Original Investigation / Orijinal Araştırma

DOI: 10.4274/tod.galenos.2025.66664
Turk J Osteoporos

Digital Rehabilitation in Parkinson's Disease: The Role of Artificial Intelligence-Assisted Exercise Training

Parkinson Hastalığında Dijital Rehabilitasyon: Yapay Zeka Destekli Egzersiz Eğitiminin Rolü

İlhan Celil Özbek[1], Erkan Özduran[2]

[1]University of Health Sciences Türkiye, Derince Training and Research Hospital, Department of Physical Medicine and Rehabilitation, Kocaeli, Türkiye
[2]Sivas Numune Hospital, Clinic of Physical Medicine and Rehabilitation, Division of Pain Medicine, Sivas, Türkiye

Abstract

Objective: The aim of this study is to evaluate the content quality, readability, reliability, understandability and applicability of ChatGPT-4o exercise recommendations for Parkinson's patients.

Materials and Methods: Questions about balance and coordination exercises for Parkinson's patients were directed to ChatGPT-4o with the literature and the experiences of physical medicine specialists. Readability was examined with simple measure of gobbledygook and the other most popular readability formula tests, understandability and applicability were examined with 5-point Likert. Content quality was evaluated with ensuring quality information for patients (EQIP), and global quality score (GQS), and reliability was evaluated with Modified DISCERN (mDISCERN) and JAMA score.

Results: It was determined that all readability scores were above the 6^{th} grade reading level (p<0.001). The understandability score was 3.80±1.01 and the applicability score was 3.73±0.96. In reliability and quality assessment, EQIP mean was determined as 50.86±3.83, GQS as 2.66±1.17, mDISCERN as 1.86±0.51. JAMA scores were 1.

Conclusion: ChatGPT-4o exercise information for Parkinson's disease patients was found to be understandable and applicable. However, the high readability level (average Flesch reading ease score of 38, equivalent to 11 years of education) limits accessibility for individuals with low health literacy, and the lack of reliability restricts safe clinical use. These limitations should be carefully considered before clinical implementation, and future improvements are required to enhance the accuracy and accessibility of AI-based health guidance.

Keywords: Artificial intelligence, ChatGPT, exercise, Parkinson's disease, quality assessment, readability

Öz

Amaç: Bu çalışmanın amacı, ChatGPT-4o tarafından Parkinson hastaları için önerilen egzersizlerin içerik kalitesini, okunabilirliğini, güvenilirliğini, anlaşılabilirliğini ve uygulanabilirliğini değerlendirmektir.

Gereç ve Yöntem: Parkinson hastalarına yönelik denge ve koordinasyon egzersizleri hakkında, literatür ve fiziksel tıp uzmanlarının klinik deneyimleri doğrultusunda oluşturulan sorular ChatGPT-4o'ya yöneltilmiştir. Okunabilirlik; anlaşılmazlığın basit ölçüsü ve diğer yaygın okunabilirlik formülleriyle değerlendirilmiştir. Anlaşılabilirlik ve uygulanabilirlik 5'li Likert ölçeğiyle değerlendirilmiştir. İçerik kalitesi; hastalar için bilgi kalitesinin sağlanması (EQIP) ve global kalite skoru (GQS) ile, güvenilirlik ise Modifiye DISCERN (mDISCERN) ve JAMA skoru ile değerlendirilmiştir.

Bulgular: Tüm okunabilirlik skorlarının 6. sınıf okuma seviyesinin üzerinde olduğu belirlenmiştir (p<0,001). Anlaşılabilirlik skoru 3,80±1,01, uygulanabilirlik skoru 3,73±0,96 olarak saptanmıştır. Güvenilirlik ve kalite değerlendirmesinde EQIP ortalaması 50,86±3,83, GQS 2,66±1,17, mDISCERN 1,86±0,51 olarak bulunmuştur. JAMA skoru ise 1'dir.

Sonuç: ChatGPT-4o tarafından Parkinson hastaları için sağlanan egzersiz bilgileri anlaşılır ve uygulanabilir bulunmuştur. Ancak, yüksek okuma düzeyi (ortalama Flesch okuma kolaylığı skoru 38, yaklaşık 11 yıllık eğitim seviyesine eşdeğer) düşük sağlık okuryazarlığına sahip bireyler için erişilebilirliği sınırlamakta ve güvenilirliğin olmaması güvenli klinik kullanımını kısıtlamaktadır. Bu sınırlamalar, klinik uygulamaya konmadan önce dikkatle değerlendirilmelidir ve yapay zeka tabanlı sağlık rehberliğinin doğruluğunu ve erişilebilirliğini artırmak için gelecekte iyileştirmeler yapılması gerekmektedir.

Anahtar kelimeler: Yapay zeka, ChatGPT, egzersiz, Parkinson hastalığı, kalite değerlendirmesi, okunabilirlik

Corresponding Author/Sorumlu Yazar: İlhan Celil Özbek MD, University of Health Sciences Türkiye, Derince Training and Research Hospital, Department of Physical Medicine and Rehabilitation, Kocaeli, Türkiye
E-mail: ilhanozbek7@gmail.com **ORCID ID:** orcid.org/0000-0003-0508-8868

Received/Geliş Tarihi: 31.07.2025 **Accepted/Kabul Tarihi:** 08.09.2025 **Epub:** 12.09.2025

Cite this article as/Atıf: Özbek İC, Özduran E. Digital rehabilitation in Parkinson's disease: the role of artificial ıntelligence-assisted exercise training. Turk J Osteoporos. [Epub Ahead of Print]

Introduction

Parkinson's disease (PD) is a chronic, progressive neurodegenerative disorder characterized by motor symptoms such as tremor, rigidity, bradykinesia, and postural instability, which significantly impair patients' functional independence and quality of life (1,2). Although pharmacological treatments are the cornerstone of symptom management, they are often insufficient in addressing motor impairments. For this reason, rehabilitation programs—including balance and coordination exercises—have emerged as essential components of PD management, aiming to maintain mobility, reduce fall risk, and improve daily living activities (3,4).

Despite the proven benefits of exercise, many PD patients encounter substantial barriers to implementing rehabilitation protocols. These include motor and cognitive challenges, limited access to specialized care, and insufficient guidance on proper exercise execution (5). Traditional rehabilitation services are often limited by workforce capacity, geographic disparities, and scheduling difficulties. Consequently, digital health technologies have gained increasing attention as tools to improve accessibility and adherence to exercise programs.

Artificial intelligence (AI), particularly large language models such as ChatGPT, has recently emerged as a novel approach to address such gaps in health education. These tools have the potential to deliver personalized, on-demand exercise recommendations, bridging the divide between clinical expertise and patient access. Recent studies have demonstrated that AI-based chatbots can assist patients with various chronic diseases—including fibromyalgia, low back pain, and spinal cord injury—by providing general information and exercise guidance with varying degrees of quality and reliability (6-12). However, limited research has focused specifically on their utility in Parkinson's rehabilitation.

Furthermore, while AI systems such as ChatGPT-4o have shown promise in improving patient engagement and understanding, concerns persist regarding the readability, reliability, and quality of the information they provide. For individuals with low health literacy, the complexity of AI-generated responses can hinder effective use. Additionally, the absence of references, author attribution, and evidence-based justification in chatbot outputs raises questions about their role as reliable health communication tools (13-15). Therefore, assessing these tools using validated readability, quality, and reliability metrics is essential before integrating them into clinical care.

This study aims to evaluate the content quality, readability, reliability, understandability, and applicability of ChatGPT-4o's exercise recommendations for patients with PD. By benchmarking the outputs against established standards and comparing them to the needs and limitations of PD patients, this study seeks to determine whether AI-assisted exercise guidance can serve as a viable complement to traditional rehabilitation methods.

Materials and Methods

Ethics Comittee Permission

This study was conducted at the University of Health Sciences Türkiye, Derince Training and Research Hospital, on April 24, 2025. The planning, execution and data collection processes of this cross-sectional study were carried out in accordance with the approval of the relevant ethics committee (Sivas Cumhuriyet University's Ethics Committee, ethics committee no: 2025-04/67, date: 24.04.2025).

Data Collection

This study evaluated the potential of ChatGPT-4o, an AI-supported language model, in delivering exercise training recommendations for patients with PD. To generate relevant queries, a pool of questions was created based on established clinical guidelines and recent meta-analyses focusing on balance and coordination exercises commonly prescribed in PD rehabilitation. The inclusion criteria prioritized clinical relevance, linguistic clarity, and alignment with functional goals such as fall prevention, mobility improvement, and postural control. Ambiguous, overly technical, or redundant questions were excluded. Two physical medicine and rehabilitation specialists (ICO and EO) independently reviewed the question set to ensure face and content validity (Table 1) (16-18).

Table 1. Questions asked to ChatGPT-4o for exercise training for Parkinson's disease patients
What is balance and which systems in the body provide balance?
What is the definition of coordination and which muscle groups is it related to?
Why does balance disorder occur in elderly people and how can it be prevented?
What are beginner level balance exercises?
How should balance exercises be performed with eyes closed?
What is the recommended duration for daily balance and coordination exercises?
How to do one-leg stand exercise and what are the benefits?
What are the appropriate coordination exercises for individuals with balance disorders?
What is the role of the vestibular system on balance?
What is proprioception and how can it be improved with balance exercises?
What are the common mistakes in balance exercises?
How can we improve the activities of daily living of patients with balance disorders?
What are the simple balance exercises we can apply at home?
How does artificial intelligence contribute to learning balance and coordination exercises?
How can health professionals use ChatGPT in balance exercises training?

To standardize the data collection process and reduce variability, all questions were entered using consistent phrasing and submitted in separate chat sessions to minimize contextual bias between responses. A dedicated ChatGPT-4o account was created for the study, and all interactions were conducted using the April 2025 version. The AI-generated responses were systematically documented and later evaluated in terms of quality, comprehensiveness, readability, and scientific accuracy (https://archive.org/details/chatgpt-parkinson-answers).

Readability Assessment

Readability assessment of the response texts was performed using two different web-based readability calculators: http://readabilityformulas.com/, https://www.online-utility.org/.

During the readability evaluation of the AI-generated texts, each calculator was evaluated separately, and their arithmetic averages were taken to record the final readability median (minimum-maximum) values.

Commonly used formulas were used to measure the readability of texts:

- Simple measure of gobbledygook (SMOG)
- Automated readability index (ARI)
- Gunning fog readability (GFOG)
- Flesch-Kincaid grade level (FKGL)
- Coleman-Liau readability index (CLI)
- The Flesch reading ease score (FRES).

Formulas and data for calculating the readability score are given in Table 2.

According to the standards set by the National Institutes of Health (NIH) and the American Medical Association (AMA), patient education materials must have a readability grade of six or below for the average individual to read. Therefore, the final readability scores we obtained were analyzed based on the sixth-grade readability level recommended by the institutions mentioned above. Accordingly, while the accepted average readability level is 6 for the other 6 formulas, 80.0 points are accepted for FRES (19,20).

Reliability Assessment

The reliability of the response texts was examined using two different scales:

1. JAMA Benchmark: Four basic criteria (transparency, authorship, timeliness and reference) were taken into account and the presence of each criterion was evaluated with 1 point and its absence with 0 points (21).

2. Modified DISCERN: Based on five basic criteria (whether the content included discussions, clarity, up-to-dateness of sources, impartiality and listing of additional sources), each criterion was evaluated out of 1 point (22).

Table 2. Readability formulas

Readability index	Description	Formula
Flesch reading ease score	It was developed to evaluate the readability of newspapers. It is best suited for evaluating school textbooks and technical manuals. The standardized test used by many US government agencies. Scores range from 0 to 100, with higher scores indicating easier readability.	I = (206.835 - (84.6 X (B/W)) - (1.015 X (W/S)))
Flesch-Kincaid grade level	Part of the Kincaid navy personnel test collection. Designed for technical documentation and suitable for a wide range of disciplines.	G = (11.8 X (B/W)) + (0.39 X (W/S)) −15.59
Simple measure of gobbledygook	It is generally suitable for middle-aged (4^{th} grade to college level) readers. While testing 100% comprehension, most formulas test about 50-75% comprehension. Most accurate when applied to documents ≥30 sentences long.	G = 1.0430 X √C + 3.1291
Gunning FOG	It was developed to help American businesses improve the readability of their writing. Applicable to many disciplines.	G = 0.4 X (W/S+((C∗/W) X 100))
Coleman-Liau index	It is designed for middle-aged (4^{th} grade to college level) readers. The formula is based on text in the grade level range of 0.4 to 16.3. It applies to many industries.	G = (−27.4004 X (E/100)) + 23.06395
Automated readability index (ARI)	ARI has been used by the military in writing technical manuals, and its calculation returns a grade level necessary for understanding.	ARI = 4.71 X l + 0.5*ASL - 21.43

W: Number of words, S: Number of sentences, B: Number of syllables, C: Number of polysyllabic words (≥3 syllables), l: Average number of letters per word, ASL: Average sentence length (number of words), E: Average number of letters per 100 words

Quality Assessment

The quality assessment of the response texts was carried out using two different methods:

1. Global quality score (GQS): This scale is scored from 1 to 5, with 1 point being considered "low quality" and 5 points being considered "very high quality" (23).

2. Ensuring quality information for patients (EQIP): In this 20-question assessment tool, "yes" was calculated as 1 point, "partially" as 0.5 points and "no" as 0 points. The results were interpreted on a scale of 0-100 (24).

Each text was independently assessed by two physical medicine and rehabilitation experts (ICO and EO) in different settings to reduce bias. In case of any differences, the assessment was re-performed and solutions were found by consensus among the experts.

In addition, the scientific accuracy of the responses was examined by two expert physicians (ICO and EO) in terms of compliance with the literature, and the criteria for understandability and applicability in terms of patient education were evaluated using a 5-point Likert scale (1: Very incomprehensible, 5: Very understandable) (25).

Statistical Analysis

Statistical analyses were conducted using SPSS software, version 24.0 for Windows (SPSS Inc., USA). Categorical data were expressed as frequencies and percentages, whereas continuous data were reported as both means with standard deviations and medians with their corresponding minimum and maximum values. Comparisons between categorical variables were carried out using Fisher's exact test and the chi-square test. For continuous variables, the Mann-Whitney U test and Wilcoxon signed-rank test were employed. To assess the level of agreement between the two readability calculators used in the study, intraclass correlation coefficients (ICC) were calculated. A two-way mixed-effects model with absolute agreement and average measures was applied, as the same two fixed tools were used to rate all items. The ICC values were computed for each readability formula (SMOG, ARI, GFOG, FKGL, CLI, FRES). A p-value of less than 0.05 was considered indicative of statistical significance.

Results

Table 3 presents the mean, standard deviation, minimum and maximum values of SMOG, ARI, GFOG, FKGL, CLI and FRES scores. SMOG score ranged from 8.58 to 15.18, with a mean of 12.35±2.43. ARI score ranged from 4.74 to 14.48, with a mean value of 10.17±3.57. GFOG scores ranged from 7.91 to 16.12, with a mean of 12.66±3.33. The scores obtained for FKGL ranged from 5.70 to 15.00, with a mean value of 11.12±3.40. CLI score ranged from 6.06 to 17.32, with a mean of 13.11±3.81. The FRES score was observed between 13.05 and 76.09 and the mean was determined as 38.81±21.84.

Table 3 includes the comparison of SMOG, ARI, GFOG, FKGL, CLI and FRES scores according to the 6th grade reading level median. The p values obtained for all scores were <0.001, indicating a statistically significant difference.

Table 4 shows the distribution of understandability and applicability scores. The understandability score ranged from 2 to 5, with an average of 3.80±1.01. Similarly, the applicability score ranged from 2 to 5, with an average of 3.73±0.96.

Table 5 provides a summary of reliability and quality scores. The EQIP score ranged from 42.30 to 66.60, with a mean of 50.86±3.83. The GQS score ranged from 1 to 4, with a mean of 2.66±1.17. The JAMA score was set to 1 in all data. The mDISCERN score ranged from 1 to 3, with a mean of 1.86±0.51. The ICC calculated in the study were as follows: 0.915 for JAMA Benchmark, 0.892 for Modified DISCERN, 0.938 for GQS, 0.927 for EQIP, 0.879 for SMOG, 0.862 for ARI, 0.895 for GFOG, 0.912 for FKGL, 0.921 for CLI and 0.934 for FRES. The ICC value for understandability score was calculated as 0.902, and the ICC value for applicability score was calculated as 0.918.

Table 3. Readability scores and comparison with 6th grade reading level median

	Minimum	Maximum	Median	Mean	SD	p
SMOG	8.58	15.18	13.41	12.35	2.43	<0.001
ARI	4.74	14.48	11.94	10.17	3.57	<0.001
GFOG	7.91	16.12	14.08	12.66	3.33	<0.001
FKGL	5.70	15.00	12.72	11.12	3.40	<0.001
CLI	6.06	17.32	14.99	13.11	3.81	<0.001
FRES	13.05	76.09	30.58	38.81	21.84	<0.001

SMOG: Simple measure of gobbledygook, ARI: Automated readability index, GFOG: Gunning fog readability, FKGL: Flesch-Kincaid grade level, CLI: Coleman-Liau index, FRES: The Flesch reading ease score, SD: Standard deviation

Table 4. Statistics of understandability and applicability scores

	Minimum	Maximum	Median	Mean	SD
Understandability (Likert)	2	5	4	3.80	1.01
Applicability (Likert)	2	5	4	3.73	0.96

SD: Standard deviation

Table 5. Statistics of reliability and quality scoress					
	Minimum	**Maximum**	**Median**	**Mean**	**SD**
EQIP	42.30	66.60	50	50.86	3.83
GQS	1	4	3	2.66	1.17
JAMA	1	1	1	1	0
mDISCERN	1	3	2	1.86	0.51

JAMA: The Journal of the American Medical Association, GQS: Global quality score, EQIP: The Ensuring quality information for patients, SD: Standard deviation, mDISCERN: Modified DISCERN

Discussion

In this study, the information provided by ChatGPT-4o on exercise training for PD patients was evaluated in terms of readability, quality, and reliability. Our analysis showed that the AI-generated texts exceeded the 6th-grade readability level recommended by the NIH and AMA, with a mean FRES score of 38, suggesting a complexity equivalent to approximately 11 years of education.

While the responses were rated as generally understandable and applicable, they displayed limitations in reliability and technical depth. The EQIP and GQS scores reflected moderate content quality, whereas the JAMA and mDISCERN scores highlighted a lack of transparency and evidence-based references. These findings align with previous studies examining AI-based health information tools for other conditions, such as fibromyalgia, low back pain, and spinal cord injury, which similarly noted issues with readability and content reliability (9-12).

The need for accessible and reliable health information for PD patients is well-documented. A recent study evaluating 60 PD websites found that only a small proportion provided clear, comprehensive, and useful information, indicating that the availability of adequate online resources remains limited (26). Similarly, studies evaluating online videos about PD, such as those by Kim et al. (27) and Al-Busaidi et al. (28), revealed that the majority of content was of mediocre quality and often lacked scientific rigor. Our findings are consistent with these observations, suggesting that even advanced AI systems like ChatGPT-4o struggle to meet the readability and reliability standards necessary for patient education.

Beyond PD, AI chatbots have been evaluated in other medical contexts. Zaleski et al. (9) reported that ChatGPT's exercise recommendations for fibromyalgia patients often lacked sufficient clarity and referenced sources, limiting their usefulness. Scaff et al. (10) also highlighted readability challenges in ChatGPT responses to common low back pain questions, describing the text as "moderately difficult". Fahy et al. (11) demonstrated similar concerns in the context of anterior cruciate ligament injuries, finding that response readability frequently exceeded the recommended level for patient materials. In a study of spinal cord injury information, Temel et al. (12) noted that AI responses lacked sufficient detail for practical application, echoing our findings for Parkinson's rehabilitation. These consistent patterns across multiple studies suggest that large language models, despite their potential, require further refinement for effective patient education.

One distinctive aspect of our study is its focus on PD-specific rehabilitation exercises. While prior research largely addressed general health information or diagnosis-related content, we specifically assessed balance and coordination exercise guidance, which are critical components of PD management (3,4). The ability of ChatGPT-4o to provide clear, structured exercise descriptions, with appropriate safety warnings and emphasis on consulting healthcare professionals, highlights its potential utility. However, the absence of detailed, step-by-step instructions and supporting references diminishes its applicability in clinical settings.

The strengths of our study include a comprehensive evaluation using multiple validated metrics and independent assessments by two rehabilitation specialists. To our knowledge, this is the first study to evaluate the readability, quality, and reliability of AI-generated content specifically for PD exercise training, providing a valuable foundation for future research.

Study Limitations

However, some limitations should be acknowledged. The scope was restricted to responses generated for predefined questions, which may not fully represent real-world patient queries. Additionally, the study did not assess the variability of ChatGPT-4o's responses over time, nor did it examine language adaptability for patients with low health literacy.

Future research should explore ways to enhance the clarity and personalization of AI-generated health information. Simplifying language, incorporating references, and providing more detailed exercise instructions could improve applicability. Moreover, integrating AI tools into supervised rehabilitation programs may increase patient adherence and safety. In addition, it may be valuable to investigate whether prompting ChatGPT to generate responses at a lower readability level (e.g., sixth grade or below, as recommended by the NIH and AMA) improves accessibility and comprehension among Parkinson's patients. Comparative studies evaluating this approach could provide meaningful insights for optimizing AI-based patient education tools.

Conclusion

ChatGPT-4o demonstrates potential as a supplementary tool for delivering exercise guidance to patients with PD. Nevertheless, our analysis revealed that its readability exceeds the NIH and AMA's recommended 6th-grade level, with an average FRES score of 38—corresponding to approximately 11 years of education.

This significantly limits accessibility for patients with lower health literacy. Furthermore, the lack of reliability reduces its suitability for direct clinical application. Addressing these shortcomings through model refinement, health-literacy–oriented design, and rigorous validation studies will be essential for transforming such AI systems into reliable and equitable aids for patient education and rehabilitation.

Ethics

Ethics Committee Approval: The planning, execution and data collection processes of this cross-sectional study were carried out in accordance with the approval of the relevant ethics committee (Sivas Cumhuriyet University's Ethics Committee, ethics committee no: 2025-04/67, date: 24.04.2025).

Informed Consent: Not applicable.

Footnotes

Authorship Contributions

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

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

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

References

1. Bloem BR, Okun MS, Klein C. Parkinson's disease. Lancet. 2021;397:2284-303.
2. Poewe W, Seppi K, Tanner CM, Halliday GM, Brundin P, Volkmann J, et al. Parkinson disease. Nat Rev Dis Primers. 2017;3:17013.
3. Abbruzzese G, Marchese R, Avanzino L, Pelosin E. Rehabilitation for Parkinson's disease: current outlook and future challenges. Parkinsonism Relat Disord. 2016;22(Suppl 1):S60-4.
4. Emig M, George T, Zhang JK, Soudagar-Turkey M. The role of exercise in Parkinson's disease. J Geriatr Psychiatry Neurol. 2021;34:321-30.
5. da Silva FC, da Rosa Iop R, Dos Santos PD, de Melo LMA, Gutierres Filho PJB, da Silva R. Effects of physical-exercise-based rehabilitation programs on the quality of life of patients with Parkinson's disease: a systematic review of randomized controlled trials. J Aging Phys Act. 2016;24:484-96.
6. Chow JC, Sanders L, Li K. Impact of ChatGPT on medical chatbots as a disruptive technology. Front Artif Intell. 2023;6:1166014.
7. Özbek İC. Evaluation of artificial intelligence supported osteoarthritis information texts: content quality and readability analysis. Turkiye Klinikleri J Phys Med Rehabil Sci. 2025;28:21-9.
8. Ozduran E, Hancı V, Erkin Y, Özbek İC, Abdulkerimov V. Assessing the readability, quality and reliability of responses produced by ChatGPT, Gemini, and Perplexity regarding most frequently asked keywords about low back pain. PeerJ. 2025;13:e18847.
9. Zaleski AL, Berkowsky R, Craig KJT, Pescatello LS. Comprehensiveness, accuracy, and readability of exercise recommendations provided by an AI-based chatbot: mixed methods study. JMIR Med Educ. 2024;10:e51308.
10. Scaff SPS, Reis FJJ, Ferreira GE, Jacob MF, Saragiotto BT. Assessing the performance of AI chatbots in answering patients' common questions about low back pain. Ann Rheum Dis. 2024;ard-2024-226202.
11. 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.
12. 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.
13. Parente H, Soares C, Ferreira MP, Cunha A, Guimarães F, Azevedo S, et al. ChatGPT's accuracy and patient-oriented answers about fibromyalgia. ARP Rheumatol. 2024;3:58-69.
14. Özbek İC, Hancı V, Özduran E. Digital guidance: quality and readability analysis of artificial intelligence-generated spondyloarthropathy texts. Turk J Osteoporos. 2025;31:12-8.
15. Magruder K, Rodriguez AN, Wong JCJ, Erez O, Piuzzi NS, Scuderi GR, et al. Assessing large language models in clinical settings: relevance, accuracy, and clarity. J Med Internet Res. 2024;26:12-20.
16. Li Y, Huang J, Wang J, Cheng Y. Effects of different exercises on improving gait performance in patients with Parkinson's disease: a systematic review and network meta-analysis. Front Aging Neurosci. 2025;17:1496112.
17. Lorenzo-García P, Cavero-Redondo I, De Arenas-Arroyo SN, Guzmán-Pavón MJ, Priego-Jiménez S, Álvarez-Bueno C. Effects of physical exercise interventions on balance, postural stability and general mobility in Parkinson's disease: a network meta-analysis. J Rehabil Med. 2024;56:10329.
18. Yau CE, Ho ECK, Ong NY, Loh CJK, Mai AS, Tan EK. Innovative technology-based interventions in Parkinson's disease: a systematic review and meta-analysis. Ann Clin Transl Neurol. 2024;11:2548-62.
19. Kara M, Ozduran E, Kara MM, Özbek İC, Hancı V. Evaluating the readability, quality, and reliability of responses generated by ChatGPT, Gemini, and Perplexity on the most commonly asked questions about ankylosing spondylitis. PLoS One. 2025;20:e0326351.
20. Ozduran E, Akkoc I, Büyükçoban S, Erkin Y, Hanci V. Readability, reliability and quality of responses generated by ChatGPT, Gemini, and Perplexity for the most frequently asked questions about pain. Medicine. 2025;104:e41780.
21. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the internet: caveant lector et viewor–let the reader and viewer beware. JAMA. 1997;277:1244-5.
22. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis—a wakeup call? J Rheumatol. 2012;3 9:899-903.
23. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Van Zanten SV. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol. 2007;102:2070-7.
24. 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.
25. Likert R. A technique for the measurement of attitudes. Arch Psychol. 1932;22:5-55.
26. Baran G. Evaluation of Parkinson's disease treatment information in internet. Cerrahpasa Med J. 2023;47:77-80.
27. Kim R, Park HY, Kim HJ, Kim A, Jang MH, Jeon B. Dry facts are not always inviting: a content analysis of Korean videos regarding Parkinson's disease on YouTube. J Clin Neurosci. 2017;46:167-70.
28. Al-Busaidi IS, Anderson TJ, Alamri Y. Qualitative analysis of Parkinson's disease information on social media: the case of YouTube™. EPMA J. 2017;8:273-7.