ANKARA MEDICAL JOURNAL

Research Article

Ankara Med J, 2025;(1):27-39 // 10.5505/amj.2025.54810

ASSESSING THE ACCURACY, READABILITY AND UNDERSTANDABILITY OF WEBSITES, CHATGPT, COPILOT, AND BARD ANSWERS ON THE RADIATION DURING PREGNANCY

Burcu Mert[1], Emre Emekli[1]

[1]Department of Radiology, Eskisehir Osmangazi University, Faculty of Medicine, Eskişehir, Türkiye

Correspondence:
Burcu Mert (e-mail: burcumert1995@gmail.com)

Submitted: 18.08.2024 // Accepted: 05.02.2025

Ankara Yıldırım Beyazıt University Faculty of Medicine
Department of Family Medicine

Abstract

Objectives: The study aims to evaluate the accuracy of answers to frequently asked questions (FAQ) about the impact of radiation during pregnancy on websites, ChatGPT, Copilot, and Bard. Secondly, to assess the readability and understandability of answers.

Materials and Methods: The answers to these questions were scored in terms of accuracy (completely correct, partially correct, incorrect). The Automated Readability Index (ARI), Flesch Reading Ease (FRE), and Gunning Fog Readability (GFR) scores were calculated. The understandability score was assessed using the Patient Education Materials Assessment Tool (PEMAT).

Results: The accuracy was calculated as 100% for the websites, 66.67% for ChatGPT, 73.33% for Copilot, and 93.33% for Bard. Readability scores ranking was ChatGPT (ARI=16.15, FRE=24.47, GFR=20.52), Copilot (ARI=14.00, FRE=37.60, GFR=18.27), websites (ARI=13.59, FRE=43.67, GFR=15.56), Bard (ARI=10.92, FRE=48.73, GFR=14.86). ChatGPT's readability was statistically the most challenging. PEMAT understandability scores were 79.53% for Bard, below the acceptable limit of 70% for others.

Conclusion: While the responses from chat tools and websites may be largely accurate, it is observed that they are not suitable for patients in terms of readability and understandability. Internet information sources should be developed, especially to ensure that the content is understandable by a broad readership.

Keywords: Artificial intelligence, ChatGPT, Bard, Copilot, Chat Tool, readability.

Introduction

Artificial intelligence (AI)-supported programs are utilized in various fields within the healthcare system, such as aiding physicians in diagnosis, determining treatment options, helping patients understand their illnesses, and providing answers to their questions. They also contribute to medical education for students.[1,2] The introduction of ChatGPT into the public sphere in November 2022 has led to increased use of AI resources by the community.[3] Following the release of ChatGPT, Microsoft Copilot, and Google Bard subsequently entered the scene, resulting in a rise in the number of AI-powered chat tools available. These models facilitate information access and integration by providing a natural language interface and enabling interactive conversations that deliver real-time, instant responses to questions. [4]

As information sources on the internet and people's access to the internet have increased, the rate of patients accessing health information online has also risen.[5] It is known that individuals conduct online research not only about their illnesses, symptoms, and treatments but also generally about health information.[6,7] The positive and negative effects of accessing web-based information about their illnesses by patients have been studied in the literature.[8] Due to the easy accessibility and ability to provide instant responses, AI-powered chat tools have become frequently used resources in various fields for information research today.[9] However, due to their novelty, there is not a sufficient number of studies available in the literature about these chat tools. Nevertheless, studies in the literature have explored the ability of AI-powered chat tools, especially ChatGPT, to respond to specific patient questions and the accuracy of their responses in certain subjects.[10-12] It is not enough for these chat tools to simply respond to questions or provide correct answers. The readability and understandability of the answers by patients are also important. While these parameters have been frequently studied about websites in the literature,[10] there is limited research in the literature evaluating the readability and understandability of AI-powered tool responses.[11-15]

Most radiological diagnostic methods involve ionizing radiation. However, in pregnant patients, these diagnostic modalities may need to be used for diagnostic and therapeutic purposes when necessary. Sometimes, patients may have used these modalities without knowing they are pregnant.[16] Therefore, pregnant or potentially pregnant patients seek answers to numerous questions regarding which modalities involve radiation, what doses can affect the fetus, and what these effects are. The study seeks to assess responses to frequently asked questions regarding the impact of radiation during pregnancy on websites, while also evaluating the accuracy rates of answers provided by ChatGPT, Copilot, and Bard. Secondly, the readability and understandability of both websites and AI-powered chat tool responses will be evaluated from the perspective of patients.

Materials and Methods

The study was planned as a methodological study. This study does not require ethics committee approval as it does not involve human participants or sensitive personal data.

Search and accuracy analysis

A search was conducted on the Google search engine using the keywords "pregnancy radiology exposure frequently asked questions." The first 20 websites that appeared in the search results were evaluated. Non-profit associations and organizations' websites that provided answers to questions related to radiation during pregnancy in a question-and-answer format were identified. The questions and answers within these websites were recorded, totaling 15 questions. These questions were individually posed to ChatGPT, Copilot, and Bard Chat tools. Due to the premium membership requirement of ChatGPT 4o, it was not preferred, and ChatGPT 4 was chosen as it is freely accessible to everyone and easily accessible for patients. To avoid influencing the dialogue, each question was asked in a new tab, and the responses were recorded. The answers to each question were randomly recorded in a single document without indicating the source. All parameters related to the answers were evaluated by an observer with no knowledge of the sources, possessing 8 years of experience in the field of radiology and 4 years of ongoing work in medical education (Appendix A).

Appendix A. Websites accessed as a result of an internet search

1	https://www.iaea.org/resources/rpop/health-professionals/radiology/pregnant-women
2	https://hps.org/physicians/radiology_pregnant_patient_qa.html
3	https://www.aafp.org/pubs/afp/issues/1999/0401/p1813.html
4	https://www.cdc.gov/nceh/radiation/emergencies/prenatalphysician.htm
5	https://www.beaumont.org
6	https://4rai.com/2016/06/08/radiology-imaging-and-pregnancy-what-you-need-to-know/
7	https://www.cdc.gov/nceh/radiation/pregnant-and-children.html
8	https://www.epa.gov/radiation/frequent-questions-radiation-medicine
9	https://www.mayoclinic.org/healthy-lifestyle/pregnancy-week-by-week/expert-answers/x-ray-during-pregnancy/faq-20058264
10	https://www.ehs.washington.edu/radiation/radiation-exposure-during-pregnancy
11	https://www.acog.org/clinical/clinical-guidance/committee-opinion/articles/2017/10/guidelines-for-diagnostic-imaging-during-pregnancy-and-lactation
12	https://ehs.virginia.edu/Radiation-Safety-Pregnancy.html
13	https://www.aapm.org/org/policies/documents/CARES_FAQs_Patient_Shielding.pdf
14	https://www.purdue.edu/ehps/rem/laboratory/Personal/preg.html
15	https://aci.health.nsw.gov.au/networks/eci/clinical/ed-factsheets/medical-imaging-in-pregnancy
16	https://www.sor.org/learning-advice/professional-body-guidance-and-publications/documents-and-publications/policy-guidance-document-library/inclusive-pregnancy-status-guidelines-for-ionising
17	https://www.rch.org.au/kidsinfo/fact_sheets/Radiation_and_pregnancy/
18	https://www.acr.org/-/media/acr/files/practice-parameters/pregnant-pts.pdf
19	https://www.insideradiology.com.au/radiation-risk-preg/
20	https://kidshealth.org/en/parents/xray-pregnancy.html

The accuracy of the answers was assessed on a scale of 2 for "completely correct," 1 for "partly correct," 0 for "incorrect," and "no answer." The methodology of the article has been illustrated in Figure 1.

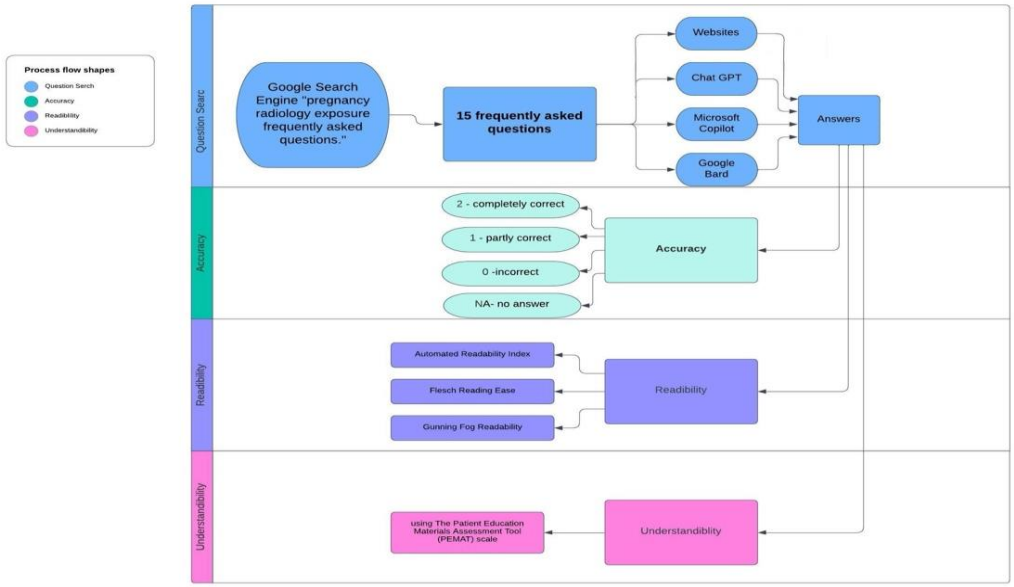

Figure 1. The methodology of the article has been illustrated

Readability scores assessment

Readability is related to how easy or difficult it is to read a text. Readability depends on factors such as the presentation of a text (for example, font choice, font size, spacing, or colors) and its context (such as syllables, words, and sentences on the page). Various scoring systems have been developed to assess readability.[17-19] The readability scores of the answers were evaluated using free web-based tools.[20] For each answer, Automated Readability Index (ARI), Flesch Reading Ease (FRE), and Gunning Fog Readability (GFR) scores were calculated. ARI score calculates readability based on the average number of characters per word and the average number of words per sentence [ARI = (4.71 * (characters/words)) + (0.5 * (words/sentences)) - 21.43]. The target is to obtain a score between 1 and 14 on this index, corresponding to the 14 grades in the United States education system. In some cases, scores above 14 can also be obtained, in which case the level is classified as college graduate.[17] FRE score is calculated based on the total number of words, sentences, and syllables in a text [FRE = (206.835 - (1.015 * (words/sentences)) - (84.6 * (syllables/words))]. According to this index, the most readable texts are those with shorter sentences and words. The scores obtained from the

FRE index are expressed as 90-100 very easy, 80-89 easy, 70-79 fairly easy, 60-69 standard, 50-59 fairly difficult, 30-49 difficult, and 0-29 very confusing.[18] GFR score is calculated based on the average number of words per sentence and the percentage of complex words in the text [GFR = 0.4 * ((words/sentences) + (percentage of complex words))]. Complex words are typically defined as words with three or more syllables or technical jargon that may be difficult for readers to understand. In this scoring, scores ranging from 6 (sixth grade) to 20 (postgraduate plus) are obtained.[19] (Figure 2).

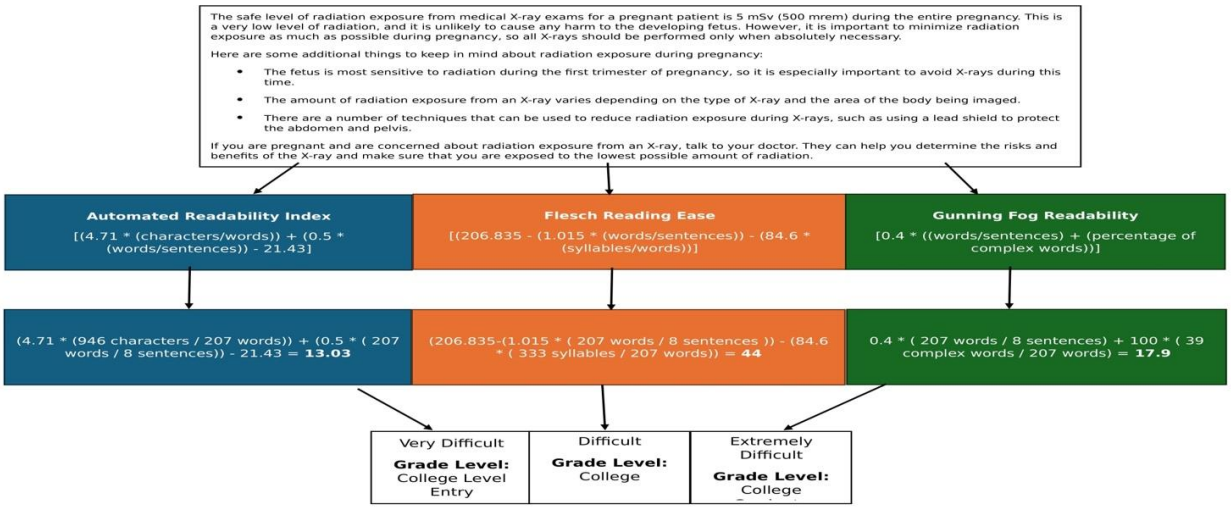

Figure 2. The calculation of readability scores for Google Bards response to the first question

Understandability score assessment

From the perspective of patients, understandability is defined as the ability of patient education materials to be understood and expressed by individuals with diverse backgrounds and varying levels of health literacy. Scales have been developed to assess this, and the most important scale used is The Patient Education Materials Assessment Tool (PEMAT).[21] Therefore, in the study, responses have been evaluated in terms of understandability using the PEMAT scale. This scale is designed to evaluate the understandability and actionability of patient education materials. It has different instruments for written and audio-visual materials.[21] In this section, considering the answers as patient education materials, the understandability scores for written materials were calculated using the PEMAT scale. The actionability part was not evaluated since it was deemed inappropriate for the study questions, and the scale includes items related to the necessity of visual materials.

Data Analysis

Statistical analysis was conducted using SPSS, version 25.0, for Windows (SPSS Inc., Chicago, IL, USA). Descriptive statistics were expressed as mean and standard deviation for continuous data and as percentages for categorical data. The Shapiro-Wilk test was employed to assess normality, and the Levene test was used to examine the homogeneity of variance. One-way ANOVA was applied to detect interactions between variables, and post-hoc tests were performed to conduct pairwise comparisons between groups. The significance level (p-value) was set at 0.05 (95% confidence interval) for all tests.

Results

Accuracy results

Through keyword-based searches on the Google search engine, four websites providing information on this topic in a question-answer format were found on the first two pages of the search results.[22-25] When excluding repeated questions from these websites, a total of 15 questions were included in the study. The questions are provided in Table 1. When assessing the accuracy of the answers to these questions, 100% (15/15) of the websites' answers, 66.67% (10/15) for ChatGPT, 73.33% (11/15) for Copilot, and 93.33% (14/15) for Bard were considered completely correct. The findings for each answer are summarized in Appendix B.

Table 1. Questions reached as a result of the website query

Q	Questions
1	Is there a safe level of radiation exposure from medical x-ray exams for a pregnant patient?
2	Can radiation cause birth defects?
3	What are the possible effects of X-rays (for fetus)*?
4	What is the "10-day rule" (about pregnancy-related radiation exposure)*?
5	How safe are x-ray exams of the chest and extremities during pregnancy?
6	What if a patient underwent an abdominal or pelvic x-ray exam before realizing that she was pregnant?
7	Is it necessary to put a lead apron over my abdomen for X-ray exams (when pregnant)*?
8	Can I undergo a CT scan while I am pregnant?
9	Is it important to know if I am pregnant or undergoing a CT scan?
10	What if a patient underwent an abdomen CT before realizing that she is pregnant?
11	Can cardiac catheterization be performed on a pregnant patient?
12	Are there recommendations regarding the termination of pregnancy after radiation exposure?
13	Can a patient become sterile after undergoing a diagnostic x-ray examination?
14	What are the effects of radiation exposure in utero?
15	Is it okay to have an MRI exam for back pain when pregnant?

*While the expressions in brackets are not expressed on websites, they are added to ensure the integrity of meaning in chat tools.

Appendix B. Accuracy and Global Quality Score scores of the answers

	Accuracy			
Questions	**ChatGPT**	**Copilot**	**Bard**	**Websites**
1	1	2	1	2
2	2	2	2	2
3	2	2	2	2
4	0	2	2	2
5	1	2	2	2
6	2	2	2	2
7	2	2	2	2
8	1	2	2	2
9	2	1	2	2
10	2	1	2	2
11	2	2	2	2
12	1	1	2	2
13	2	2	2	2
14	2	2	2	2
15	2	1	2	2
Mean	1.6	1.73	1.93	2

Readability scores

Readability scores for all answers are summarized in Table 2. When the average of the readability scores for all three measures was considered, the difficulty level was ranked as ChatGPT, Copilot, websites, and Bard. In ARI scores for ChatGPT were higher than Bard and websites' answers ($p < 0.001$, $p = 0.001$, respectively), and scores for Copilot were also statistically higher than Bard scores ($p = 0.004$). For FRE scores for ChatGPT were lower than Bard and website answers ($p < 0.001$, $p < 0.001$, respectively). FRE scores differ from the others, with an increase indicating easier readability. For GFR scores for ChatGPT were higher than Bard and website answers ($p < 0.001$, $p < 0.001$, respectively), while Copilot scores were also higher than Bard scores ($p = 0.014$) (Figure 3).

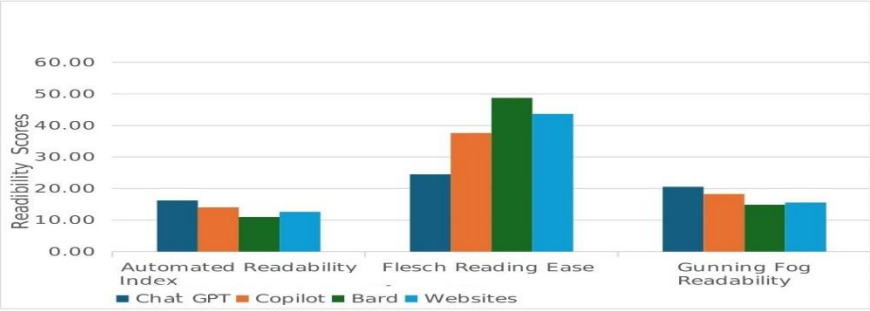

Figure 3. Comparison of readability scores among Chat Tools

Table 2. Readability scores of the answers.

Questions	ChatGPT			Copilot			Bard			Websites		
	ARI	FRE	GFR	ARI	FRE	GFR	ARI	FRE	GFR	ARI	FRE	GFR
1	17,56	12	23,9	19,20	15	23,7	13,03	44	17,9	14,30	30	18,90
2	17,78	18	20,90	12,01	48	17,20	9,05	49	13,6	9,62	65	11,2
3	15,20	19	19,00	10,54	57	14,60	6,66	71	10,4	17,25	34	17,8
4	13,90	46	16,70	17,17	28	21,10	11,15	48	15,3	14,19	36	18,1
5	15,20	27	20,30	11,36	51	15,30	12,03	46	16	9,64	45	13,9
6	16,26	24	20,9	14,1	36	19,8	10,2	45	14,9	10,49	49	14,33
7	15,19	32	18,7	14,9	47	17	13,3	42	16,1	12,14	51	14,9
8	15,78	37	20,8	11,3	53	16,1	9,69	61	13,6	11,62	52	14
9	14,82	35	18,2	11,2	48	15,7	12,2	52	16,3	11,16	38	16,8
10	17,99	20	22,3	16,5	31	20,4	12	57	13,6	10,23	49	14,5
11	19,75	9	24	15,1	12	21,4	10,1	39	15,5	16,24	39	18,7
12	15,85	23	20,6	19,7	15	23,2	11,3	37	16	13,55	42	16,5
13	16,5	20	22,1	11,8	38	18,6	10,7	46	15,3	11,3	45	12,7
14	16,73	11	21,6	14,9	32	18,1	13,4	28	16,9	16,18	34	18
15	13,69	34	17,8	10,3	53	11,9	8,99	66	11,5	10,95	46	13,1
Mean	16,15	24,47	20,52	14,00	37,60	18,27	10,92	48,73	14,86	12,59	43,67	15,56

ARI: Automated Readability Index, FRE: Flesch Reading Ease, GFR: Gunning Fog Readability

PEMAT understandability scores

When evaluating the PEMAT scores for the answers, the average scores were calculated as 66.13 for the websites, 66.43 for ChatGPT, 68.53 for Copilot, and 79.53 for Bard. The PEMAT scores for each answer are summarized in Table 3. The Bard PEMAT score was statistically significantly higher than websites, ChatGPT, and Copilot (p=0.014, p=0.014, p=0.010, respectively).

Table 3. PEMAT understandability scores of the answers.

Questions	ChatGPT	Copilot	Bard	Websites
1	64	91	73	67
2	56	70	82	78
3	82	56	82	56
4	No Answer	70	64	56
5	58	78	82	67
6	44	67	67	40
7	78	67	78	78
8	67	80	89	56
9	67	70	78	67
10	82	70	82	67
11	60	56	83	67
12	56	56	87	85
13	78	60	73	92
14	82	70	82	60
15	56	67	91	56
Mean	66,43	68,53	79,53	66,13

Discussion

In the study, when examining the accuracy of answers provided by chat tools to questions about the effects of radiation during pregnancy, it was found that the accuracy and quality of the answers were high, similar to findings in the literature.[1,26,27] Literature studies have generally been conducted using ChatGPT. For instance, in a study related to basal cell carcinoma, the answers were evaluated as appropriate and inappropriate, with an appropriate answer rate of 84%.[1] Similarly, in a study on breast cancer prevention, this rate was 88%,[27], and in a study on cardiovascular disease prevention, it was 84%.[26] In this study, 66.67% of ChatGPT answers were evaluated as completely accurate. When an unanswered question was excluded, the appropriate answer rate was calculated as 93.33%, similar to the literature. In a study related to lung cancer, ChatGPT answered 70.8% of the questions completely accurately, similar to our study.[28] In the same study, Copilot had an accuracy rate of 61.7%, and Bard had 51.7% accuracy. In our study, Copilot also similarly provided completely accurate answers to 73.33% of the questions. However, in this study, the accuracy rate for Bard was 93.33%, which was higher than the literature.[28] This could be attributed to the use of Bard's experimental version in the literature study. Overall, considering all these data, it can be concluded that chat tools generally provide appropriate answers to patient questions, and most of the answers are correct. However, in addition to these findings, a study using ChatGPT for clinical radiological information indicated that the majority of references could not be found, and only a small portion of references contained correct information to answer the questions.[29] This suggests that caution should be exercised, particularly as the difficulty level of questions increases when dealing with chat tool responses.

An important finding of this study is that the readability scores of the chat tool responses were significantly higher than the community average at school grade level eight. This adversely affects the comprehensibility of responses on chat tools and websites from the users' perspective. In a conducted study comparing ChatGPT and websites in terms of readability, ChatGPT was found to have statistically higher GFR scores. Additionally, the study stated that both websites and ChatGPT had readability scores above the average.[15] In an evaluation of patient-oriented information on abdominal aortic aneurysms obtained from websites, a study found an average FRE score of 39, considered difficult.[30] Various studies investigating readability scores have consistently found them to be high for the general population.[11-13] In this study, the average FRE scores range from 24.47 to 48.73, categorizing these values as difficult and very confusing. Another noteworthy finding in this study is that, despite high scores on all platforms, ChatGPT has more unfavorable scores in all three readability systems, especially when compared to websites and Bard. Considering that chat tools acquire information from the internet, these findings may not be surprising; however, it was observed that Bard presented information in a relatively more readable manner.

The answers provided to the questions are aimed at increasing the patient's level of knowledge. In this context, these responses can be considered as patient education materials. Accordingly, in the study, the PEMAT scores of the responses were calculated. A threshold of 70% has previously been suggested for the material to be considered suitable for patient education.[21] There are limited studies in the literature conducted from this perspective. In a study related to obstructive sleep apnea, ChatGPT and Bard were compared, and they were evaluated with PEMAT in terms of understandability. ChatGPT scores were found to be statistically higher than Bard's.[14] In another study, ChatGPT and website data were compared, and although the understandability scores were above 70%, they were found to be higher for websites.[15] In this study, PEMAT understandability scores for ChatGPT, Copilot, and websites were similar and remained below the 70% threshold. For Bard, the PEMAT score was calculated as 79.5%, which is statistically higher compared to the others. Although there are not many studies conducted on this topic in the literature, it is noteworthy that different results have been obtained in studies focusing on understandability.

This study has some limitations. Firstly, the accuracy of responses and PEMAT scores were evaluated by a single observer, even though experienced, which could introduce bias. Secondly, due to the nature of chat tools, they do not provide the same response each time. Therefore, repeated answers to the same questions may yield different results in some parameters. Lastly, ChatGPT, Bard, and Copilot have a constraint as they lack image processing capabilities for both inputs and outputs. As a result, all responses in this study received a score of zero on PEMAT questions assessing visual aids, limiting the highest achievable PEMAT understandability score to 92%.

The study evaluated data from websites and three commonly used chat tools. Similar to the literature, the responses from all platforms to the questions were largely appropriate, and most questions were answered completely accurately. Bard is identified as superior in terms of readability and understandability compared to other information sources. However, considering that many patients and their relatives perceive the internet as a source of information, our results indicate that responses from chat tools and websites are not suitable for patients in terms of readability and understandability. In this regard, chat tools and internet information sources should be developed, especially to ensure that the content is understandable to the majority of readers and to provide equal access to health information. While these platforms may have potential in terms of accuracy, it should be kept in mind that they may produce content that is deficient and difficult to understand in terms of patient education.

Ethical Considerations: The study was planned as a methodological study. This study does not require ethics committee approval as it does not involve human participants or sensitive personal data.

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

References

1. Trager MH, Queen D, Bordone LA, et al. Assessing ChatGPT responses to common patient queries regarding basal cell carcinoma. *Arch Dermatol Res.* 2023;315:2979-81 (doi: 10.1007/s00403-023-02705-3).
2. Ibrahim H, Liu F, Asim R, et al. Perception, performance, and detectability of conversational artificial intelligence across 32 university courses. *Sci Rep.* 2023;13:12187 (doi: 10.1038/s41598-023-43998-8).
3. Farina M, Lavazza A. ChatGPT in society: emerging issues. *Front Artif Intell*. 2023;6 (doi: 10.3389/frai.2023.1130913).
4. Kantor J. ChatGPT, large language models, and artificial intelligence in medicine and health care: A primer for clinicians and researchers. *JAAD Int*. 2023;13:168-9 (doi: 10.1016/j.jdin.2023.07.011).
5. Van Riel N, Auwerx K, Debbaut P, et al. The effect of Dr Google on doctor-patient encounters in primary care: a quantitative, observational, cross-sectional study. *BJGP Open* 2017;1:bjgpopen17X100833 (doi: 10.3399/bjgpopen17X100833).
6. Santana S, Lausen B, Bujnowska-Fedak M, et al. Informed citizen and empowered citizen in health: results from a European survey. *BMC Fam Pract*. 2011;12:20 (doi: 10.1186/1471-2296-12-20).
7. Schwartz KL, Roe T, Northrup J, et al. Family medicine patients' use of the Internet for health information: a MetroNet study. *J Am Board Fam Med.* 2006;19:39-45 (doi: 10.3122/jabfm.19.1.39).
8. Lee K, Hoti K, Hughes JD, et al. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. *J Med Internet Res.* 2014;16:e262 (doi: 10.2196/jmir.3706).
9. Al Shboul MKI, Alwreikat A, Alotaibi FA. Investigating the use of ChatGPT as a novel method for seeking health information: A qualitative approach. *Sci Technol Libr*. 2023:1-10 (doi: 10.1080/0194262X.2023.2250835).
10. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: A systematic review of indicators, criteria, tools, and evaluation results. *J Assoc Inf Sci Technol.* 2015;66:2071-84 (doi: 10.1002/asi.23311).
11. Doğan L, Özçakmakcı GB, Yılmaz IE. The Performance of Chatbots and the AAPOS Website as a Tool for Amblyopia Education. *J Pediatr Ophthalmol Strabismus*. 2024;25:1-7 (doi: 10.3928/01913913-20240409-01).
12. Tepe M, Emekli E. Assessing the Responses of Large Language Models (ChatGPT-4, Gemini, and Microsoft Copilot) to Frequently Asked Questions in Breast Imaging: A Study on Readability and Accuracy. *Cureus.* 2024;16(5):e59960. (doi: 10.7759/cureus.59960).

13. Tepe M, Emekli E. Decoding medical jargon: The use of AI language models (ChatGPT-4, BARD, microsoft copilot) in radiology reports. Patient Educ Couns. 2024;126:108307. (doi: 10.1016/j.pec.2024.108307.)
14. Cheong RCT, Unadkat S, Mcneillis V, et al. Artificial intelligence chatbots as sources of patient education material for obstructive sleep apnoea: ChatGPT versus Google Bard. *Eur Arch Otorhinolaryngol*. 2023;2 (doi: 10.1007/s00405-023-08319-9).
15. McCarthy CJ, Berkowitz S, Ramalingam V, et al. Evaluation of an artificial intelligence chatbot for delivery of IR patient education material: A comparison with societal website content. *J Vasc Interv Radiol*. 2023:34;1760-8 (doi: 10.1016/j.jvir.2023.05.037).
16. Wiles R, Hankinson B, Benbow E, et al. Making decisions about radiological imaging in pregnancy. *BMJ.* 2022;377:e070486 (doi: 10.1136/bmj-2022-070486).
17. Joseph P, Silva NA, Nanda A, et al. Evaluating the Readability of Online Patient Education Materials for Trigeminal Neuralgia. *World Neurosurg.* 2020;144:e934-e938 (doi: 10.1016/j.wneu.2020.09.123).
18. Eleyan D, Othman A, Eleyan A. Enhancing Software Comments Readability Using Flesch Reading Ease Score. *Information*. 2020;11:430. (doi: 10.3390/info11090430).
19. Gunning R. The Fog index after twenty years. *Int J Bus.* 1969;6:3-13 (doi: 10.1177/002194366900600202).
20. Readability formulas [Internet]. https://readabilityformulas.com/readability-scoring-system.php (Accessed: 22.10.2023).
21. Shoemaker SJ, Wolf MS, 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).
22. Radiology and the pregnant patient Q&A. Health Physics Society. [Internet]. https://hps.org/physicians/radiology_pregnant_patient_qa.html (Accessed: 15.10.2023).
23. X-Rays, pregnancy and you. Food and Drug Administration. https://www.fda.gov/radiation-emitting-products/medical-x-ray-imaging/x-rays-pregnancy-and-you (Accessed: 15.10.2023).
24. Pregnancy and radiation - what patients need to know. International Atomic Energy Agency. https://www.iaea.org/resources/rpop/patients-and-public/pregnant-women (Accessed: 15.10.2023).
25. Radiation exposure during pregnancy. American Academy of Family Physicians. https://www.aafp.org/pubs/afp/issues/2010/0901/p494.html (Accessed: 15.10.2023).
26. Sarraju A, Bruemmer D, Van Iterson E, et al. Appropriateness of Cardiovascular Disease Prevention Recommendations Obtained From a Popular Online Chat-Based Artificial Intelligence Model. *JAMA*. 2023:329;842-4 (doi: 10.1001/jama.2023.1044).
27. Haver HL, Ambinder EB, Bahl M, et al. Appropriateness of Breast Cancer Prevention and Screening Recommendations Provided by ChatGPT. *Radiology*. 2023;307:e230424. doi: 10.1148/radiol.230424.

28. Rahsepar AA, Tavakoli N, Kim GHJ, et al. How AI responds to common lung cancer questions: ChatGPT vs Google Bard. *Radiology*. 2023;307:e230922 (doi: 10.1148/radiol.230922).
29. Wagner MW, Ertl-Wagner BB. Accuracy of information and references using ChatGPT-3 for retrieval of clinical radiological information. *Can Assoc Radiol J*. 2023:8465371231171125 (doi: 10.1177/08465371231171125).
30. Bailey,MA, Coughlin PA, Sohrabi S, et al. Quality and readability of online patient information for abdominal aortic aneurysms. *J Vasc Surg.* 2012;56:21-6. (doi: 10.1016/j.jvs.2011.12.063).