Research Article / *Araştırma Makalesi*

Ege Journal of Medicine / *Ege Tıp Dergisi* 2026; 65 (3): 519-526

Evaluating the potential of GPT-4.1 mini and GPT-5 for postmastectomy lymphedema patient education: A physician-rated comparative study

Postmastektomi sonrası lenfödemde hasta eğitimine yönelik GPT-4.1 mini ve GPT-5 yanıtlarının kullanışlılık ve güvenilirlik açısından hekim değerlendirmesi

Benil Nesli Ata[1] Ece Cinar[2] Sibel Eyigor[2]
[1]Department of Physical Medicine and Rehabilitation, İzmir City Hospital, İzmir-Türkiye
[2]Department of Physical Medicine and Rehabilitation, Ege University School of Medicine, İzmir-Türkiye

ABSTRACT

Aim: To compare the reliability and usefulness of GPT-4.1 mini and GPT-5 in postmastectomy lymphedema patient education.

Materials and Methods: Twenty patient-centered questions were developed from reputable sources. Responses from both models were independently rated by three blinded physiatrists using 7-point Likert scales for reliability and usefulness.

Results: GPT-5 demonstrated higher usefulness and reliability scores compared to GPT-4.1 mini ($p = 0.005$ and $p = 0.021$, respectively).

Conclusions: Both models produced reliable and useful responses for lymphedema education. GPT-5 demonstrated higher usefulness and reliability scores, although these differences appear to reflect incremental improvements rather than substantial differences in clinical applicability. GPT-4.1 mini may already support basic patient education tasks. Future research should integrate patient usability data and real-time clinical resources to optimize AI-assisted rehabilitation.

Keywords: breast cancer, lymphedema, patient education, artificial intelligence, large language models.

ÖZ

Amaç: *Bu çalışmanın amacı, postmastektomiye bağlı lenfödemde hasta eğitimine yönelik GPT-4.1 mini ve GPT-5 büyük dil modellerinin güvenilirlik ve kullanışlılık açısından karşılaştırılmasıdır.*

Gereç ve Yöntem: *Güvenilir bilimsel ve klinik kaynaklardan yararlanılarak hasta odaklı yirmi soru oluşturulmuştur. Her iki model tarafından üretilen yanıtlar, güvenilirlik ve kullanışlılık düzeyleri açısından 7 puanlı Likert ölçeği kullanılarak üç kör fiziyatrist tarafından bağımsız olarak değerlendirilmiştir. Veriler Wilcoxon testi kullanılarak analiz edilmiştir.*

Bulgular: *GPT-5'in kullanışlılık ve güvenilirlik puanlarının GPT-4.1 mini'ye göre anlamlı derecede daha yüksek olduğu saptanmıştır (sırasıyla $p = 0,005$ ve $p = 0,021$).*

Sonuç: *Her iki büyük dil modeli de lenfödem hasta eğitimi için güvenilir ve kullanışlı yanıtlar üretmiştir. GPT-5 istatistiksel olarak anlamlı avantajlar göstermiş olmakla birlikte, bu farklılıkların klinik uygulanabilirlik açısından belirgin bir üstünlükten ziyade sınırlı iyileşmeleri yansıttığı düşünülmektedir. GPT-4.1 mini'nin temel hasta eğitimi gereksinimlerini karşılayabilecek potansiyele sahip olduğu görülmektedir. Gelecekte yapılacak çalışmaların hasta kullanım deneyimi ve gerçek zamanlı klinik kaynakları entegre etmesi önerilmektedir.*

Anahtar Sözcükler: *Meme kanseri, lenfödem, hasta eğitimi, yapay zekâ, büyük dil modelleri*

Corresponding author: Benil Nesli Ata
Department of Physical Medicine and Rehabilitation, Izmir City Hospital, Izmir-Türkiye
E-mail: *drbenilnesli@gmail.com*
Application date: 18.02.2026 Accepted: 15.06.2026

INTRODUCTION

The integration of artificial intelligence (AI) into clinical workflows is transforming the way healthcare information is generated, accessed, and disseminated. As AI continues to advance, its utility as a data resource and decision-support tool in the medical field is becoming increasingly prominent. Among the most well-known AI systems is Chat Generative Pre-trained Transformer (ChatGPT), a large language model (LLM) developed by OpenAI and released in November 2022. ChatGPT is capable of generating contextually relevant, human-like text, making it a popular tool among researchers, clinicians, and patients (1, 2).

In March 2023, OpenAI released GPT-4, which demonstrated significant improvements over GPT-3.5 in terms of response quality, coherence, and accuracy (3). These systems are trained on massive datasets derived from publicly available web sources, books, and articles, but not from specialized clinical databases (4, 5). As a result, although ChatGPT can deliver plausible-sounding medical information, concerns remain regarding its tendency to generate incorrect or misleading content, commonly referred to as “AI hallucination” (5).

In April 2025, OpenAI introduced GPT-4.1 mini, and in August 2025, GPT-5 was released as a newer generation model (6, 7). Newer-generation models are generally expected to improve response consistency and reduce factual errors; however, their performance may vary by clinical task and context (8, 9). More broadly, the convergence of AI and medicine has been proposed as an important direction for healthcare (10).

Breast cancer-related lymphedema (BCRL) is a common and often chronic complication following breast cancer surgery or radiation therapy, with incidence rates ranging from 14% to 54% depending on diagnostic criteria and follow-up period (11). BCRL can manifest as early as a few weeks post-treatment or even decades later, profoundly affecting patients' physical function, emotional well-being, and quality of life (12). Physical exercise and return-to-work experiences are also important considerations in cancer survivorship (13). The International Society of Lymphology identifies complex decongestive therapy (CDT) as the international gold standard for lymphedema management. CDT includes manual lymphatic drainage, compression bandaging, therapeutic exercises, and skin care (14).

In addition to clinical interventions, patient education plays a pivotal role in successful lymphedema management. Educational strategies improve patients’ awareness of risk factors, early signs and symptoms, treatment options, and self-management practice (14). However, numerous studies have found that a large proportion of patients do not receive adequate information about their condition or its management (15, 16). This gap highlights the need for scalable and accessible educational resources.

Given the increasing global demand for rehabilitation services and the shortage of trained specialists, AI-based tools such as ChatGPT offer a promising opportunity to supplement conventional patient education (17). Previous research has shown that ChatGPT can provide generally accurate and helpful responses to patient questions in various conditions such as laryngopharyngeal reflux and COVID-19-related anosmia (18,19).

To our knowledge, no prior study has systematically evaluated the performance of GPT-4.1 mini and GPT-5 in the context of postmastectomy lymphedema education. This study therefore aims to assess and compare the reliability and usefulness of responses generated by these two state-of-the-art models. The findings are expected to provide timely insights into the feasibility of integrating different LLMs into digital rehabilitation education.

MATERIALS and METHODS

This study was conducted in September 2025 in a tertiary care physical medicine and rehabilitation setting. The study did not involve patients, patient data, biological material, or clinical interventions. Physician ratings were obtained exclusively for the evaluation of AI-generated text. Therefore, ethics committee approval was not required in accordance with institutional policies for non-interventional research. Twenty patient-centered questions were developed to evaluate the performance of ChatGPT-based LLMs in postmastectomy lymphedema patient education. The questions were formulated following a systematic review of frequently asked patient questions and educational materials obtained

from reputable and widely used sources. These sources included the websites of recognized organizations such as the American Cancer Society, National Cancer Institute, BreastCancer.org, and the Lymphatic Education & Research Network. All sources were selected based on their clinical credibility, evidence-based content, and patient accessibility (20–23).

Key domains of postmastectomy lymphedema care were predefined by consensus among three specialists in physical medicine and rehabilitation with expertise in lymphedema management. These domains included etiology, symptoms, preventive strategies, treatment modalities (e.g., CDT, compression therapy, exercise), daily living adaptations, and commonly misunderstood concepts. At least one question per domain was formulated with a focus on patient-centered phrasing to simulate real-world information-seeking behavior.

To minimize potential bias, three independent physiatrists, blinded to model identity and not involved in question design, evaluated the ChatGPT responses independently. This ensured objectivity in the assessment process.

Questions were systematically crafted around key themes: causes, symptoms, treatments, prevention, physical activity, and negatively phrased control questions. Additional inquiries were included for skin care, weight management, bandage therapy, compression garments, and pneumatic compression therapy. Daily living aspects such as exercise, housework, and travel were also covered. Four negatively phrased questions were included to assess whether the models appropriately rejected incorrect assumptions. The questions directed to ChatGPT, along with the evaluations from reviewer 1 (R1), reviewer 2 (R2), and reviewer 3 (R3), are presented in Table-1. Inter-rater reliability among physician evaluators and ChatGPT outputs was assessed using the intraclass correlation coefficient (ICC).

All questions were submitted individually in English through the ChatGPT web interface (chat.openai.com). GPT-4.1 mini and GPT-5 were accessed via the ChatGPT web interface in September 2025, and identical prompts were submitted under the same conditions. The same set of 20 questions was submitted to both models under identical conditions, and the responses were recorded verbatim. Screenshots and transcripts were stored immediately after generation to avoid later modifications.

Responses were assessed using 7-point Likert-type scales for reliability and usefulness, adapted from Uz et al. (Table-2 and Table-3) (24). Reliability scoring focused on factual correctness and consistency with clinical guidelines, while usefulness scoring evaluated clarity, completeness, and patient applicability. Higher scores indicated greater reliability and usefulness. All screenshots of the responses were independently reviewed by three physiatrists with expertise in lymphedema, who were blinded to each other's ratings.

All statistical analyses were conducted using SPSS version 27 (IBM, New York, USA). Descriptive statistics were expressed as mean ± standard deviation. Comparisons between GPT-4.1 mini and GPT-5 were performed using the Wilcoxon signed-rank test due to the ordinal nature of the data and the paired study design. A p value of <0.05 was considered statistically significant.

Table-1 Reviewer Scores for Usefulness and Reliability of ChatGPT Responses to 20 Lymphedema-Related Questions

	GPT-4.1 mini Usefulness			GPT-5 Usefulness			GPT-4.1 mini Reliability			GPT-5 Reliability		
	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3
1. What causes lymphedema?	4	4	4	5	6	6	3	4	5	6	5	6
2. When does lymphedema most commonly develop?	3	5	5	6	6	5	2	4	6	6	6	5
3. What are the symptoms of lymphedema?	6	6	6	7	6	6	6	6	6	7	6	6
4. How to prevent lymphedema?	6	6	6	6	6	5	6	6	7	6	6	5
5. Is there a treatment for lymphedema?	5	5	6	6	5	6	5	6	6	6	6	5
6. Can lymphedema be cured completely?	2	5	6	5	6	6	3	6	6	5	6	6
7. How is lymphedema treated?	6	6	6	6	6	6	6	6	6	6	6	5
8. How to care for the skin of a lymphedema patient?	7	6	7	7	6	6	7	6	6	7	5	6
9. Is weight gain important for lymphedema?	5	6	6	7	6	6	5	5	5	7	6	6
10. How is bandage therapy used in the treatment of lymphedema?	5	6	6	6	6	6	5	6	5	6	6	6
11. How to use pressure gloves/garments to treat lymphedema?	7	6	6	7	6	6	7	6	6	7	6	6
12. Can pneumatic pumps be used in the treatment of lymphedema?	2	6	5	3	6	6	2	6	4	3	6	6
13. Can I do sports with lymphedema?	3	5	6	3	6	6	3	6	5	4	6	6
14. What exercises should I perform to treat my lymphedema?	6	5	6	6	6	7	6	6	6	6	6	6
15. If I stand without moving my arm in lymphedema, will the swelling go down?	6	6	6	6	6	6	6	6	6	6	6	6
16. Can I do housework if I have lymphedema?	6	5	5	5	6	6	6	6	5	5	6	5
17. Can I travel if I have lymphedema?	6	6	7	6	6	6	6	6	6	6	6	5
18. Can herbals be used safely in the treatment of lymphedema?	5	7	6	7	6	6	5	6	5	7	6	6
19. Can lymphedema be cured with surgical treatment completely?	5	5	6	6	6	4	5	5	6	6	6	4
20. Is there a medication specifically used to treat lymphedema?	3	7	4	6	6	6	3	6	4	6	6	6

Table-2 Reliability Score

1. Completely unsafe: None of the information provided can be verified from medical sources or contains inaccurate and incomplete information.
2. Very unsafe: Most of the information cannot be verified from medical sources or are partially correct but contains important incorrect or incomplete information.
3. Relatively reliable: The majority of the information provided are verified from medical scientific sources, but there are some important incorrect or incomplete information
4. Reliable: Most of the information provided are verified from medical scientific sources, but there are some minor inaccurate or incomplete information.
5. Relatively very reliable: Most of the information provided are verified from medical scientific sources, and there is very little incorrect or incomplete information.
6. Very reliable: Most of the information provided are verified from medical scientific sources, and there is almost no inaccurate or incomplete information.
7. Absolutely reliable: All of the information provided are verified from medical scientific sources, and there is no inaccurate or incomplete information or missing information.

Table-3 Usefulness Score

1. Not useful at all: Unintelligible language, contradictory information, and missing important information. Not useful for users.

2. Very little useful: Partly clear language is used. Some important information are missing or incorrect. For users, limited use is possible.
3. Relatively useful: Clear language is used. Most important information are mentioned, but some important information are incomplete or incorrect. Useful for users.
4. Partly useful: Clear language is used. Some important information are missing or incorrect, but most important information are addressed. Somewhat useful for users.
5. Moderately useful: Clear language is used and most important information are covered, but some important information are still incomplete or incorrect. Useful for users.
6. Very useful: Clear language is used. All important information are mentioned, but some unimportant information or details are also mentioned. Very useful for users.
7. Extremely useful: Clear language is used and all important information is mentioned. Extremely useful to patients, additional information and resources are also provided.

RESULTS

Inter-rater reliability analysis demonstrated moderate agreement among the three evaluators for both usefulness (ICC = 0.599, 95% CI 0.281–0.814) and reliability (ICC = 0.510, 95% CI 0.134–0.771) (Table-5).

Descriptive analysis showed that GPT-5 had higher mean scores compared to GPT-4.1 mini for both usefulness and reliability. The mean usefulness score was 5.87 ± 0.42 for GPT-5 and 5.43 ± 0.79 for GPT-4.1 mini. Similarly, the mean reliability score was 5.80 ± 0.38 for GPT-5 and 5.37 ± 0.81 for GPT-4.1 mini.

Comparison of the two models using the Wilcoxon signed-rank test revealed that GPT-5 had higher usefulness scores compared to GPT-4.1 mini at the overall score level (p = 0.005). Similarly, reliability scores were also significantly higher for GPT-5 (p = 0.021) (Table-4).

Qualitative examination of the responses identified several notable differences between the models. GPT-5 provided a clearer distinction between primary and secondary lymphedema and more explicitly referred to complex decongestive therapy as a key component of lymphedema management. In contrast, GPT-4.1 mini included statements regarding the possible use of diuretics, which may lead to misunderstanding in a patient education context. These observations were not subjected to statistical testing and should therefore be interpreted as descriptive findings.

The negatively phrased control questions were generally handled appropriately by both models. Responses emphasized that lymphedema cannot be completely cured, including through surgical treatment, that arm inactivity is not an effective strategy for reducing swelling, and that there is no medication specifically indicated for the treatment of lymphedema.

Table-4. Comparison of GPT-4.1 mini and GPT-5 in terms of usefulness and reliability using the Wilcoxon signed-rank test

Outcome	GPT-4.1 mini	GPT-5	p-value
Usefulness	5.43 ± 0.79	5.87 ± 0.42	**0.005**
Reliability	5.37 ± 0.81	5.80 ± 0.38	**0.021**

Values are presented as mean ± SD.

Table-5. Inter-rater reliability analysis for usefulness and reliability scores

Measure	Cronbach's Alpha	ICC (95% CI)
Usefulness	0.634	0.599 (0.281–0.814)
Reliability	0.548	0.510 (0.134–0.771)

ICC: Intraclass correlation coefficient; CI: Confidence interval.

DISCUSSION

This study contributes to the growing literature on AI-generated patient education content by evaluating responses related to postmastectomy lymphedema. While prior research has mainly addressed common conditions, our findings provide condition-specific insights in the field of rehabilitation.

Our findings provide preliminary evidence suggesting the potential of ChatGPT as a supplementary tool for educating BCRL. While LLMs such as ChatGPT have demonstrated impressive capabilities, their deployment in clinical settings remains limited by critical constraints. ChatGPT has achieved remarkable global reach, with over 200 million weekly active users (25). This widespread use highlights the need for healthcare professionals to understand both the capabilities and limitations of AI-based platforms. Although ChatGPT responses received relatively high usefulness and reliability scores in our study,

many scholars agree that it is not yet ready for unregulated use in clinical decision-making or unsupervised patient education (26).

Studies evaluating ChatGPT across medical domains suggest that newer versions provide modest improvements over earlier models. For example, GPT-4 has shown slightly improved accuracy in patient education tasks such as laryngopharyngeal reflux and interventional radiology (19, 32). Large language models have also demonstrated the capacity to encode substantial clinical knowledge (33). Our study extends this line of research by comparing GPT-4.1 mini and GPT-5 in BCRL education. Although GPT-5 demonstrated statistically higher scores in overall usefulness and reliability, these differences likely reflect incremental improvements in response quality rather than substantial differences in clinical applicability. From a practical perspective, both models achieved relatively high scores, suggesting their potential to support basic patient education tasks.

Several qualitative differences were also observed between the models. GPT-5 provided a clearer distinction between primary and secondary lymphedema and more explicitly referred to complex decongestive therapy as a key component of lymphedema management, consistent with current recommendations (14). In contrast, GPT-4.1 mini did not emphasize complex decongestive therapy in one treatment-related response and included statements regarding the possible use of diuretics in selected situations. As diuretics are not routinely recommended for lymphedema treatment, such statements may be open to misinterpretation in a patient education setting. These qualitative observations may help explain some of the differences observed in overall usefulness and reliability scores.

Our findings also suggest that models such as GPT-4.1 mini may already provide meaningful support for fundamental patient education tasks, raising the question of whether more advanced models are truly necessary in this specific context. Nevertheless, larger and more diverse studies are needed to determine whether advanced models such as GPT-5 offer additional benefits in patient education. To our knowledge, direct comparisons between recent ChatGPT versions in condition-specific patient education remain limited.

Although we did not directly compare ChatGPT with other information sources, widely used platforms for health information have shown variable quality and limitations in accuracy and evidence-based content (28, 29). These findings highlight the need for future studies comparing AI-generated responses with other commonly used information sources in terms of reliability, comprehensiveness, and safety in patient education.

However, AI-generated content is not without risks. Although it provides accessible and user-friendly responses, ChatGPT lacks the capacity for nuanced, personalized care required in complex conditions such as BCRL. Oversimplification or omission of individual factors may compromise patient safety (30). In addition, cultural sensitivity and patient-specific contexts are essential for effective health education, yet current AI systems do not adequately address these aspects (31).

In resource-limited settings, where healthcare professionals face high patient volumes and time constraints, AI tools such as ChatGPT may serve as valuable adjuncts for delivering basic educational information. By supporting patient engagement and self-management in BCRL, these tools may improve efficiency and help reduce healthcare burden. Global health organizations, including the World Health Organization (WHO), have also highlighted both the potential and risks of AI in such contexts (27).

Given the well-documented psychosocial burden of BCRL, AI-enabled tools such as GPT-4.1 mini or GPT-5 may support patient understanding and engagement. Prior studies have linked lymphedema-related distress to poorer psychosocial functioning and reduced quality of life (34–36), while digital education interventions have shown benefits in reducing anxiety and depression (37, 38). Although our study did not assess psychosocial outcomes, the improved clarity and accessibility of AI-generated responses may contribute to patient education when appropriately integrated into survivorship care. Similar studies evaluating AI-generated health information, such as Ayers et al., have also relied on a limited number of expert raters, reflecting the exploratory nature of this field (39).

This study has several limitations that should be considered. First, the study was limited to a predefined set of questions, which may restrict the

generalizability of the findings. In addition, the absence of patient-centered usability data limits the ability to evaluate trust, comprehension, and engagement with AI-generated responses. Furthermore, the lack of real-world patient interaction prevents assessment of how these responses are perceived and applied in actual clinical contexts. Finally, the small number of expert raters (n=3) limits the statistical power of the analysis and further restricts the generalizability of the results. The inter-rater agreement was moderate, which may be attributed to the subjective interpretation of response quality despite the use of standardized evaluation criteria.

Future studies should incorporate more representative patient perspectives and objective measures of comprehension and behavior change. In addition, improving transparency in AI training data, ensuring regular model updates, and applying domain-specific fine-tuning in rehabilitation medicine may enhance the reliability and safety of these tools. While ChatGPT shows promise, its use should remain under human supervision, particularly given the risk of "AI hallucinations," where models may generate confident but inaccurate responses (6).

CONCLUSION

In conclusion, GPT-4.1 mini and GPT-5 demonstrated promising potential as reliable and useful adjuncts for patient education in postmastectomy lymphedema. Although GPT-5 showed higher usefulness and reliability scores, these differences likely reflect incremental improvements in response quality rather than substantial differences in clinical applicability.

Our findings suggest that both models achieved relatively high usefulness and reliability scores, indicating their potential to support fundamental patient education tasks. These results highlight the potential role of AI-based tools in improving access to health information, particularly in settings where conventional patient education is limited.

However, these systems should not be used as standalone sources of medical advice. Their optimal use lies in supporting patient education under the supervision of healthcare professionals and in combination with conventional educational materials. Given the potential risks, including inaccurate or misleading responses, careful integration into clinical practice is essential.

Further studies incorporating real-world patient perspectives and objective outcome measures are needed to better define the role of advanced AI models in patient education.

We confirm that this manuscript has not been published or submitted elsewhere.

There is no conflict of interest among the authors.
No funding was received for this study.

Conflict of Interest: The authors declare that there is no real, potential, or perceived conflict of interest for this article

References

1. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol 2019;28:73-81.
2. Thirunavukarasu AJ, Ting DSJ, Elangovan K, et al. Large language models in medicine. Nat Med 2023;29:1930-40.
3. 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.
4. Xue VW, Lei P, Cho WC. The potential impact of ChatGPT in clinical and translational medicine. Clin Transl Med 2023;13:e1216.
5. van Dis EAM, Bollen J, Zuidema W, et al. ChatGPT: five priorities for research. Nature 2023;614:224-6.
6. OpenAI [homepage on the Internet]. Introducing GPT-4.1. San Francisco (CA): OpenAI; 2025 Apr [cited 2026 Jan 19]. Available from: https://openai.com/index/introducing-gpt-4-1
7. OpenAI [homepage on the Internet]. Introducing GPT-5. San Francisco (CA): OpenAI; 2025 Aug [cited 2026 Jan 19]. Available from: https://openai.com/index/introducing-gpt-5
8. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023;11:887.
9. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLoS Digit Health 2023;2:e0000198.
10. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44-56.

11. Donahue PMC, MacKenzie A, Filipovic A, et al. Advances in the prevention and treatment of breast cancer-related lymphedema. Breast Cancer Res Treat 2023;200:1-14.
12. Basha MA, Aboelnour NH, Alsharidah AS, et al. Effect of exercise mode on physical function and quality of life in breast cancer-related lymphedema: a randomized trial. Support Care Cancer 2022;30:2101-10.
13. Groeneveld IF, de Boer AG, Frings-Dresen MH. Physical exercise and return to work: cancer survivors' experiences. J Cancer Surviv 2013;7:237-46.
14. Executive Committee of the International Society of Lymphology. The diagnosis and treatment of peripheral lymphedema: 2020 consensus document of the International Society of Lymphology. Lymphology 2020;53:3-19.
15. Perdomo M, Davies C, Levenhagen K, et al. Patient education for breast cancer-related lymphedema: a systematic review. J Cancer Surviv 2023;17:384-98.
16. Borman P, Yaman A, Yasrebi S, et al. The importance of awareness and education in patients with breast cancer-related lymphedema. J Cancer Educ 2017;32:629-33.
17. Peng S, Wang D, Liang Y, Xiao W, Zhang Y, Liu L. AI-ChatGPT/GPT-4: An Booster for the Development of Physical Medicine and Rehabilitation in the New Era! Ann Biomed Eng 2024;52:462-6.
18. Sina EM, Campbell DJ, Duffy A, et al. Evaluating ChatGPT as a patient education tool for COVID-19-induced olfactory dysfunction. OTO Open 2024;8:e70011.
19. Lechien JR, Carroll TL, Huston MN, et al. ChatGPT-4 accuracy for patient education in laryngopharyngeal reflux. Eur Arch Otorhinolaryngol 2024;281:2547-52.
20. American Cancer Society [homepage on the Internet]. Lymphedema after breast cancer [cited 2026 Jan 19]. Available from: https://www.cancer.org/
21. National Cancer Institute [homepage on the Internet]. Lymphedema (PDQ®)–Patient Version [cited 2026 Jan 19]. Available from: https://www.cancer.gov/
22. BreastCancer.org [homepage on the Internet]. Lymphedema after breast cancer [cited 2026 Jan 19]. Available from: https://www.breastcancer.org/
23. Lymphatic Education & Research Network [homepage on the Internet]. [cited 2026 Jan 19]. Available from: https://lymphaticnetwork.org/
24. Uz C, Umay E. "Dr ChatGPT": is it a reliable and useful source for common rheumatic diseases? Int J Rheum Dis 2023;26:1343-9.
25. Reuters. OpenAI says ChatGPT's weekly users have grown to 200 million [serial on the Internet]. 2024 Aug 29 [cited 2026 Jan 19]. Available from: https://www.reuters.com/technology/artificial-intelligence/openai-says-chatgpts-weekly-users-have-grown-200-million-2024-08-29/
26. Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. Nat Med 2022;28:31-8.
27. World Health Organization. Ethics and governance of artificial intelligence for health. Geneva: World Health Organization; 2021.
28. Küçükakkaş O, İnce B. Can YouTube be used as an educational tool in lymphedema rehabilitation? Arch Physiother 2022;12:5.
29. Kandemir D, Temiz Z, Aydın Sayılan A. YouTube as a source of information on breast cancer-related lymphedema exercises: a content, quality, and reliability analysis. Health Care Women Int 2025;46:1206-20.
30. Millenson ML, Baldwin JL, Zipperer L, Singh H. Beyond Dr. Google: the evidence on consumer-facing digital tools for diagnosis. Diagnosis (Berl) 2018;5:95-105.
31. Temel MH, Erden Y, Bağcıer F. Information quality and readability: ChatGPT's responses to the most common questions about spinal cord injury. World Neurosurg 2024;181:e1138-44.
32. Scheschenja M, Viniol S, Bastian MB, Wessendorf J, König AM, Mahnken AH. Feasibility of GPT-3 and GPT-4 for in-depth patient education prior to interventional radiological procedures: a comparative analysis. Cardiovasc Intervent Radiol 2024;47:245-50.
33. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature 2023;620:172-80.
34. Eaton LH, Narkthong N, Hulett JM. Psychosocial issues associated with breast cancer-related lymphedema: a literature review. Curr Breast Cancer Rep 2020;12:216-24.
35. Jørgensen MG, Toyserkani NM, Hansen FG, Bygum A, Sørensen JA. The impact of lymphedema on health-related quality of life up to 10 years after breast cancer treatment. NPJ Breast Cancer 2021;7:70.
36. Dominick SA, Natarajan L, Pierce JP, Madanat H, Madlensky L. The psychosocial impact of lymphedema-related distress among breast cancer survivors in the WHEL Study. Psychooncology 2014;23:1049-56.
37. Ambrosi E, Mezzalira E, Canzan F, Leardini C, Vita G, Marini G, Longhini J. Effectiveness of digital health interventions for chronic conditions management in European primary care settings: systematic review and meta-analysis. Int J Med Inform 2025;196:105820.
38. Johnson E, Corrick S, Isley S, et al. Mind-body internet and mobile-based interventions for depression and anxiety in adults with chronic physical conditions: a systematic review of RCTs. PLoS Digit Health 2024;3:e0000435.
39. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023;183:589-96.