Received: 23 Jun 2025 | **Accepted:** 21 July 2025 **Genel Tıp Derg. Volume** 35/**Issue** 4 (August), 775-781
DOI: 10.54005/geneltip.1725576

ORIGINAL ARTICLE

Artificial or Authentic? A Comparative Study of Article Titles Generated by Humans and AI in Physical Medicine and Rehabilitation

Yapay mı Gerçek mi? Fiziksel Tıp ve Rehabilitasyon Alanında İnsanlar ve Yapay Zeka Tarafından Oluşturulan Makale Başlıkları Üzerine Karşılaştırmalı Bir Çalışma

[1]Lütfiye PARLAK, [1]Funda LEVENDOĞLU, [1] Elif BALEVI BATUR, [1]Ezgi AKYILDIZ TEZCAN, [1]İlknur ALBAYRAK GEZER

[1]MD, Assist. Prof., Selcuk University Faculty of Medicine, Department of Physical Medicine and Rehabilitation, Konya, Türkiye
E-mail: drlutfiyeparlak@gmail.com
[1]MD, Prof., Selcuk University Faculty of Medicine, Department of Physical Medicine and Rehabilitation, Konya, Türkiye,
E-mail: flevendoglu@gmail.com
[1]MD, Assoc. Prof., Selcuk University Faculty of Medicine, Department of Physical Medicine and Rehabilitation, Konya, Türkiye
E-mail: elifbalevi@hotmail.com
[1]MD, Assist. Prof., Selcuk University Faculty of Medicine, Department of Physical Medicine and Rehabilitation, Konya, Türkiye
E-mail: drezgiakyildiz@gmail.com
[1]MD, Prof., Selcuk University Faculty of Medicine, Department of Physical Medicine and Rehabilitation, Konya, Türkiye
E-mail: ilknur.ftr@gmail.com

Correspondence

Lütfiye PARLAK, MD, Assistant Prof.
Selcuk University Faculty of Medicine, Konya 42130 Selçuklu, Konya, Türkiye

E-Mail: drlutfiyeparlak@gmail.com

How to cite ?

Parlak L., Levendoğlu F., Balevı Batur E., Akyıldız Tezcan E., Albayrak Gezer İ., Artificial or Authentic? A Comparative Study of Article Titles Generated by Humans and AI in Physical Medicine and Rehabilitation, Genel Tıp Derg. 2025;35(4):775-781

ABSTRACT

Aim: To examine expert preferences and perceptions regarding whether article titles in the field of physical medicine and rehabilitation are generated by humans or artificial intelligence (AI), and to evaluate the selected titles in terms of scientific accuracy, comprehensibility, originality, and attractiveness.
Materials and Methods: Thirty-two article titles published in a Q1 journal indexed in the Web of Science database were selected. For each, an alternative AI-generated title was created using the ChatGPT-4o model. A survey was administered to 15 academic participants with an H-index ≥3. For each title pair, participants indicated their preference and the reason for their choice. The data were analyzed using the Statistical Package for Social Sciences (SPSS), version 22.0 software.
Results: A total of 480 title choices were recorded, with 58.1% favoring AI-generated titles. This difference was statistically significant (p=0.008). Overall, 93.3% of participants preferred AI titles more frequently. The most cited reason for preference was "comprehensibility" (47.3%). No statistically significant difference was found based on demographic variables.
Conclusions: AI-generated titles were found to be strong in terms of clarity and attractiveness. However, "scientific accuracy" was cited more frequently in favor of original titles. This indicates that while AI is effective in generating appealing titles, human contribution remains necessary for ensuring scientific rigor. Hybrid models combining AI and human input may offer a balanced approach to scientific title generation.

Keywords: Academic writing, artificial ıntelligence, physical medicine and rehabilitation

ÖZ

Amaç: Bu çalışma, fiziksel tıp ve rehabilitasyon alanında yayımlanan bilimsel makale başlıklarının insan mı yoksa yapay zekâ (YZ) tarafından üretildiğine dair akademisyen tercihlerini ve bu başlıklara yönelik algılarını incelemektedir. Başlıkların bilimsel doğruluk, anlaşılırlık, özgünlük ve dikkat çekicilik gibi nitelikler açısından değerlendirilmesi amaçlanmıştır.
Gereç ve Yöntemler: Web of Science veritabanında yer alan Q1 kategorisindeki bir dergide yayımlanan 32 makale başlığı seçilmiş ve her biri için ChatGPT-4o tarafından alternatif bir YZ başlığı üretilmiştir. Hazırlanan anket, H-indeksi ≥3 olan 15 akademisyene uygulanmıştır. Katılımcılar, her başlık çifti için tercihlerini belirtmiş ve tercihlerinin gerekçesini seçmiştir. Veriler SPSS v22 programı ile analiz edilmiştir.
Bulgular: Toplam 480 tercih yapılmış ve bunların %58,1'i YZ başlıkları yönünde olmuştur. Bu fark istatistiksel olarak anlamlı bulunmuştur (p=0.008). Katılımcıların %93,3'ü YZ başlıklarını daha fazla tercih etmiştir. En sık tercih gerekçesi "anlaşılırlık" (%47,3) olarak belirlenmiştir. Demografik değişkenlere göre anlamlı bir fark bulunmamıştır.
Sonuç: YZ başlıkları anlaşılırlık ve çekicilik açısından güçlü bulunmuştur. Ancak bilimsel doğruluk gerekçesi daha çok orijinal başlıklarda öne çıkmıştır. Bu durum, başlık üretiminde YZ'nin etkili ancak insan katkısının hâlen gerekli olduğunu göstermektedir. İnsan-YZ işbirliğine dayalı hibrit modellerin daha dengeli sonuçlar sağlayabileceği düşünülmektedir.

Anahtar Kelimeler: Yapay Zeka, Akademik Yazım, Fiziksel Tıp ve Rehabilitasyon

Peer-Review: Double anonymized - Two External
Plagiarism Checks: Yes - intihal.net
Complaints: geneltip@selcuk.edu.tr

Selcuk University Press
Genel Tıp Dergisi | e-ISSN: 2602-3741
https://dergipark.org.tr/tr/pub/geneltip
https://yayinevi.selcuk.edu.tr/

INTRODUCTION

The title of an article is the first window of an academic work to the outside world. It is critical both to attract the attention of readers and to increase the availability of the article on digital platforms (1). Artificial Intelligence (AI) is defined as a branch of computer science focusing on the development of systems that can perform tasks usually requiring human intelligence and is divided into different technical fields (2). In recent years, AI, with its current approaches such as deep learning, has become one of the focal points of research aiming to bring data assets to a new level of understanding and use (3).

With AI playing an increasing role in academic writing, even creative but technical processes such as title generation have come under the influence of automation. Large language models developed in recent years are capable not only of generating text but also of creating unique and meaningful structures such as scientific titles (4). It has been shown that AI-assisted authoring tools can create "impressive", 'entertaining', and even "scientifically valuable" content, especially in the production of titles. However, these contents have some limitations in terms of depth of meaning, contextual relevance, and educational value when compared to human-generated titles (4).

On the other hand, AI systems for knowledge synthesis, such as tools like DistillerSR, have been shown to save considerable time and reduce human errors in title and abstract screening in systematic review processes (5). However, human-specific qualities such as the specificity of the language used in title generation, contextual intelligence, and scientific intuition are still limited in current AI systems. General assessments of the use of AI in educational settings argue that human-robot collaboration is the most efficient model (6). In this context, the role of human intelligence and AI is being redefined in a creative but technical process, such as "headline writing".

This study aims to examine the preferences of the participants regarding whether the titles of scientific articles published in the field of physical medicine and rehabilitation are created by humans or AI and their perceptions of these titles in terms of scientificity, comprehensibility, creativity, and attractiveness.

MATERIALS and METHODS

The sampling process in this study was carried out through the journals belonging to the category of "Rehabilitation" in the Web of Science (WoS) database. On 05.06.2025, a total of 171 journals belonging to this category were identified. Among these, 55 journals with a Journal Impact Factor (JIF) value above 2 for 2023 were identified. Within the scope of this study, only the content obtained from publicly available articles published in the Web of Science (WoS) database was analyzed. No human or animal subjects were used, no personal data were collected, and no intervention was made to the participants. Therefore, the study does not require ethics committee approval.

Sample selection was based on a probabilistic sampling approach. In this context, repeatability was ensured by using the set.seed() function through R Studio software, and a simple random sampling method was applied. As a result of randomization, the European Journal of Physical and Rehabilitation Medicine, which

is in the Q1 category and has a Journal Impact Factor (JIF) value of 3.4 in 2023, was selected as the sample.

Original research and review articles published in the last two issues of the selected journal were analyzed, and the titles of these articles were evaluated in two different ways. Before starting the study, preliminary research was conducted to rule out the possibility that the 32 original titles in the sample had been previously proposed by any AI tool (especially ChatGPT). In this context, the GPT-4o model was made to generate titles based on the general content definition, and the generated titles were compared with the original titles. As a result of the screening, it was determined that none of the 32 titles in the sample overlapped with the titles suggested by AI. Thus, the risk that the original titles evaluated in the study were previously generated by AI was excluded, and the validity of the distinction regarding the source of the titles was ensured. One option of the abstracts in the questionnaire contains the original title of the article; the other option contains the AI-generated version of the same title. The ChatGPT-4o (GPT-4 Omni) model, developed by OpenAI and announced in 2024, was used to generate the AI-assisted titles. For each original title, the model was only given a short and directive prompt: 'Suggest a single scientific title alternative to this title', and the first and only title generated by the model was directly included in the survey form without any human intervention or reorganization. In this way, it was ensured that the evaluation was made based on the titles suggested by the model 'in one go'. In each item of the questionnaire, the AI-generated title and the original title were randomly ordered, thus differentiating the layout of the titles in each question. The academic demographic information of the participants was collected through the same form; each pair of titles was presented as a question and the participants were asked to first select their preferred title and then indicate the reason underlying this preference by marking one of the criteria such as 'comprehensibility', 'originality', 'scientific accuracy' and 'attractiveness'. The questionnaire was directed to academic experts in the field of physical medicine and rehabilitation with a qualified publication history and an H-index ≥3. Participants were not given any preliminary information about the study.

Within the scope of the present study, only the content obtained from publicly available articles published in the Web of Science (WoS) database was analyzed. No human or animal subjects were used, no personal data were collected, and no intervention was made to the participants. Therefore, the study does not require ethics committee approval.

Data analysis was performed using the Statistical Package for Social Sciences, version 22.0 software (SPSS, IBM Corp., Armonk, NY, USA). In descriptive statistics, continuous variables are presented as mean±standard deviation (SD), and categorical variables are presented as frequency (n) and percentage (%). The conformity of preference rates to normal distribution was evaluated by the Shapiro-Wilk test, and it was determined that the distribution was not normal ($p<0.05$).

The difference between the participants' AI title preference rates and the original title preference rates was analyzed with

the Wilcoxon signed-ranks test because of the paired structure. Whether the AI title preference rates differ according to demographic variables (gender, academic title, h-index, etc.) was evaluated with the Mann-Whitney U test. Participants' reasons for title preference were compared according to title type (AI/ORIGINAL), and whether the distribution of reasons was different was analyzed using the Pearson chi-square test.

Finally, the McNemar's test, requiring pairwise matched nominal data analysis, was applied to determine whether preferences based on "comprehensibility" and "originality", the two most frequently reported justifications, were related to each other within the same title pair. Only title pairs with a sufficient number of observations ($n \geq 10$) were included in this analysis. For all statistical tests, the significance level was set as $p < 0.05$.

RESULTS

A total of 15 academicians participated in the study. Demographic data of the participants are presented in Table 1. The participants made a total of 480 choices over 32 title pairs. Of these choices, 279 (58.1%) were AI-generated titles and 201 (41.9%) were original titles. According to the Wilcoxon signed-rank test results, the preference rate of AI titles was statistically significantly higher than the original titles ($n=15$, $W=15.0$, $p=0.008$). Of the participants, 93.3% ($n=14$) were detected to prefer AI titles more.

It was examined whether the AI title preference rates differed significantly according to the demographic characteristics of the participants. As a result of Mann-Whitney U tests conducted according to variables such as gender, academic title, H-index, and academic seniority, no statistically significant difference was found (all $p > 0.40$). In addition, there was no significant difference between the AI preference rates of female and male participants ($p=0.44$).

Among the reasons given by the participants after making 480 choices, the most frequently reported reason was "comprehensibility" ($n=227$, 47.3%). This was followed by "attractiveness" ($n=98$, 20.4%), "scientific accuracy" ($n=63$, 13.1%), 'originality' ($n=51$, 10.6%), and "other" reasons ($n=41$, 8.5%).

In 279 choices made in favor of AI titles, "comprehensibility" was cited as a reason for preference by 44.8% ($n=125$), and in 201 choices made in favor of original titles, 50.7% ($n=102$). In AI titles, the reason for attractiveness was reported more frequently (24.0%), while scientific accuracy and other reasons came to the fore in original titles.

The relationship between title type (AI/ORIGINAL) and the distribution of justifications was analyzed using the chi-square test, and no statistically significant difference was found ($p=0.13$). This finding reveals that the distribution of justifications is similar regardless of the title type.

The preference frequency of "comprehensibility" and "originality" justifications was analyzed comparatively for the same pair of titles. This analysis was performed using the McNemar test for title pairs with a sufficient number of data ($n \geq 10$). In 13 out of 32 title pairs (40.6%), "clarity" was reported significantly more often than "originality" ($p < 0.05$). It was observed that "originality" was not significantly dominant in any of the title pairs.

These results show that the participants prioritized the clarity, simplicity, and comprehensibility of the titles in their decision-making process. Although features such as originality and attractiveness were occasionally reported as justifications, the most dominant factor in their decision processes was the clarity of the titles.

Table 1. Demographic and Professional Characteristics of Survey Participants

Variables	Values n (%)
Number of Participants	15
Average Age (years) (±SD)	42.0±6.4
Female/Male	9/6 (60/40)
Academic Titles	
Assist. Professor	6 (40)
Assoc. Dr.	5 (33.3)
Prof. Dr.	4 (26.7)
Academic Seniority	
0-5 years	6 (40.0)
6-10 years	3 (20.0)
11-15 years	5 (33.3)
>15 years	1 (6.7)
SCI/SCI-E Publications	
1-5 publications	4 (26.7)
6-10 publications	3 (20.0)
>10 publications	7 (46.7)
H Index	
h≤4	9 (60)
h>4	6 (40)
Number of journals participants have served as reviewers	
1-5 magazines	7 (46.7)
>5 magazines	7 (46.7)
No	1 (6.7)

n: Number of participants, SD: Standard deviation

DISCUSSION

The findings of this study show that AI-generated headlines are preferred over original headlines. 93.3% of the participants preferred AI headlines; the most frequently cited reason for these preferences was "clarity" (44.8%), followed by "attractiveness" (24%). There was no statistically significant difference between the preferred title type and the reasons for preference, revealing that the participants' justifications are distributed similarly regardless of the title type.

Similarly, in the study by Marlow&Wood, the headings created by GPT-3 were evaluated positively in terms of attractiveness and comprehensibility, and their educational potential was reported to be high (4). In another study, Hamel et al. stated that AI can play not only a supportive but also a guiding role in the production of scientific content (5). In this context, the high preference for titles produced without human contribution in our study confirms the effectiveness of AI in academic writing processes.

In Sayar's review on the use of AI in medicine, it is stated that AI contributes at many stages from diagnosis to treatment and is effective in data processing processes, including scientific text production (7). Since academic titles are key to shaping the reader's first impression and facilitating access to the text, it is important to use AI effectively in this process. Given the multidisciplinary and clinically nuanced nature of PMR, AI-generated titles prioritizing clarity and comprehensibility may facilitate better accessibility of scientific findings across professional subfields. However, alignment with domain-specific terminology remains a challenge, highlighting the ongoing need for expert oversight in title generation.

In the study by Yıldıran et al., it was reported that AI titles were more preferred in some

subtopics (e.g., hand surgery); however, in general, the original titles were still superior (8). This suggests that subject-specific terminological precision may not always be achieved without human input. Similarly, in our study, the criterion of scientific accuracy was less decisive in the choice of AI titles.

The fact that there is no significant difference between the demographic variables of the participants and their AI title preferences reveals that this preference is shaped by the structural features of the title rather than individual characteristics. The fact that variables such as gender, academic seniority, number of publications, and H-index remain insignificant supports that AI-supported titles are accepted at a similar rate regardless of the level of academic expertise.

As noted in the Putra & Khodra study, AI-generated headlines can sometimes lack terminological integrity. This necessitates human supervision in areas requiring contextual evaluation (9). This finding is consistent with the low rate of "scientific accuracy" justification in favor of AI in our study.

In the same study, it was emphasized that although AI-generated content is highly preferred by users, human expertise is needed, especially in terms of "interdisciplinary consistency" and "terminological precision"(9). This finding coincides with the fact that the rationale for scientific accuracy was mentioned to a limited extent in our study.

Similarly, in another study, it was stated that although the summaries and titles created by AI were found to be more impressive formally, the level of contextual accuracy and alignment with field-specific terms was open to criticism (10). Therefore, it is suggested that human supervision should be maintained in AI-assisted text generation.

The strengths of the study are that it examines the perception of AI-supported headlines in the field of physical medicine and rehabilitation with a systematic and comparative approach. This study makes a meaningful contribution to the literature by revealing the potential of AI to generate linguistically appealing titles in the field of Physical Medicine and Rehabilitation. It also stands out as one of the pioneering studies in this area. Furthermore, it underscores the indispensable role of expert oversight in maintaining scientific accuracy and terminological precision. The principle of randomization in sample selection and the evaluation of AI-generated titles in response to a single original prompt enhances the reliability of the method. However, the study has some limitations: The limited number of participants, sampling headlines from only one journal, evaluating headlines only at the linguistic level rather than at the contextual level, and using only one AI model (ChatGPT-4o) are the main factors limiting the generalizability of the results obtained. Furthermore, relying exclusively on a single AI model may limit the applicability of findings across other language systems. Future studies comparing outputs from different AI systems could offer more comprehensive insights into their respective strengths and weaknesses.

In conclusion, this study reveals that AI-assisted headlines in the field of physical medicine and rehabilitation are effective in academic writing, especially in terms of qualities such as attractiveness and comprehensibility, but human intervention

is still needed in terms of scientific accuracy and contextual integrity. To address this shortcoming in the future, the use of hybrid systems based on human-AI collaboration in title generation has the potential to optimize both scientific accuracy and presentation quality.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We sincerely thank all the professors who generously dedicated their valuable time to completing our study.

REFERENCES

1.Jamali HR, Nikzad M. Article title type and its relation with the number of downloads and citations. Scientometrics. 2011;88(2):653-61.

2.Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, et al. Deep learning: a primer for radiologists. Radiographics. 2017;37(7):2113-31.

3.Sardanelli F, Castiglioni I, Colarieti A, Schiaffino S, Di Leo G. Artificial intelligence (AI) in biomedical research: discussion on authors' declaration of AI in their articles' title. Eur Radiol Exp. 2023;7(1):2.

4.Marlow R, Wood D. Ghost in the machine or monkey with a typewriter-generating titles for Christmas research articles in The BMJ using artificial intelligence: observational study. BMJ. 2021;375:e067732.

5.Hamel C, Hersi M, Kelly SE, Tricco AC, Straus S, Wells G, et al. Guidance for using artificial intelligence for title and abstract screening while conducting knowledge syntheses. BMC Med Res Methodol. 2021;21(1):285.

6.Chen L, Chen P, Lin Z. Artificial intelligence in education: A review. IEEE Access. 2020;8:75264-78.

7.Sayar B. Tıp alanında yapay zekânın kullanımı: araştırma makalesi. Acta Med Ruha. 2023;1(1):27-33.

8.Yıldıran G, Demirtaş S, Akdeniz H, Tosun Z. Which is superior in creating scientific article titles in plastic surgery: human intelligence or artificial intelligence? Genel Tip Derg. 2025;35(2):222-5.

9.Putra JWG, Khodra ML. Automatic title generation in scientific articles for authorship assistance: a summarization approach. J ICT Res Appl. 2017;11(3):253-67.

10.Huang J, Tan M. The role of ChatGPT in scientific communication: writing better scientific review articles. Am J Cancer Res. 2023;13(4):1148.