Original Investigation / Orijinal Araştırma

DOI: 10.4274/tod.galenos.2024.76743
Turk J Osteoporos 2025;31(1):12-8

Digital Guidance: Quality and Readability Analysis of Artificial Intelligence-Generated Spondyloarthropathy Texts

Dijital Rehberlik: Yapay Zeka ile Oluşturulan Spondiloartropati Metinlerinin Kalite ve Okunabilirlik Analizi

İlhan Celil Özbek[1], Volkan Hancı[2], Erkan Özduran[3]

[1]University of Health Sciences Türkiye, Kocaeli Derince Training and Research Hospital, Clinic of Physical Medicine and Rehabilitation, Kocaeli, Türkiye
[2]Dokuz Eylül University, Faculty of Medicine, Department of Anaesthesia and Reanimation, İzmir, Türkiye
[3]Sivas Numune Hospital, Clinic of Physical Medicine and Rehabilitation, Division of Pain Medicine, Sivas, Türkiye

Abstract

Objective: The aim of this study was to comprehensively evaluate the quality and readability of the content of artificial intelligence (AI)-generated texts about spondyloarthropathy (SpA).

Materials and Methods: The most frequently searched keywords related to the SpA-group were identified through Google Trends. The keywords were sequentially entered into AI chatbots (ChatGPT, Bard, Copilot). The Ensuring Quality Information for Patients (EQIP) tool was used to assess the clarity of information and quality of writing. Flesch-Kincaid readability tests (reading-ease and grade-level) and Gunning Fog index (GFI) were used to assess the readability of the texts.

Results: The mean EQIP score of the texts was 66.44. The mean Flesch-Kincaid reading ease score was 38.06. The mean score for Flesch-Kincaid grade level is 11.38. The mean GFI score is 13.91. Our study concludes that the AI chatbots' responses on SpA are generally of "good quality with minor problems". It was determined that the texts produced were complex enough to require approximately 11 years of training. When the quality and readability characteristics of the texts generated by the AI chatbots were compared, the EQIP scores of the texts generated by Copilot were higher than those generated by both ChatGPT and Bard ($p<0.001$, $p=0.004$, respectively). Furthermore, ChatGPT-generated texts were found to require a higher level of education than those generated by both Copilot and Bard ($p=0.002$, $p=0.004$, respectively).

Conclusion: This study reveals that AI chatbots' texts about SpA have certain shortcomings in terms of quality and readability. As a result, it emphasizes that online resources and AI tools play an important role in information delivery in the healthcare field, but quality and readability control should be ensured. This can facilitate patients' access to accurate, reliable, and comprehensible information.

Keywords: Spondyloarthropathy, artificial intelligence chatbot, ChatGPT, Bard, Copilot, quality assessment, readability

Öz

Amaç: Bu çalışmanın amacı, spondiloartropati (SpA) ile ilgili yapay zeka destekli oluşturulan metinlerin içeriğinin kalitesini ve okunabilirliğini kapsamlı bir şekilde değerlendirmektir.

Gereç ve Yöntem: Google Trends üzerinden SpA grubu ile ilgili en sık aranan anahtar kelimeler belirlendi. Belirlenen anahtar kelimeler sırayla yapay zeka sohbet robotlarına (ChatGPT, Bard, Copilot) girildi. Bilginin netliği ve yazım kalitesi açısından değerlendirmek için Hastalar için Kaliteli Bilgi Sağlama aracı (EQIP) kullanıldı. Metinlerin okunabilirliğini değerlendirmek için Flesch-Kincaid okunabilirlik testleri (okuma kolaylığı ve sınıf düzeyi) ve Gunning Fog indeksi (GFI) kullanıldı.

Bulgular: Metinlerin EQIP skoru ortalama değerleri 66,44'tür. Flesch-Kincaid okuma kolaylığı skoru ortalama değeri ise 38,06'dır. Flesch-Kincaid sınıf düzeyi için ortalama skor 11,38'dir. GFI skoru ortalaması ise 13,91'dir. Çalışmamız, yapay zeka sohbet robotlarının SpA konusundaki yanıtlarının genel olarak "küçük sorunlarla birlikte iyi kaliteli" olduğu sonucuna varmaktadır. Üretilen metinlerin yaklaşık 11 yıl eğitim gerektirecek karmaşıklıkta olduğu belirlendi. Yapay zeka sohbet robotlarına oluşturduğu metinler kalite ve okunabilirlik özellikleri karşılaştırıldığında, Copilot tarafından üretilen metinlerin EQIP skorları, hem ChatGPT hem de Gemini tarafından üretilenlere göre anlamlı derecede daha yüksekti (sırasıyla, $p<0,001$, $p=0,004$). Ayrıca, ChatGPT tarafından üretilen metinlerin, hem Copilot hem de Gemini tarafından üretilenlere göre daha yüksek bir eğitim seviyesi gerektirdiği belirlendi (sırasıyla, $p=0,002$, $p=0,004$).

Corresponding Author/Sorumlu Yazar: İlhan Celil Özbek, University of Health Sciences Türkiye, Kocaeli Derince Training Research Hospital, Clinic of Physical Medicine and Rehabilitation, Kocaeli, Türkiye
E-mail: ilhanozbek7@gmail.com **ORCID ID:** orcid.org/0000-0003-0508-8868
Received/Geliş Tarihi: 18.08.2024 **Accepted/Kabul Tarihi:** 30.09.2024 **Publication Date/Yayınlanma Tarihi:** 20.03.2025

Cite this article as/Atıf: Özbek İC, Hancı V, Özduran E. Digital guidance: quality and readability analysis of artificial intelligence-generated spondyloarthropathy texts. Turk J Osteoporos. 2025;31(1):12-8

Abstract

Sonuç: Bu çalışma, yapay zeka sohbet robotlarının SpA hakkındaki metinlerinin kalite ve okunabilirlik konusunda belirli eksikliklerin bulunduğunu ortaya koymaktadır. Sonuç olarak, çevrimiçi kaynakların ve yapay zeka araçlarının sağlık alanında bilgi sunumunda önemli bir rol oynadığını, ancak kalite ve okunabilirlik kontrolünün sağlanması gerektiğini vurgulamaktadır. Bu, hastaların doğru, güvenilir ve anlaşılır bilgilere erişimini kolaylaştırabilir.
Anahtar kelimeler: Spondiloartropati, yapay zeka sohbet robotu, ChatGPT, Bard, Copilot, kalite değerlendirmesi, okunabilirlik

Introduction

Spondyloarthropathy (SpA) is a term used to describe a group of diseases that share various both hereditary and clinical characteristics. Common characteristics of SpA include axial skeleton involvement, peripheral arthritis, enthesitis, dactylitis, acute anterior uveitis, psoriasis, or inflammatory bowel disease. This group of diseases is classified as axial or peripheral based on the predominant clinical feature. The axial form is characterized by involvement of the spine and/or sacroiliac joints and includes subtypes such as ankylosing spondylitis and non-radiographic axial spondyloarthritis, whereas the peripheral form is characterized by peripheral arthritis, enthesitis, and/or dactylitis (1-3).

SpA typically begins in the third decade of life and is a significant group of diseases that can cause chronic pain and disability (4). Prevalence studies usually do not include imaging and HLA-B27 testing, making it difficult to determine the exact prevalence of SpA. However, studies in North America estimate the prevalence of SpA to be between 0.4% and 1.3% (5). Another study found that the global prevalence of SpA varies between 0.21% and 1.61% in different geographical regions (6).

Artificial intelligence (AI) is the evolution of algorithms designed to perform tasks associated with intelligent behavior. These algorithms encompass many areas such as natural language understanding, image recognition, decision-making, problem-solving, and learning from experience (7). In the healthcare sector, AI is utilized in various areas such as medical imaging, diagnosis, drug development, patient monitoring, and robot-assisted surgery (8).

Recent studies show that the use of AI-powered chatbots is on the rise (9). These robots are designed to generate appropriate and consistent responses to user inputs, addressing patients' needs, resolving their questions, providing health information, and assisting with appointment scheduling (10,11). However, there are uncertainties and reliability issues when obtaining health-related information online. Additionally, individuals with limited understanding of medical terms may struggle to assess the reliability and validity of the information they acquire (12). Therefore, it is crucial for patients to access information that is accessible, comprehensible, and reliable. Well-structured and trustworthy information can help patients learn about their diseases, understand treatment options, and implement preventive measures (13,14).

There are numerous studies in the literature investigating the quality and readability of health information related to medical conditions. However, there is no study in the literature that evaluates the health information generated by AI chatbots for the SpA group. The aim of this study is to comprehensively evaluate the quality and readability of AI-generated texts related to SpA.

Materials and Methods

The study was conducted on May 10, 2024, at the Medical Faculty Hospital of our University. No human or animal participants were included in this study; Hence, ethical approval was not required. Similar studies in the literature have followed the same approach Since this study did not involve patient intervention, individual patient consent was not required (15).

The most frequently searched keywords related to SpA, ankylosing spondylitis, psoriatic arthritis, enteropathic arthritis, and reactive arthritis were identified using Google Trends. Before starting the searches, all browser data were completely cleared to ensure the results were not influenced. The search criteria were set to include data from 2004 to the present, covering the entire world and all categories. The most relevant keywords were selected from the related queries section of the results. The twenty-five most frequently used keywords were recorded for each search, except for enteropathic arthritis. Nine keywords were obtained for the enteropathic arthritis query. Exclusion criteria for the study included repetitive and irrelevant terms, which were removed from the analysis. In total, thirty keywords were identified (Table 1). The number of keywords to be evaluated was determined considering similar studies in the literature (12,15,16).

Three separate accounts were created for the AI chatbots Bard Version 2.0.0 (https://bard.google.com/), Copilot (https://copilot.microsoft.com/), and ChatGPT (https://chat.openai.com/) dedicated to this study. The selected thirty keywords were entered sequentially into the chat interfaces of the AI chatbots. Each keyword was processed to lead to a separate interaction on different chat pages to minimize the potential impact of previous queries and responses. The resulting responses were systematically documented for subsequent analysis, focusing particularly on quality, comprehensiveness, and readability. Texts were copied into Microsoft Office Word 2016 (Microsoft Corporation, Redmond, WA) and saved. Marks such as options and bullet points were removed during the evaluations. All answers were recorded on the internet. (Access adress: https://archive.org/details/19_20240703_202407/gemini/1/, https://archive.org/details/5_20240703_202407/chatgpt/1/ https://archive.org/details/6_20240703/copilot/1/)

Table 1. Most searched keywords related to spondyloarthropathy group		
Ankylosing spondylitis	Undifferentiated spondyloarthropathy	Reactive arthritis symptoms
Ankylosing spondylitis pain	Inflammatory spondyloarthropathy	Reactive arthritis treatment
Ankylosing spondylitis arthritis	Spondyloarthropathy symptoms	Reactive arthritis
Ankylosing spondylitis symptoms	Spondyloarthropathy treatment	Reactive arthritis causes
Ankylosing spondylitis treatment	Spondyloarthropathy	Septic arthritis
Ankylosing spondylitis test	Seronegative arthritis	Enteropathic arthritis
Ankylosing spondylitis disease	Sacroiliitis	Reactive arthritis diagnosis
Psoriatic spondyloarthropathy	Psoriatic arthritis pain	Psoriatic arthritis signs
Psoriasis arthritis	Psoriatic arthritis treatment	Methotrexate psoriatic arthritis
Psoriatic arthritis symptoms	Psoriasis and psoriatic arthritis	Psoriatic arthritis nails

Evaluation of the Texts

The obtained 90 texts were evaluated for clarity and writing quality using the Ensuring Quality Information for Patients (EQIP) tool. A form containing 20 EQIP items was used to evaluate the texts (17). Each item was assessed with responses of "yes", "partly", "no", or "not applicable" (N/A).

Since access permission was required for the health services contact number information and the responses were not produced in PDF format for the reader to take notes, these criteria were not evaluated (11). In addition, supporting the generated responses with visuals is another criterion that was not evaluated for Copilot and ChatGPT, which are text-based AI models.

The total score was calculated by assigning 1 point for "yes" responses, 0.5 points for "partly" responses, and 0 points for "no" responses. Items marked "not applicable" were excluded from the total number of items. The overall score was then divided by the number of valid items and expressed as a percentage. The EQIP score was categorized according to the score ranges recommended in the EQIP development publication: sources scoring between 76% and 100% were classified as "well-written and high-quality", those scoring between 51% and 75% as "good quality with minor issues", those scoring between 26% and 50% as having "serious quality issues", and those scoring between 0% and 25% as having "severe quality issues" (18).

Each text was independently evaluated by two physical medicine and rehabilitation specialists (İ.C.Ö and E.Ö.) in separate settings to minimize bias. In case of any discrepancies, the assessment was carried out again and a solution was found by consensus among the experts.

To assess the readability of the texts, the Flesch-Kincaid readability (FKRE) tests (readability ease and grade level) and the Gunning Fog index (GFI) were utilized. Texts were evaluated using a calculator (https://readabilityformulas.com/readability-scoring-system.php).

The FKRE ease score is calculated using the formula: 206.835-(1.015 x average sentence length)-(84.6 x average syllables per word). The higher the score on the test, the more readable the content is. A score below 30 indicates a reading level comparable to that of university graduates.

The Flesch-Kincaid grade level (FKGL) Score is calculated using the formula: 0.39 x (total words/Total sentences) + 11.8 x (total syllables/total words) - 15.59. The result indicates the educational level of the audience the text is aimed at. For example, a result of 10 and above suggests the text is aimed at a high school level audience (19).

The GFI is an assessment based on sentence length and the complexity of words. GFI is calculated using the formula: (number of words/number of sentences)+[(number of words with three or more syllables x 100)/(number of words)] x 0.4. According to the formula, shorter sentences indicate better readability. A score above 12 indicates a difficult text to read (19).

Readability scores were analysed and compared with the sixth grade readability level recommended by the American Medical Association and the National Institutes of Health. The accepted readability level for the FKRE formula was 80.0, whereas for the other 2 formulae it was 6 (20).

Statistical Analysis

Version 27.0 of the Statistical Package for the Social Sciences was used to analyze the study data. For normally distributed variables, descriptive statistics were shown as mean±standard deviation; For non-normally distributed variables, they were shown as median (minimum-maximum). Both visually (using probability plots and histograms) and analytically (using the Kolmogorov-Smirnov test) was the normality of the variable distribution evaluated.

The Kruskal-Wallis test was used to compare more than two groups when the data were non-normally distributed. The Mann-Whitney U test was used for pairwise comparisons, and the Bonferroni correction was used. Intraclass correlation coefficient (ICC) analysis was performed to determine the consistency in EQIP assessments. P-values of less than 0.05 were used to classify results as statistically significant.

Results

When examining the countries with the highest search frequencies related to SpA, the top three are New Zealand, Australia, and the United Kingdom (Figure 1). Similarly, for searches related to reactive arthritis and enteropathic arthritis,

the leading countries are the United Kingdom, New Zealand, and Australia. For ankylosing spondylitis, the top three countries are Australia, New Zealand, and Ireland. In searches for psoriatic arthritis, Germany, Austria, and Switzerland rank the highest.

Table 2 presents the mean, standard deviation, median, minimum, and maximum values of the EQIP, FKRE, FKGL, and GFI scores. The EQIP scores of the texts range from 54.14 to 78.12, with an average of 66.44. The FKRE scores range from 0 to 60.60, with an average score of 38.06. The FKGL scores range from 7.5 to 24.5, with an average score of 11.38. The GFI scores range from 8.61 to 26.38, with an average score of 13.91.

Table 3 contains the median, minimum, and maximum values of the EQIP, FKRE, FKGL, and GFI scores for the texts generated by the AI chatbots. Significant statistical differences were found in the EQIP, FKRE, FKGL, and GFI scores of the texts created by the AI chatbots ($p<0.001$, $p<0.001$, $p=0.001$, $p=0.003$, respectively) (Table 3).

According to the results of the pairwise group comparisons, after Bonferroni correction, the EQIP scores of the texts generated by the Copilot chatbot were found to be significantly higher than those generated by both the ChatGPT and Bard chatbots ($p<0.001$ and $p=0.004$, respectively).

In terms of FKRE scores, the texts produced by the ChatGPT chatbot were found to be significantly lower than those produced by both the Copilot and Bard chatbots ($p=0.005$ and $p<0.001$, respectively). Similarly, for FKGL scores, the texts generated by the ChatGPT chatbot were significantly higher than those produced by both the Copilot and Bard chatbots ($p=0.002$ and $p=0.004$, respectively).

Additionally, the GFI scores of the texts generated by the Copilot chatbot were found to be significantly higher than those generated by both the ChatGPT and Bard chatbots ($p=0.003$ and $p=0.007$, respectively) (Table 3).

When the median readability scores of all AI (ChatGPT, Copilot and Gemini) responses were compared with the sixth grade reading level, a statistically significant difference was observed in all scores compared to the sixth grade level ($p<0.001$). According to all scores, their answers had a readability above the sixth grade level (Table 4). The ICCs for EQIP were 0.904 for ChatGPT, 0.896 for Copilot, 0.873 for Gemini ($p<0.001$).

Figure 1. Interest in spondyloarthropathy-related searches across countries: 2004-2023 (based on Google Trends data)

Discussion

Our study concludes that the responses of AI chatbots regarding SpA are generally of "good quality with minor issues". It was determined that the average FKRE score was 38 and the texts produced were complex enough to require approximately 11 years of training. This is the first study to evaluate the quality and readability of responses generated by AI chatbots for the most frequently searched keywords related to the SpA group.

When examining the countries with the highest search frequencies related to SpA, the top three are New Zealand, Australia, and the United Kingdom. Similarly, for searches related to reactive arthritis and enteropathic arthritis, the

Tablo 2. Statistics of EQIP, FKRE, FKGL and GFI scores

	Minimum	Maximum	Median	Mean	Standard deviation
Ensuring Quality Information for Patients score	54.14	78.12	66.66	66.44	5.52
The Flesch-Kincaid reading ease score	0	60.60	41.35	38.06	12.06
The Flesch-Kincaid grade level score	7.5	24.5	10.75	11.38	2.66
Gunning Fog index score	8.61	26.38	13.25	13.91	3.2

EQIP: Ensuring Quality Information for Patients, FKRE: Flesch-Kincaid readability, FKGL: Flesch-Kincaid grade level, GFI: Gunning Fog index

Table 3. Comparison of EQIP, FKRE, FKGL and GFI Scores of texts generated by artificial intelligence chatbots

Median (min-max)	ChatGPT	Copilot	Bard	p-value
Ensuring Quality Information for Patients score	63.63 (54.54-71.87)[a]	72.72 (59.37-78.12)[b]	66.66 (54.54-75)[a]	<0.001
The Flesch-Kincaid reading ease score	31.65 (0-50.4)[a]	42.35 (21.1-59.1)[b]	43 (28.6-60.6)[b]	<0.001
The Flesch-Kincaid grade level score	12.35 (9.3-24.5)[a]	10.5 (8-12.3)[b]	10.45 (7.5-13.5)[b]	0.001
Gunning Fog index score	14.22 (10.9-26.38)[a]	12.33 (8.61-15.24)[b]	13.87 (9.16-19.94)[a]	0.003

[a,b]superscripts indicate the difference between groups. There is no difference in groups with a common letter
EQIP: Ensuring Quality Information for Patients, FKRE: Flesch-Kincaid readability, FKGL: Flesch-Kincaid grade level, GFI: Gunning Fog index, Min-max: Minimum-maximum

Table 4. Comparison of FKRE, FKGL and GFI Scores of texts generated by artificial intelligence chat robots according to the 6th grade reading level median

Median	ChatGPT	p-value	Copilot	p-value	Gemini	p-value
Ensuring Quality Information for Patients score	63.63	<0.001	72.72	<0.001	66.66	<0.001
The Flesch-Kincaid reading ease score	31.65	<0.001	42.35	<0.001	43.00	<0.001
The Flesch-Kincaid grade level score	12.35	<0.001	10.50	<0.001	10.45	<0.001
Gunning Fog index score	14.22	<0.001	12.33	<0.001	13.87	<0.001

FKRE: Flesch-Kincaid readability, FKGL: Flesch-Kincaid grade level, GFI: Gunning Fog index

leading countries are the United Kingdom, New Zealand, and Australia. For ankylosing spondylitis, the top three countries are Australia, New Zealand, and Ireland. In searches for psoriatic arthritis, Germany, Austria, and Switzerland rank the highest. These findings indicate how the tendency to access information on different types of SpA varies across countries. The research highlights the importance of geographical differences in awareness and access to information regarding these specific medical conditions. These data suggest that global health education and information efforts should focus more on specific regions.

Our study concludes that the responses of the three different AI chatbots are generally of "good quality with minor issues". The EQIP evaluations showed that all the texts reviewed followed a logical order, had a clear design, and addressed the reader respectfully and personally. However, some of the texts received zero points on certain evaluation criteria. We believe that even small improvements in these areas could elevate the texts from the "good quality" category to the "well-written and high-quality" category.

In intergroup comparisons, it was found that the EQIP scores of the texts generated by Copilot were significantly higher than those of the texts generated by ChatGPT and Bard. A determining factor for this difference could be that Copilot included references at the end of each text. It was observed that approximately half of the Bard texts included references, whereas ChatGPT did not include any references. Additionally, another factor contributing to the difference is that the majority of Bard's responses were supported by visuals. In a study evaluating different AI chatbots about erectile dysfunction, it was similarly observed that the EQIP scores of texts produced by Copilot were higher than those produced by ChatGPT and Bard (12).

Accessible, accurate, and easily understandable information is crucial in supporting individuals coping with SpA. High-quality and straightforward texts help patients understand the complexity of their condition, the available treatment options, and preventive measures. However, complex and difficult-to-understand online health information can lead to misunderstandings and even health risks (21).

In a study by Fahy et al. (22) evaluating ChatGPT responses related to anterior cruciate ligament injury, it was found that there were readability problems. Similarly, in a study examining responses related to spinal cord injury, it was observed that ChatGPT caused difficulties in terms of readability (16). Similar to our results, other studies in the literature also found that there were readability problems (15,23). In intergroup comparisons, the texts generated by ChatGPT required a higher educational level compared to those produced by Copilot and Bard. The results of a different study evaluating AI chatbots on erectile dysfunction were similar to our findings (12). To solve this problem, the importance of evaluating the quality of texts produced especially in the field of health with indices such as EQIP and readability indices such as FKRE, FKRL, GFI should be emphasized by teaching AI. In order to make the necessary arrangements, improvements should be made and audited in the database. These improvements will be a step towards ensuring patient safety while increasing health literacy. When these conditions are met, it can make patients more aware of the acceptance of the disease, the importance of treatment and the control of the process.

We did not find a study evaluating the responses of AI chatbots for the SpA group in the literature; However, other research in this area has provided us with several important findings. For example, a study analyzing YouTube videos related to SpA in terms of quality and reliability found that there are useful videos as well as misleading videos, and that these videos often contain inaccurate clinical features and unproven alternative treatments (24). Another study on the quality and readability of online information about ankylosing spondylitis found that less than half of the websites had high-quality content and that the average readability levels of the websites were lower than recommended (25). These findings underscore the need for SpA patients and healthcare professionals to be cautious when accessing online information.

In today's world, there is an increasing tendency for patients to seek information about health issues through online resources and AI-based chat tools (26). However, research indicates that these online resources are inadequate in terms of quality and readability (27-31). According to the results of our study, it is necessary to improve the quality and readability of AI chatbots as well. Consequently, patients and their families may suffer due to access to incorrect information (32). Therefore, ensuring the accuracy, quality, and readability of health information is of great importance. Compliance with quality and readability standards facilitates patients' access to reliable information and enhances health literacy (33). However, each patient is unique, and the treatment process requires a personalized approach. Therefore,

online resources and AI tools cannot replace healthcare professionals (34,35). The importance of the physician-patient relationship should always be emphasized.

Although the number of keywords evaluated in our study is approximately the same level as similar studies, there may be limitations in making generalizations.

Study Limitations

This limitation can be considered as a constraint of our study. Additionally, only English keywords were evaluated in the study. Evaluating keywords in different languages can broaden the scope of the results. Another limitation of our study is the use of a single calculator to evaluate the readability of websites. In the study conducted by Gül et al. (20), the correlation between different calculators was assessed, and medium strong correlation correlation results were obtained. Therefore, we also chose to use a single calculator.

Conclusion

This study reveals that AI chatbots' texts about SpA have certain shortcomings in terms of quality and readability. In conclusion, it emphasizes that online resources and AI tools play an important role in information delivery in the healthcare field, but quality and readability control should be ensured. This can facilitate patients' access to accurate, reliable and comprehensible information.

Ethics

Ethics Committee Approval: Since this study was not conducted on humans, it did not require an ethics committee certificate.

Informed Consent: Since this study was not conducted on humans, patient consent was not required.

Footnotes

Authorship Contributions

Concept: İ.C.Ö., V.H., E.Ö., Design: İ.C.Ö., V.H., E.Ö., Data Collection or Processing: İ.C.Ö., V.H., E.Ö., Analysis or Interpretation: İ.C.Ö., Literature Search: İ.C.Ö., Writing: İ.C.Ö.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study has received no financial support

References

1. Ivanova M, Zimba O, Dimitrov I, Angelov AK, Georgiev T. Axial spondyloarthritis: an overview of the disease. Rheumatol Int. 2024;44:1607-19.
2. Hay CA, Packham J, Prior JA, Mallen CD, Ryan S. Barriers and facilitators in diagnosing axial spondyloarthritis: a qualitative study. Rheumatol Int. 2024;44:863-84.
3. Bagcier F, Yurdakul OV, Ozduran E. Top 100 cited articles on ankylosing spondylitis. Reumatismo. 2021;72:218-27.
4. Sieper J, Poddubnyy D. Axial spondyloarthritis. Lancet. 2017;390:73-84.
6. Stolwijk C, van Onna M, Boonen A, van Tubergen A. Global prevalence of spondyloarthritis: a systematic review and meta-regression analysis. Arthritis Care Res (Hoboken). 2016;68:1320-31.
7. van Hartskamp M, Consoli S, Verhaegh W, Petkovic M, van de Stolpe A. Artificial intelligence in clinical health care applications: viewpoint. Interact J Med Res. 2019;8:e12100.
8. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017;69S:S36-40.
9. Skjuve M, Brandzaeg PB (2019). Measuring user experience in chatbots: An approach to interpersonal communication competence. Internet Science: INSCI 2018 International Workshops, St. Petersburg, Russia, October 24–26, 2018, Revised Selected Papers 5: Springer;113–120.
10. Chow JCL, Sanders L, Li K. Impact of ChatGPT on medical chatbots as a disruptive technology. Front Artif Intell. 2023;6:1166014.
11. Li J, Dada A, Puladi B, Kleesiek J, Egger J. ChatGPT in healthcare: a taxonomy and systematic review. Comput Methods Programs Biomed. 2024;245:108013.
12. Şahin MF, Ateş H, Keleş A, Özcan R, Doğan Ç, Akgül M, et al. Responses of five different artificial intelligence chatbots to the top searched queries about erectile dysfunction: a comparative analysis. J Med Syst. 2024;48:38.
13. Chapman L, Brooks C, Lawson J, Russell C, Adams J. Accessibility of online self-management support websites for people with osteoarthritis: a text content analysis. Chronic Illn. 2019;15:27-40.
14. Varady NH, Dee EC, Katz JN. International assessment on quality and content of internet information on osteoarthritis. Osteoarthritis Cartilage. 2018;26:1017-26.
15. Erden Y, Temel MH, Bağcıer F. Artificial intelligence insights into osteoporosis: assessing ChatGPT's information quality and readability. Arch Osteoporos. 2024;19:17.
16. Temel MH, Erden Y, Bağcıer F. Information uality and readability: ChatGPT's responses to the most common questions about spinal cord injury. World Neurosurg. 2024;181:e1138-44.
17. Moult B, Franck LS, Brady H. Ensuring quality information for patients: development and preliminary validation of a new instrument to improve the quality of written health care information. Health Expect. 2004;7:165-75.
18. Ladhar S, Koshman SL, Yang F, Turgeon R. Evaluation of online written medication educational resources for people living with heart failure. CJC Open. 2022;4:858-65.
19. Benzer A. A step towards a Turkish artificial intelligence based readability formula. Journal of Research and Experience. 2020;5:47-82.
20. Gül Ş, Erdemir İ, Hanci V, Aydoğmuş E, Erkoç YS. How artificial intelligence can provide information about subdural hematoma: assessment of readability, reliability, and quality of ChatGPT, BARD, and perplexity responses. Medicine (Baltimore). 2024;103:e38009.
21. Daraz L, Morrow AS, Ponce OJ, Farah W, Katabi A, Majzoub A, et al. Readability of online health information: a meta-narrative systematic review. Am J Med Qual. 2018;33:487-92.
22. Fahy S, Oehme S, Milinkovic D, Jung T, Bartek B. Assessment of quality and readability of information provided by ChatGPT in relation to anterior cruciate ligament injury. J Pers Med. 2024;14:104.
23. Abou-Abdallah M, Dar T, Mahmudzade Y, Michaels J, Talwar R, Tornari C. The quality and readability of patient information provided by ChatGPT: can AI reliably explain common ENT operations? Eur Arch Otorhinolaryngol. 2024;281:6147-53.
24. Kocyigit BF, Koca TT, Akaltun MS. Quality and readability of online information on ankylosing spondylitis. Clin Rheumatol. 2019;38:3269-74.
25. Elangovan S, Kwan YH, Fong W. The usefulness and validity of English-language videos on YouTube as an educational resource for spondyloarthritis. Clin Rheumatol. 2021;40:1567-73.

26. Gravina AG, Pellegrino R, Cipullo M, Palladino G, Imperio G, Ventura A, et al. May ChatGPT be a tool producing medical information for common inflammatory bowel disease patients' questions? An evidence-controlled analysis. World J Gastroenterol. 2024;30:17-33.
27. Kocyigit BF, Akaltun MS. Does YouTube provide high quality information? Assessment of secukinumab videos. Rheumatol Int. 2019;39:1263-8.
28. Sasse M, Ohrndorf S, Palmowski A, Wagner AD, Burmester GR, Pankow A, et al. Digital health information on autoinflammatory diseases: a YouTube quality analysis. Rheumatol Int. 2023;43:163-71.
29. Tolu S, Yurdakul OV, Basaran B, Rezvani A. English-language videos on YouTube as a source of information on self-administer subcutaneous anti-tumour necrosis factor agent injections. Rheumatol Int. 2018;38:1285-92.
30. Pamukcu M, Izci Duran T. Are YouTube videos enough to learn anakinra self-injection? Rheumatol Int. 2021;41:2125-31.
31. Erkin Y, Hanci V, Ozduran E. Evaluating the readability, quality and reliability of online patient education materials on transcutaneuous electrical nerve stimulation (TENS). Medicine (Baltimore). 2023;102:e33529.
32. Younis HA, Eisa TAE, Nasser M, Sahib TM, Noor AA, Alyasiri OM, et al. A systematic review and meta-analysis of artificial ıntelligence tools in medicine and healthcare: applications, considerations, limitations, motivation and challenges. Diagnostics (Basel). 2024;14:109.
33. Coşkun AB, Elmaoğlu E, Buran C, Alsaç SY. Integration of Chatgpt and E-Health Literacy: opportunities, challenges, and a look towards the future. Journal of Health Reports and Technology. 2024;10.
34. 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.
35. Javaid M, Haleem A, Singh RP. ChatGPT for healthcare services: an emerging stage for an innovative perspective. BenchCouncil Transactions on Benchmarks, Standards and Evaluations. 2023;3:100105.