Lymphology 59 (2026) 27-34

DIGITAL EXPERTS IN LYMPHEDEMA: ASSESSING THE QUALITY AND READABILITY OF RESPONSES FROM CHATGPT AND GEMINI

C.S. Pirincci, E. Cihan, E. Uzelpasacı, A.B. Pacacı

University of Health Sciences, Gulhane Faculty of Physiotherapy and Rehabilitation, Ankara, Turkey (CSP,EU,ABP); Selcuk University, Vocational School of Health Science Physiotherapy, Konya, Turkey (EC).

ABSTRACT

This study evaluated the quality and readability of responses provided by ChatGPT and Gemini to frequently asked questions related to lymphedema. Ten frequently asked questions about lymphedema were selected by expert therapists for submission to ChatGPT and Gemini. The initial responses from ChatGPT and Gemini were recorded without follow-up queries. Five independent experts (therapists) specialized in this field evaluated responses from ChatGPT and Gemini using a four-point rating scale. Readability levels were analyzed using the Flesch-Kincaid Grade Level through WordCalc software. Results indicated that except for questions 3 and 6, ChatGPT and Gemini provided similar responses, with significant differences observed in those two questions (p = 0.014). ChatGPT's answers to questions 1, 2, 3, 4, 8, 9, and 10 were found to be more readable than Gemini's, while Gemini's answers to questions 5, 6, and 7 were more readable. Overall, approximately 70% of ChatGPT's responses were found to be easier to read compared to Gemini. The results of this study indicate that both AI search engines offer comparable responses to research questions. However, when evaluating clarity and accessibility of the answers, ChatGPT was found to be more understandable and user-friendly than Gemini.

Keywords: Artificial intelligence; Cancer; ChatGPT; Gemini; Lymphedema

INTRODUCTION

Breast cancer is the most common type of cancer in women worldwide, and in Turkey, it accounts for 47.7% of all cancers in women (1, 2). Survival rates among breast cancer patients in developed countries have improved in recent decades (3). However, as women live longer after a breast cancer diagnosis, they may experience side effects of cancer treatment, such as breast cancer-related lymphedema (4).

Breast cancer-related lymphedema is characterized by abnormal accumulation of protein-rich fluid in the interstitial tissues of the arm, hand, or chest wall, often developing as a consequence of breast cancer surgery and/or radiation therapy. Symptoms of lymphedema can emerge at any point following breast cancer treatment, with several risk factors contributing to its onset (5). These include a higher number of lymph nodes removed, a larger primary tumor size, multi-field radiation exposure, elevated body mass index (BMI), and tumors located in the upper outer quadrant of the breast (6). Lymphedema not only affects physical health but also impairs daily activities, diminishes quality of life, and negatively impacts psychological well-being (7-9).

Previous investigations have shown that

breast cancer patients often have limited awareness of lymphedema, emphasizing the need for better education on risk factors and preventive strategies, particularly at diagnosis or in the early years following treatment (10, 11). Recent studies suggest that lymphedema knowledge may play a crucial role in influencing the risk of developing the condition, indicating that higher awareness could potentially reduce its occurrence (12,13). Lymphedema patients can obtain information about their condition from various sources, including healthcare professionals, patient support groups, books, brochures, and online resources or websites (14-16). Although the exact rate at which patients seek out these sources is unclear, they play a crucial role in enhancing patients' understanding of lymphedema, its management, and preventive strategies. However, the accuracy and reliability of the information may vary, highlighting the importance of guidance from healthcare professionals (16).

Patient use of online resources for health information is increasing, especially with the recent growing integration of artificial intelligence (AI) (17,18). AI can be defined as a system's capability to accurately analyze external data, derive insights from it, and apply that knowledge to accomplish specific objectives and tasks by adapting dynamically (19). Although studies in the literature have explored the use of AI for patient education in conditions such as hypertension, obstructive sleep apnea, and aortic stenosis (20-22), to the best of our knowledge, no research has been conducted specifically for lymphedema patients. Therefore, it is essential to assess the readability, accuracy, and reliability of information obtained from AI regarding the diagnosis, prevention, and treatment of lymphedema in lymphedema patients. The aim of this study was to compare the information obtained from AI with that provided by healthcare professionals in order to evaluate the accuracy, quality, and readability of AI-generated health content concerning the diagnosis, prevention, and treatment of lymphedema. The findings of this study may support the safe, comprehensible, and evidence-based use of AI in patient education, particularly ensuring that individuals with lymphedema have access to high-quality and reliable health information.

MATERIALS AND METHODS

Question Collection Process

As an initial step in question collection, various artificial intelligence (AI) chatbots were analyzed. The study included ChatGPT-4 Turbo (OpenAI) and Gemini 2.0 Flash (Google). To minimize potential bias and accurately simulate public inquiries, new accounts were created. A non-corporate computer was utilized, and browser history was cleared prior to data collection. Each question was entered in a newly initiated chat session in both ChatGPT and Gemini to prevent AI models from referencing prior interactions. Additionally, the "temporary chat" feature in ChatGPT was enabled to disable conversation history for each query. To further eliminate potential biases related to AI recognition of repeated inputs, all questions were manually typed rather than copied and pasted.

Question Collection

On February 26, 2025, the AI-based chatbot systems were given the prompt: "Can you list the top 20 questions that people with lymphedema ask you around the world?" This method was chosen to identify frequently asked questions by leveraging the large-scale data access and user interaction capabilities of AI. Rather than relying solely on clinical experience, this AI-driven approach was intended to reflect global trends and common information needs among individuals with lymphedema. However, it is acknowledged that this method may carry an inherent risk of bias due to the nature of the AI training data.

The initial responses from both chatbots were recorded, and the resulting questions were consolidated into a single question pool. Two physiotherapist researchers then independently reviewed the pool. Based on predetermined criteria—namely, the frequency of question occurrence across both AI outputs, clinical relevance, and applicability to patient

TABLE 1
Research Questions Generated by AI and Finalized for Study

No	Question
1	What is lymphedema and why does it occur?
2	What are the symptoms of lymphedema?
3	How is lymphedema diagnosed?
4	Are there different stages of lymphedema?
5	Is there a definitive treatment for lymphedema?
6	What methods are used to treat lymphedema?
7	How is manual lymphatic drainage for lymphedema performed?
8	How is complex decongestive therapy applied?
9	What are lymphedema exercises and how are they performed?
10	Can lymphedema be treated surgically?

education—10 questions were selected. This dual-review strategy aimed to ensure that the final selection represented both widespread user concerns identified by AI and clinically meaningful topics relevant to physiotherapy practice (*Table 1*).

Response Collection Process

The questions determined by the researchers were entered into the chat interface of the AI chatbots. The initial responses provided for each question were recorded without generating additional prompts or requesting further clarification.

Evaluation

A research team of five physiotherapists was involved in the assessment phase. To assess the quality of the responses, the 4-point system proposed by Mika et al. was used (23). The scale is rated from 1 to 4, with 1 indicating a perfect answer and 4 representing an inadequate response.

Ethics Committee

Our study did not involve any human research and does not require 'Ethics Committee Approval' due to the design of the study (24, 25).

Statistical Analysis

Statistical analyses were performed using SPSS (IBM Corp., Armonk, NY, USA). Qualitative scores obtained from a 4-point Likert-type scale were treated as ordinal data. Comparisons between Gemini and ChatGPT were conducted using the Mann–Whitney U test due to the absence of a normal distribution and the small sample size. A p-value of <0.05 was considered statistically significant for all tests.

The readability level of the responses was assessed using the WordCalc online readability test (https://www.wordcalc.com/readability/). Readability was assessed using the Flesch-Kincaid Grade Level, a widely used readability metric that estimates the educational level required to understand a text based on sentence

TABLE 2
Responses of Artificial Intelligence Applications to Research Questions

No	Quality of answer				z	p
	CHATGPT		GEMINI			
	Median	IQR	Median	IQR		
1	1.00	1.00	1.00	1.00	0.000	1.000
2	1.00	0.50	2.00	0.50	-1.800	0.072
3	2.00	0.50	1.00	0.00	-2.449	0.014*
4	1.00	1.00	2.00	1.00	-0.600	0.549
5	2.00	0.50	2.00	1.00	-0.516	0.606
6	3.00	0.50	2.00	0.00	-2.449	0.014*
7	2.00	0.50	2.00	0.50	0.000	1.000
8	2.00	0.00	2.00	0.00	0.000	1.000
9	2.00	0.00	3.00	1.00	-1.964	0.050
10	2.00	1.00	2.00	0.00	-1.500	0.134

length and word complexity. Higher grade levels indicate increased textual complexity and lower readability. This metric was selected due to its widespread use in biomedical and educational research and its suitability for objectively comparing readability across large language model responses. To minimize potential scoring bias during the evaluation, outputs from ChatGPT and Gemini were anonymized before being submitted for evaluation.

RESULTS

Results demonstrated that except for questions 3 and 6 in which significant differences were obtained [question 3 (z = -2.449, p = 0.014) and question 6 (z = -2.449, p = 0.014)], both artificial intelligence models provided similar answers. *Table 2* displays responses to the questions.

Results also demonstrated that readability scores of ChatGPT-4's answers to questions 1, 2, 3, 4, 8, 9 and 10 were lower than Gemini's, while the readability scores of Gemini's answers to questions 5, 6 and 7 were lower than ChatGPT's. Approximately 70% of the responses provided by ChatGPT were found to be easier to read compared to those of Gemini. The Fleche-Kincaid Degree levels of the questions are shown in *Figure 1.*

DISCUSSION

Findings of this study show that both AI search engines provide similar answers to the research questions. In addition, in terms of the comprehensibility levels of the answers given to the questions, ChatGPT was found to be more comprehensible and accessible than Gemini. In the current literature, responses of artificial intelligence models in terms of comprehensibility and explainability are addressed in various dimensions. Pirkle et al. in pediatric orthopedic cases found that ChatGPT and

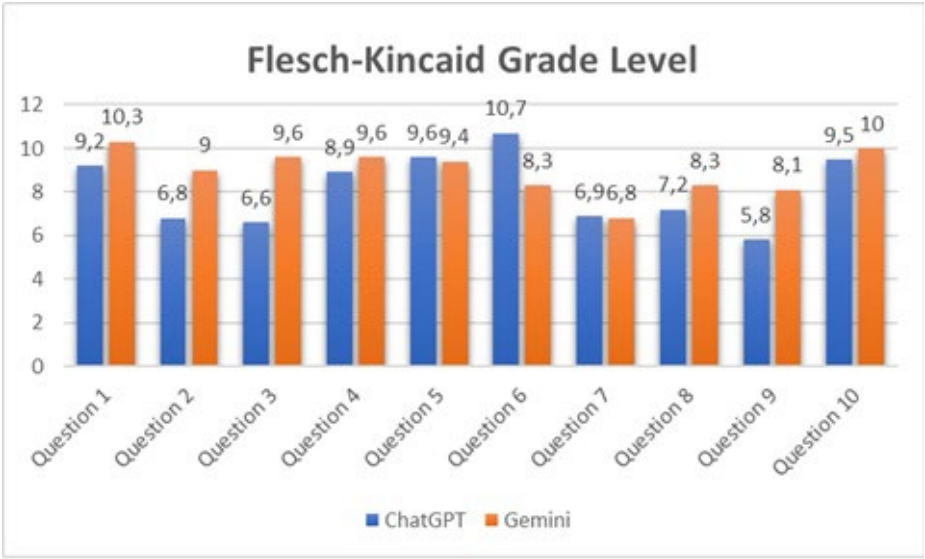

Fig. 1. *Results of the Flesch-Kincaid Grade Level analysis for readability of responses to each of the 10 questions.*

Gemini gave similar answers to the Clinical Practice Guidelines of the Academy of Orthopedic Surgeons (26). In the study conducted by Abdul Sami et al. in pediatric radiology, it was concluded that ChatGPT's answers to the questions were more accurate than Gemini (27). Another study demonstrated that overall performance of ChatGPT in providing information to caregivers of pediatric cancer patients was higher than Google Bard (Gemini Pro), Microsoft Bing Chat, and Google SGE, and that it could contribute to the information acquisition processes of caregivers of pediatric cancer patients (28). Engin et al. evaluated the effectiveness of ChatGPT, BingAI, and Gemini in informing families about retinopathy of prematurity and found that ChatGPT achieved the highest score in terms of accuracy and reliability, while Gemini had a high score in terms of readability, but its reliability was considered lower (29). In another study, 98 oncological cases were retrospectively analyzed and four main decision points such as the need for additional imaging studies, chemotherapy, surgery and radiotherapy were determined. ChatGPT and Gemini showed moderate agreement with the multidisciplinary tumor board in surgical and radiotherapy decisions. ChatGPT also showed moderate agreement on chemotherapy decisions, while Gemini showed lower agreement (30). In general, ChatGPT was found to be superior to Gemini on most key points. These studies demonstrate that when performances of search engines in different patient populations are analyzed in the literature, it is found that artificial intelligence models have an important potential in providing medical information and decision making for patients and their relatives. However, it is noteworthy that there are differences in the accuracy, reliability, and readability of different search engines.

CONCLUSION

This study found that both artificial intelligence search engines gave similar answers to research questions. Regarding the comprehensibility of the answers, it was concluded that ChatGPT was higher than Gemini for many questions. As a result, our study draws attention with the feature of being the first study in the field of lymphedema. It is concluded that lymphedema patients and their relatives may prefer ChatGPT and Gemini to have information about this disease, but the information obtained should not be used alone in medical decision making.

Limitations

This study represents the first attempt to compare the quality and readability of responses from ChatGPT and Gemini, two advanced artificial intelligence systems, concerning lymphedema. However, our study has several limitations. The analysis focused on the 10 most frequently asked questions about lymphedema posed to ChatGPT and Gemini. These questions may not comprehensively reflect all the needs and concerns of lymphedema patients and in the future a broader question pool could be developed. As these AI models rely on their respective datasets, the responses are limited by the data on which both models were trained. Furthermore, the analysis was conducted based on general responses, without considering the underlying causes of the disease or patient profiles. In additionally inter-rater reliability (IRR) was not calculated in this study. The fact that only initial responses were evaluated (whereas users might ask follow-up questions) is another potential limitation. Although the inclusion of a five-member expert panel in the study is close to the upper limit of numerical standards in the literature (31,32), the fact that the experts came from similar clinical disciplines may have created a disadvantage. This situation may have resulted in a limited diversity of perspectives, particularly regarding lymphedema, where multidisciplinary views are of critical importance in diagnostic and management processes.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this study, the authors used ChatGPT and Gemini to identify questions and receive responses to the questions asked. After using this tool/service, the authors revised the content as needed, but the responses were not edited as the purpose of this study was to evaluate the effectiveness of ChatGPT and Gemini in the first place, as described in the paper. The authors take full responsibility for the content of the publication. AI support was utilized for text language correction.

FUNDING

The authors received no financial support for the research and/or authorship of this article.

CONFLICT OF INTEREST

All authors declare no financial conflicts of interest exist with respect to the authorship and/or publication of this article.

REFERENCES

1. Sung, H, J Ferlay, RL Siegel, et al: Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer. J. Clin. 71 (2021), 209-249. doi: 10.3322/caac.21660
2. Türkyılmaz, MO, A Kavak, S Ergun, et al: Türkiye Kanser İstatistikleri (2017). Ankara: Türkiye Cumhuriyeti Sağlık Bakanlığı, Halk Sağlığı Genel Müdürlüğü (2018). Available from:https://hsgm.saglik.gov.tr/depo/birimler/kanserdb/istatistik/Turkiye_Kanser_Istatistikleri_2017.pdf.
3. Kerr, AJ, D Dodwell, P McGale, et al: Adjuvant and neoadjuvant breast cancer treatments: A systematic review of their effects on mortality. Cancer. Treat. Rev. 105 (2022), 102375. doi: 10.1016/j.ctrv.2022.102375
4. Bernas, M, SR Thiadens, P Stewart, et al: Secondary lymphedema from cancer therapy. Clin. Exp. Metastasis 39 (2022), 239-247. doi: 10.1007/s10585-021-10096-w
5. Shen, A, Q Lu, X Fu, et al: Risk factors of unilateral breast cancer-related lymphedema: An updated systematic review and meta-analysis of 84 cohort studies. Support. Care. Cancer 31 (2023), 18. doi: 10.1007/s00520-022-07508-2
6. Shen, A, Q Lu, L Zhang, et al: Risk factors of breast cancer-related lymphoedema: Protocol of an umbrella review. BMJ. Open 13 (2023), 070907. doi: 10.1136/bmjopen-2022-070907
7. Jørgensen, MG, NM Toyserkani, FG Hansen, et al: The impact of lymphedema on health-related quality of life up to 10 years after breast cancer treatment. NPJ. Breast. Cancer 7 (2021), 70. doi: 10.1038/s41523-021-00276-y

8. Pirinççi, CŞ, E Cihan, F Yaman. Assessing body awareness and upper extremity functionality in breast cancer survivors with and without lymphedema: A comparative analysis with healthy controls. Support. Care. Cancer. 33 (2025), 86. doi: 10.1007/s00520-024-09138-2
9. Orhan, C, S Özgül, G Nakip, et al: Meme kanseri tedavisiyle ilişkili lenfödemi olan hastalarda lenfödem şiddetinin yaşam kalitesi, üst ekstremite fonksiyonu ve fiziksel aktivite düzeyi üzerindeki etkileri. Anatol. Clin. J. Med. Sci. 24 (2019), 189-198. doi: 10.21673/anadoluklin.554019
10. Fu, MR, CM Chen, J Haber, et al: The effect of providing information about lymphedema on the cognitive and symptom outcomes of breast cancer survivors. Ann. Surg. Oncol. 17 (2010), 1847-1853. doi: 10.1245/s10434-010-0941-3
11. Kwan, ML, L Shen, JR Munneke, et al: Patient awareness and knowledge of breast cancer-related lymphedema in a large, integrated health care delivery system. Breast. Cancer. Res. Treat. 135 (2012), 591-602. doi: 10.1007/s10549-012-2199-x
12. Perdomo, M, C Davies, K Levenhagen, et al: Patient education for breast cancer–related lymphedema: A systematic review. J. Cancer. Surviv. 17 (2023), 384-398. doi: 10.1007/s11764-022-01262-4
13. Borman, P, A Yaman, S Yasrebi, et al: The importance of awareness and education in patients with breast cancer-related lymphedema. J. Cancer. Educ. 32 (2017), 629-633.
14. Deng, J, M Fu, J Armer, et al: Self-reported information sources and perceived knowledge in individuals with lymphedema. Lymphology 46 (2013), 173-183.
15. Ridner, SH: Pretreatment lymphedema education and identified educational resources in breast cancer patients. Patient. Educ. Couns. 61 (2006), 72-79. doi: 10.1016/j.pec.2005.02.009
16. Johnson, AR, AF Doval, SA Egeler, et al: A multimetric evaluation of online Spanish health resources for lymphedema. Ann. Plas. Surg. 82 (2019), 255-261. doi: 10.1097/SAP.0000000000001762
17. Seth, AK, CR Vargas, DJ Chuang, et al: Readability assessment of patient information about lymphedema and its treatment. Plast. Reconstr. Surg. 137 (2016), 287-295. doi: 10.1097/01.prs.0000475747.95096.ab
18. Eldaly, AS, FR Avila, RA Torres-Guzman, et al: Artificial intelligence and lymphedema: State of the art. J. Clin. Transl. Res. 8 (2022), 234.
19. Fritsch, SJ, A Blankenheim, A Wahl, et al: Attitudes and perception of artificial intelligence in healthcare: A cross-sectional survey among patients. Digit. Health 8 (2022). doi: 10.1177/20552076221116772
20. Almagazzachi, A, A Mustafa, AE Sedeh, et al: Generative artificial intelligence in patient education: ChatGPT takes on hypertension questions. Cureus 16 (2024). doi: 10.7759/cureus.53441
21. Cheong, RCT, S Unadkat, V McNeillis, et al: Artificial intelligence chatbots as sources of patient education material for obstructive sleep apnoea: ChatGPT versus Google Bard. Eur. Arch. Otorhinolaryngol. 281 (2024), 985-993. doi: 10.1007/s00405-023-08319-9
22. Rouhi, AD, YK Ghanem, L Yolchieva, et al: Can artificial intelligence improve the readability of patient education materials on aortic stenosis? A pilot study. Cardiol. Ther. 13 (2024), 137-147. doi: 10.1007/s40119-023-00347-0
23. Mika, AP, JR Martin, SM Engstrom, et al: Assessing ChatGPT responses to common patient questions regarding total hip arthroplasty. J. Bone. Joint. Surg. 105 (2023), 1519-1526. doi: 10.2106/JBJS.23.00209
24. Pan, A, D Musheyev, D Bockelman, et al: Assessment of artificial intelligence chatbot responses to top searched queries about cancer. JAMA. Oncol. 9 (2023), 1437-1440. doi: 10.1001/jamaoncol.2023.2947
25. Johnson, D, R Goodman, J Patrinely, et al: Assessing the accuracy and reliability of AI-generated medical responses: An evaluation of the ChatGPT model. Res. Sq. (2023), 2566942. doi: 10.21203/rs.3.rs-2566942/v1
26. Pirkle, S, J Yang, TJ Blumberg. Do ChatGPT and Gemini provide appropriate recommendations for pediatric orthopaedic conditions? J. Pediatr. Orthop. 45 (2025), 66-71. doi: 10.1097/BPO.0000000000002797
27. Abdul Sami, M, M Abdul Samad, K Parekh, et al: Comparative accuracy of ChatGPT 4.0 and Google Gemini in answering pediatric radiology text-based questions. Cureus 16 (2024), 70897. doi: 10.7759/cureus.70897
28. Sezgin, E, DI Jackson, AB Kocaballi, et al: Can large language models aid caregivers of pediatric cancer patients in information seeking? A cross-sectional investigation.

Cancer Med. 14 (2025), 70554. doi: 10.1002/cam4.70554
29. Durmaz Engin, C, E Karataş, T Öztürk. Exploring the role of ChatGPT-4, BingAI, and Gemini as virtual consultants to educate families about retinopathy of prematurity. Children 11 (2024). doi: 10.3390/children11060750
30. Hernández-Flores, LA, JB López-Martínez, JJ Rosales-de-la-Rosa, et al: Assessment of challenging oncologic cases: A comparative analysis between ChatGPT, Gemini, and a multidisciplinary tumor board. J. Surg. Oncol. 131 (2025) 1562-1570. doi: 10.1002/jso.28121
31. Lee, TC, K Staller, V Botoman, et al : ChatGPT Answers Common Patient Questions About Colonoscopy. Gastroenterology 165 (2023), 509-11. doi: 10.1053/j.gastro.2023.04.033
32. Merç, P, CŞ Pirinççi, E Cihan: Evaluation of AI chatbots for patient education and information on chronic obstructive pulmonary disease. Heart Lung 75 (2026), 21-25. doi: 10.1016/j.hrtlng.2025.09.002

Cansu Sahbaz Pirincci, Assoc. Prof., PT
University of Health Sciences,
Gulhane Faculty of Physiotherapy and Rehabilitation,
Ankara, Türkiye
Phone: +90 312 567 15 00
E-mail: cansusahbaz@gmail.com,
cansu.sahbazpirincci@sbu.edu.tr