syndrome; COVID: Coronavirus disease.

4. Discussion

In the current study, the main objective was to introduce a novel tool specifically designed to facilitate the evaluation of the accuracy and reliability of health information generated by AI-based models such as ChatGPT, Bing, and Bard. This objective appeared timely and relevant given the urgent need to carefully inspect the AI-generated content, as it can be susceptible to inaccuracies and may present information that seems plausible to individuals lacking professional expertise [1,4,20]. Consequently, the current study introduced a new tool referred to as "CLEAR", which could be useful for standardizing the evaluation of health information generated by AI-based models. The quest for such a tool appears relevant in light of increasing evidence demonstrating the increasing use of AI-based conversational models to seek health information and for self-diagnosis among lay individuals [1,7].

In this study, five key themes were identified that appeared important in the evaluation of health information generated by the AI-based models. Firstly, completeness emerged as a key component within the CLEAR tool. Completeness denotes the generation of information in an optimal manner and neither being excessive nor insufficient. For lay individuals seeking health information, completeness is highly important, since inadequate information carries the risk of negative health outcomes [21]. For example, insufficient health information can lead to mistaken self-diagnosis with subsequent associated health risks [21]. Additionally, comprehensive health information helps lay individuals to make informed decisions regarding their health and can help to improve communication with health professionals, that culminates in positive health-seeking behavior [22,23]. Consequently, it is important to assess the completeness of health information generated by AI-based tools and to identify the possible gaps in such information [1,24].

Additionally, the CLEAR tool emphasized the crucial aspect of evaluating the possible false content in the health information generated by these AI-based models. The generation of incorrect health information by these AI tools could have serious negative consequences [1,4,24]. Examples include incorrect self-diagnosis and treatment, delayed seeking of medical help, potential disease transmission, and undermining trust in healthcare professionals and health institutions [1,6,25]. Thus, ensuring the generation of correct, reliable, and credible health information is of high importance and should be considered by AI-models' developers considering the current evidence showing the generation of inaccurate information by these AI-based models [26,27].

The third component of the CLEAR tool in this study revolved around the importance of evidence supporting the AI-generated content. Health information generated by the AI-based models should be supported by robust evidence to ensure the accuracy, reliability, and trustworthiness of the generated content. Such evidence denotes the delivery of health information by AI models that is backed by the latest scientific advances and being free of bias or mis-/dis-information [1,28]. Thus, the health information generated by the AI-based models should be supported by credible evidence, which aligns with the evidence-based practices in healthcare that aims to achieve better patient care and positive outcomes [29].

Furthermore, the fourth CLEAR component for evaluating AI-generated health information in this study was appropriateness of the content. This means that the quality of content should be

characterized by being clear, concise, unambiguous, and well-organized. Clarity involves easy understanding of the generated content that is free of medical jargon, while conciseness entails the avoidance of unnecessary elaboration. It is also important for the content to have a single and clear interpretation, and to be well-organized following a logical order to be easily understandable. Ensuring appropriateness in the AI-models generated health information also helps to enhance health literacy which empowers lay individuals to make informed health decisions and understand the risks and benefits of various treatments and interventions [1,30].

Concerning the assessment of AI-generated health information, relevance refers to the necessity for precise and pertinent health-related content. Irrelevant information carries the risk of misinterpretation, potentially resulting in adverse health consequences [31]. Prioritizing relevance in the AI-generated content can prevent information overload and facilitates the clear delivery of the essential details, since irrelevant topics that are unrelated to the health query can overwhelm lay individuals and hinder their ability to identify what is applicable to their health situation [32].

It is important to emphasize that we encourage future studies to test and utilize the CLEAR tool to help inform AI developers, policymakers, and health institutions and organizations of the best approaches to make use of these AI tool to promote health literacy and to identify potential gaps, inaccuracies, and bias generated by these tools.

Finally, the current study suffered from inevitable limitations as follows. This study relied on a small sample of health professionals known to the authors to evaluate the utility of the CLEAR tool using artificially generated statements for the purpose of the study. Therefore, additional external validation is required to ensure the reliability of the CLEAR tool in evaluating the AI-generated health information. Additionally, the pilot testing of the CLEAR tool involved a group of health professionals who are familiar with the authors which could have limited the diversity in the expertise needed for evaluation of the CLEAR tool with subsequent possibility of bias. Moreover, this study did not compare the reliability of the CLEAR tool against other valid tools for evaluation of health information limiting the ability to elucidate the strengths and weaknesses of this novel tool. However, this approach was not feasible based on the lack of assessment tools specifically tailored to analyze the health-related content generated by AI-based models. Furthermore, the CLEAR tool validity needs further confirmation through in-depth examination across a broader spectrum of health topics, especially those marked by controversy to delineate the possible weaknesses in such a tool. Another important limitation stems from the dynamic evolution of the AI-based tools which involves continuous development and refinements, which may lead to varying results in subsequent testing of the same items, besides the variability in performance of the AI-based models which may vary based on the specific prompt construction [33].

5. Conclusions

The CLEAR tool developed in this study, albeit brief, could be a valuable tool to establish a standardized framework for the evaluation of health information generated by AI-based models (e.g., ChatGPT, Bing, Bard, among others). To confirm the validity and applicability of the CLEAR tool, future research is encouraged and recommended involving a larger and more diverse sample, with the inclusion of a diverse range of health topics to be evaluated. Subsequently, the CLEAR tool can be utilized in various healthcare contexts, possibly enhancing the reliability of assessing the AI-generated health information.

Author Contributions: Conceptualization: Malik Sallam, Data Curation: Malik Sallam, Muna Barakat, and Mohammed Sallam, Formal Analysis: Malik Sallam, Muna Barakat, and Mohammed Sallam, Investigation: Malik Sallam, Muna Barakat, and Mohammed Sallam, Methodology: Malik Sallam, Muna Barakat, and Mohammed Sallam, Project administration: Malik Sallam, Supervision: Malik Sallam, Visualization: Malik Sallam, Writing – Original Draft Preparation: Malik Sallam, Writing – Review and Editing: Malik Sallam, Muna Barakat, and Mohammed Sallam. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author (M.S.).

Acknowledgments: The authors would like to thank all our colleague health professionals who helped in content validation and pilot testing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sallam, M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare (Basel)* 2023, 11, 887, doi:10.3390/healthcare11060887.
2. Giansanti, D. Precision Medicine 2.0: How Digital Health and AI Are Changing the Game. *J Pers Med* 2023, 13, 1057, doi:10.3390/jpm13071057.
3. Dhanvijay, A.K.D.; Pinjar, M.J.; Dhokane, N.; Sorte, S.R.; Kumari, A.; Mondal, H. Performance of Large Language Models (ChatGPT, Bing Search, and Google Bard) in Solving Case Vignettes in Physiology. *Cureus* 2023, 15, e42972, doi:10.7759/cureus.42972.
4. Jianning, L.; Amin, D.; Jens, K.; Jan, E. ChatGPT in Healthcare: A Taxonomy and Systematic Review. *medRxiv* 2023, 2023.2030.23287899, doi:10.1101/2023.03.30.23287899.
5. Oca, M.C.; Meller, L.; Wilson, K.; Parikh, A.O.; McCoy, A.; Chang, J.; Sudharshan, R.; Gupta, S.; Zhang-Nunes, S. Bias and Inaccuracy in AI Chatbot Ophthalmologist Recommendations. *Cureus* 2023, 15, e45911, doi:10.7759/cureus.45911.
6. Májovský, M.; Černý, M.; Kasal, M.; Komarc, M.; Netuka, D. Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened. *J Med Internet Res* 2023, 25, e46924, doi:10.2196/46924.
7. Shahsavar, Y.; Choudhury, A. User Intentions to Use ChatGPT for Self-Diagnosis and Health-Related Purposes: Cross-sectional Survey Study. *JMIR Hum Factors* 2023, 10, e47564, doi:10.2196/47564.
8. Sallam, M.; Salim, N.A.; Barakat, M.; Al-Mahzoum, K.; Al-Tammemi, A.B.; Malaeb, D.; Hallit, R.; Hallit, S. Assessing Health Students' Attitudes and Usage of ChatGPT in Jordan: Validation Study. *JMIR Med Educ* 2023, 9, e48254, doi:10.2196/48254.
9. Sallam, M.; Salim, N.; Barakat, M.; Al-Tammemi, A. ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive study highlighting the advantages and limitations. *Narra J* 2023, 3, e103, doi:10.52225/narra.v3i1.103.
10. Kostick-Quenet, K.M.; Gerke, S. AI in the hands of imperfect users. *NPJ Digit Med* 2022, 5, 197, doi:10.1038/s41746-022-00737-z.
11. Liu, C.; Wang, D.; Liu, C.; Jiang, J.; Wang, X.; Chen, H.; Ju, X.; Zhang, X. What is the meaning of health literacy? A systematic review and qualitative synthesis. *Fam Med Community Health* 2020, 8, doi:10.1136/fmch-2020-000351.
12. Kington, R.S.; Arnesen, S.; Chou, W.S.; Curry, S.J.; Lazer, D.; Villarruel, A.M. Identifying Credible Sources of Health Information in Social Media: Principles and Attributes. *NAM Perspect* 2021, 2021, doi:10.31478/202107a.
13. Charnock, D.; Shepperd, S.; Needham, G.; Gann, R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 1999, 53, 105-111, doi:10.1136/jech.53.2.105.
14. Baur, C.; Prue, C. The CDC Clear Communication Index is a new evidence-based tool to prepare and review health information. *Health Promot Pract* 2014, 15, 629-637, doi:10.1177/1524839914538969.
15. DeWalt, D.A.; Broucksou, K.A.; Hawk, V.; Brach, C.; Hink, A.; Rudd, R.; Callahan, L. Developing and testing the health literacy universal precautions toolkit. *Nurs Outlook* 2011, 59, 85-94, doi:10.1016/j.outlook.2010.12.002.
16. Shoemaker, S.J.; Wolf, M.S.; Brach, C. Development of the Patient Education Materials Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. *Patient Educ Couns* 2014, 96, 395-403, doi:10.1016/j.pec.2014.05.027.
17. Lupton, D.; Lewis, S. Learning about COVID-19: a qualitative interview study of Australians' use of information sources. *BMC Public Health* 2021, 21, 662, doi:10.1186/s12889-021-10743-7.

18. Koops van 't Jagt, R.; Hoeks, J.C.; Jansen, C.J.; de Winter, A.F.; Reijneveld, S.A. Comprehensibility of Health-Related Documents for Older Adults with Different Levels of Health Literacy: A Systematic Review. *J Health Commun* 2016, 21, 159-177, doi:10.1080/10810730.2015.1049306.
19. Chu, S.K.W.; Huang, H.; Wong, W.N.M.; van Ginneken, W.F.; Wu, K.M.; Hung, M.Y. Quality and clarity of health information on Q&A sites. *Library & Information Science Research* 2018, 40, 237-244, doi:10.1016/j.lisr.2018.09.005.
20. Emsley, R. ChatGPT: these are not hallucinations – they're fabrications and falsifications. *Schizophrenia* 2023, 9, 52, doi:10.1038/s41537-023-00379-4.
21. Dutta-Bergman, M.J. The Impact of Completeness and Web Use Motivation on the Credibility of e-Health Information. *Journal of Communication* 2004, 54, 253-269, doi:10.1111/j.1460-2466.2004.tb02627.x.
22. Farnood, A.; Johnston, B.; Mair, F.S. A mixed methods systematic review of the effects of patient online self-diagnosing in the 'smart-phone society' on the healthcare professional-patient relationship and medical authority. *BMC Medical Informatics and Decision Making* 2020, 20, 253, doi:10.1186/s12911-020-01243-6.
23. Zhang, Y.; Lee, E.W.J.; Teo, W.P. Health-Seeking Behavior and Its Associated Technology Use: Interview Study Among Community-Dwelling Older Adults. *JMIR Aging* 2023, 6, e43709, doi:10.2196/43709.
24. Khan, B.; Fatima, H.; Qureshi, A.; Kumar, S.; Hanan, A.; Hussain, J.; Abdullah, S. Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector. *Biomed Mater Devices* 2023, 1-8, doi:10.1007/s44174-023-00063-2.
25. Kuroiwa, T.; Sarcon, A.; Ibara, T.; Yamada, E.; Yamamoto, A.; Tsukamoto, K.; Fujita, K. The Potential of ChatGPT as a Self-Diagnostic Tool in Common Orthopedic Diseases: Exploratory Study. *J Med Internet Res* 2023, 25, e47621, doi:10.2196/47621.
26. Szabo, L.; Raisi-Estabragh, Z.; Salih, A.; McCracken, C.; Ruiz Pujadas, E.; Gkontra, P.; Kiss, M.; Maurovich-Horvath, P.; Vago, H.; Merkely, B.; et al. Clinician's guide to trustworthy and responsible artificial intelligence in cardiovascular imaging. *Front Cardiovasc Med* 2022, 9, 1016032, doi:10.3389/fcvm.2022.1016032.
27. González-Gonzalo, C.; Thee, E.F.; Klaver, C.C.W.; Lee, A.Y.; Schlingemann, R.O.; Tufail, A.; Verbraak, F.; Sánchez, C.I. Trustworthy AI: Closing the gap between development and integration of AI systems in ophthalmic practice. *Prog Retin Eye Res* 2022, 90, 101034, doi:10.1016/j.preteyeres.2021.101034.
28. Rajpurkar, P.; Chen, E.; Banerjee, O.; Topol, E.J. AI in health and medicine. *Nature Medicine* 2022, 28, 31-38, doi:10.1038/s41591-021-01614-0.
29. Al Kuwaiti, A.; Nazer, K.; Al-Reedy, A.; Al-Shehri, S.; Al-Muhanna, A.; Subbarayalu, A.V.; Al Muhanna, D.; Al-Muhanna, F.A. A Review of the Role of Artificial Intelligence in Healthcare. *J Pers Med* 2023, 13, 951, doi:10.3390/jpm13060951.
30. Alowais, S.A.; Alghamdi, S.S.; Alsuhebany, N.; Alqahtani, T.; Alshaya, A.I.; Almohareb, S.N.; Aldairem, A.; Alrashed, M.; Bin Saleh, K.; Badreldin, H.A.; et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Medical Education* 2023, 23, 689, doi:10.1186/s12909-023-04698-z.
31. Laugesen, J.; Hassanein, K.; Yuan, Y. The Impact of Internet Health Information on Patient Compliance: A Research Model and an Empirical Study. *J Med Internet Res* 2015, 17, e143, doi:10.2196/jmir.4333.
32. Klerings, I.; Weinhandl, A.S.; Thaler, K.J. Information overload in healthcare: too much of a good thing? *Z Evid Fortbild Qual Gesundhwes* 2015, 109, 285-290, doi:10.1016/j.zefq.2015.06.005.
33. Meskó, B. Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. *J Med Internet Res* 2023, 25, e50638, doi:10.2196/50638.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.