ORIGINAL ARTICLE

CONTENTS

AI-driven rehabilitation: evaluation of ChatGPT-4o for generating personalized physical rehabilitation plans in comorbid patients

Yaroslav Mykhalko[1], Svitlana Dyditska[2], Larisa Balatska[2], Felix Filak[1], Yelyzaveta Rubtsova[1]
[1]UZHHOROD NATIONAL UNIVERSITY, UZHHOROD, UKRAINE
[2]YURIY FEDKOVYCH CHERNIVTSI NATIONAL UNIVERSITY, CHERNIVTSI, UKRAINE

ABSTRACT

Aim: To evaluate the performance of ChatGPT-4o in creating personalized physical rehabilitation plans for comorbid patients.

Materials and Methods: ChatGPT-4o was employed to generate physical rehabilitation plans for 50 clinical cases of comorbid patients. These plans were evaluated independently by two experts according to 6 criteria using a 5-point Likert scale. Experts also classified each plan regarding its suitability for use into 3 categories: "Completely unsuitable for use", "Suitable for use with corrections", "Completely suitable for use". Statistical analysis included the Mann–Whitney U test, intraclass correlation coefficient (ICC) and linear weighted Cohen's kappa (k_w). The statistical significance was set at $p<0.05$

Results: The overall mean score of ChatGPT-4o generated rehabilitation plans was 4.30±0.28 with the highest scores for respiratory and musculoskeletal pathology (4.37±0.36 and 4.33±0.24, respectively). Among the evaluation criteria, the highest indicators were observed for Clinical accuracy and Safety (4.59±0.59 and 4.41±0.71, respectively). 72.00% of the generated plans were classified as "Suitable for use with corrections". None of the plans were identified as "Completely unsuitable for use". The agreement percentage ranged from 84% to 90%, ICC values were 0.80-0.86, and overall suitability k_w was 0.77

Conclusions: LLM-generated rehabilitation plans show promise as supportive tools in clinical practice, but they are not yet at a stage where they can be implemented without expert review and modification. The high overall inter-rater reliability provides confidence in the evaluation process, while also highlighting areas for improvement in both the LLM's performance and the assessment methodology.

KEY WORDS: ChatGPT-4o, large language model, performance, physical rehabilitation

Wiad Lek. 2025;78(4):753-759. doi: 10.36740/WLek/203850 DOI

INTRODUCTION

Over the past few years, the phrases "Artificial Intelligence" and "Large Language Models" have become increasingly common both in everyday conversation and in the pages of scientific journals. According to Cambridge dictionary, the large language model (LLM) is "a complex mathematical representation of language that is based on very large amounts of data and allows computers to produce language that seems similar to what a human might say" [1]. These models represent a significant advancement in artificial intelligence, particularly in the realm of natural language processing (NLP). They have been trained on extensive datasets using deep learning algorithms, particularly transformer architectures. LLMs are capable of performing various NLP tasks, including recognizing, generating, translating, and summarizing text. The training process involves learning statistical relationships from vast amounts of text data, allowing the model to understand context, syntax, and semantics inherent in human language. The architecture of LLMs typically includes multiple layers of neural networks that process input data in parallel, significantly enhancing their efficiency compared to earlier models like recurrent neural networks (RNNs).

Over the last 60 years LLMs made a great journey from the first pattern matching ELIZA to modern GPT-4 and LLaMA that are pushing the boundaries of AI and are capable of writing, conversing, summarizing, and translating with a sophistication that closely mirrors human intelligence. The number of released LLMs is constantly rising every year. Thus, in only 2023 more than 20 such models were introduced [2, 3].

LLMs are successfully used in various fields of activity from media and entertainment to financial management. The adoption of LLMs is projected to increase significantly across various industries from 2022 to 2026. While the technology sector leads with an expected adoption rate of 85% by 2026, other industries such as finance, retail, and manufacturing are also anticipated to see substantial growth in LLM integration [4].

Table 1. Evaluation criteria and 5-point likert scale for assessing ChatGPT-4o-generated rehabilitation plans

Criterion and it's description	5-point Likert scale
Individualization: The extent to which the plan was tailored to the specific needs and conditions of the patient.	1 - No personalization evident 2 - Minimal consideration of patient-specific factors 3 - Basic tailoring to patient needs 4 - Well-tailored plan with some unique considerations 5 - Highly personalized plan addressing all patient-specific factors
Clinical accuracy: The correctness of the interventions proposed, based on current clinical guidelines and best practices.	1 - Contains significant clinical errors 2 - Some clinical inaccuracies present 3 - Generally accurate but lacks depth 4 - Clinically sound with minor oversights 5 - Demonstrates high-level clinical knowledge and accuracy
Safety: Assessment of potential risks and the inclusion of appropriate safety precautions.	1 - Potentially harmful recommendations 2 - Inadequate safety considerations 3 - Basic safety measures included 4 - Comprehensive safety protocols with minor gaps 5 - Exceptional attention to safety, covering all potential risks
Progressive design: The logical progression of rehabilitation activities over the one-month period, ensuring a gradual increase in intensity and complexity.	1 - No progression in difficulty or intensity 2 - Minimal progression evident 3 - Basic progression structure present 4 - Well-structured progression with some refinement needed 5 - Optimal progression design tailored to patient capabilities
Feasibility and accessibility: The practicality of implementing the plan given the patient's likely resources, environment, and support system.	1 - Impractical or inaccessible for the patient 2 - Limited consideration of patient's circumstances 3 - Moderately feasible and accessible 4 - Highly feasible with minor accessibility concerns 5 - Perfectly aligned with patient's resources and circumstances
Focus on the result: The clarity and specificity of the expected outcomes, with a focus on measurable improvements in the patient's condition.	1 - No clear goals or outcome measures 2 - Vague or inappropriate goals 3 - Basic outcome-oriented approach 4 - Clear, relevant goals with some room for improvement 5 - Comprehensive, patient-centered goals with clear outcome measures

The use of this technology in healthcare is also promising. It was shown that LLMs can provide differential diagnoses and suggest potential treatments, improve the accuracy and speed of medical decision-making, transcribe and summarize patient interactions, answer patient queries, provide health advice, identify new drug candidates and predict their potential efficacy, create personalized treatment plans by analyzing individual patient data etc [5-9]. The evaluation of the LLMs application in health care is carried out by researchers from different positions. Thus, the most common task is assessing medical knowledge, while making diagnoses and educating patients are less common. More than 80% of studies focused on question answering, with fewer on summarization, conversational dialogue, and translation. Accuracy of LLMs responses is the primary focus in the majority of studies, while fairness, bias, toxicity, robustness, and deployment considerations are less frequently measured. Among all medical specialties LLMs application in internal medicine, surgery, and ophthalmology were the most studied areas [6]. Physical rehabilitation is one of the medical fields that is often overlooked by LLMs researchers.

Physical rehabilitation plays a vital role in the comprehensive treatment approach for patients across a wide spectrum of diseases and conditions. It serves as a critical component in restoring function, improving quality of life, and promoting overall well-being. Effective rehabilitation can significantly reduce healthcare costs by preventing complications, reducing hospital readmissions, and decreasing dependency on long-term care services.

Physical rehabilitation specialists encounter several challenges in their practice, which impact patient care and outcomes. This includes but is not limited to complexity and diversity of patient conditions, resource limitations, patient adherence to rehabilitation programs, telerehabilitation techniques, designing personalized rehabilitation plans [10, 11]. Involving the LLMs in routine practice of physical rehabilitation specialists may help to overcome these issues and significantly increase the efficacy of rehabilitation programs. Although the integration of such tools offers promising advantages, it is crucial to consider potential drawbacks and constraints associated with their use. Paramount among these considerations is

Table 2. ChatGPT-4o performance in generating rehabilitation plans across various pathology classes and evaluation criteria, M±SD

Pathology class	Individualization	Clinical accuracy	Safety	Progressive design	Feasibility and accessibility	Focus on Results	Overall Mean
Musculoskeletal	4.35±0.60	4.55±0.61	4.60±0.61	4.30±0.75	4.00±0.71	4.20±0.60	4.33±0.24
Nervous	4.35±0.81	4.70±0.47	4.60±0.68	4.00±0.79	3.95±0.69	4.00±0.73	4.27±0.23
Respiratory	4.40±0.75	4.75±0.44	4.55±0.60	4.10±0.72	3.85±0.75	4.55±0.69	4.37±0.36
Cardiovascular	4.50±0.51	4.45±0.60	4.50±0.69	4.30±0.57	4.05±0.83	3.95±0.89	4.29±0.26
Digestive	4.20±0.70	4.50±0.76	3.80±0.70	4.35±0.75	4.60±0.6	4.00±0.65	4.24±0.30
Overall Mean	4.36±0.67	4.59±0.59*	4.41±0.71#	4.21±0.71	4.09±0.75	4.14±0.74	4.30±0.28

Note. * - the difference is statistically significant compared to all other criteria ($p<0.05$),
- the difference is statistically significant compared to all criteria except Individualization ($p<0.05$).

the necessity to guarantee the accuracy, effectiveness, and clinical appropriateness of rehabilitation programs developed with LLM assistance.

AIM

The aim of this study was to evaluate the performance of ChatGPT-4o in creating personalized physical rehabilitation plans for comorbid patients.

MATERIALS AND METHODS

A total of 50 clinical cases were utilized in this study. Each case contained a comprehensive description of the patient's current condition, past medical history, primary diagnosis, and accompanying comorbidities. The cases were carefully selected to represent a diverse range of medical conditions, with an equal distribution across five major physiological systems: musculoskeletal system (10 cases), nervous system (10 cases), respiratory system (10 cases), cardiovascular system (10 cases), digestive system (10 cases). All patient data used in the clinical cases were anonymized to protect privacy.

LLM ChatGPT-4o (OpenAI) was employed to generate individualized physical rehabilitation plans for all 50 clinical cases. It was prompted to create plans that had to cover a one-month period and be tailored to each patient's specific condition and needs.

Two experts in physical rehabilitation independently evaluated the rehabilitation plans generated by the LLM on six key criteria: Individualization, Clinical accuracy, Safety, Progressive design, Feasibility and accessibility, and Focus on result. A 5-point Likert scale was employed for each criterion (Table 1).

To ensure consistency in scoring, the experts were provided with detailed rubrics describing the expectations for each score level across all criteria. They underwent a calibration session before the evaluation process to align their understanding of the scoring system.

In addition to the detailed evaluation, the experts provided a final conclusion regarding the overall suitability for use of each rehabilitation plan using the following scale: "Completely unsuitable for use", "Suitable for use with corrections", "Completely suitable for use". If the experts' conclusions differed, the worst one was taken into account when calculating the final frequency of the rehabilitation plan's overall suitability.

Statistical analysis was performed to assess the ChatGPT-4o performance. To analyze the effectiveness of the studied LLM in the generation of physical rehabilitation plans, mean scores were calculated. The mean scores presented as M±SD. The Mann–Whitney U test was used to compare the differences between the evaluation criteria. Additionally, the frequency of each overall suitability assessment was tabulated.

Inter-rater reliability was assessed using agreement percentage, intraclass correlation coefficient (ICC) with 95% confidence intervals (CI) and linear weighted Cohen's kappa (k_w) to ensure consistency between the two experts' evaluations. ICC values <0.5 were considered as poor reliability, 0.5-0.75 – moderate reliability, 0.75-0.9 – good reliability, and >0.90 – excellent reliability [12]. k_w values were interpreted as <0.0, poor; 0.0-0.2, slight; 0.2-0.4, fair; 0.4-0.6, moderate; 0.6-0.8, substantial; and 0.8-1.0 almost perfect agreement [13].

Mean (M) and standard deviations (SD) were calculated using statistical package Statistica 12 (TIBCO Software Inc., USA). k_w values were calculated using Microsoft Excel 2016 software (Microsoft Corporation, Redmond, WA, USA), while the ICCs with their 95% CIs were computed using the web-based tool StatsToDo. (https://www.statstodo.com/IntraclassCorrelation.php). The statistical significance of all tests was set at $p<0.05$.

RESULTS

The physical rehabilitation plans generated by ChatGPT-4o, based on the provided clinical case descriptions,

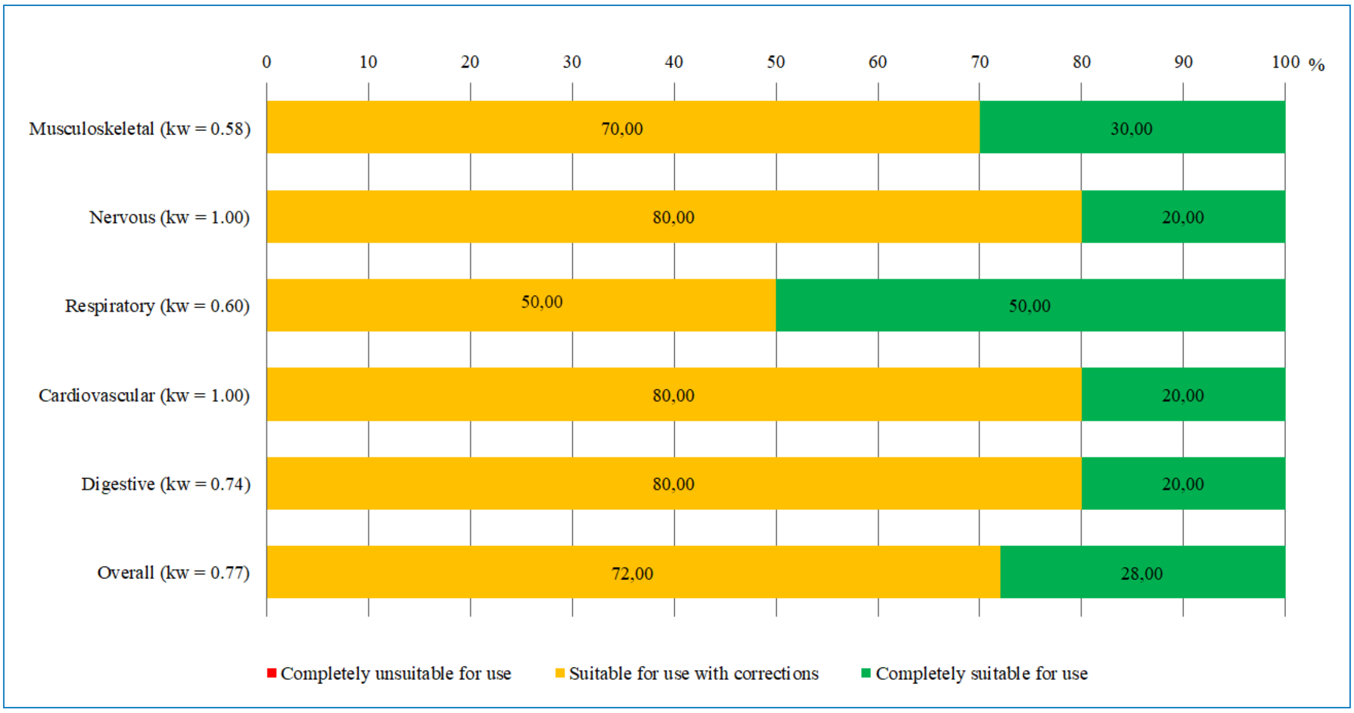

Fig. 1. Frequency distribution of rehabilitation plans' overall suitability assessments across the pathology groups (%) with inter-rater agreement (Cohen's weighted kappa, k_w).

were highly evaluated by the experts. The average ratings were consistently above 4.0 on the proposed 5-point Likert scale, both across the pathology classes and evaluation criteria with the overall mean score of 4.30±0.28 (Table 2).

In terms of pathology classes, LLM showed the highest efficiency in developing rehabilitation plans for patients with primary pathology of respiratory and musculoskeletal systems (4.37±0.36 and 4.33±0.24, respectively). The effectiveness in the development of rehabilitation plans for pathologies of other systems was slightly worse. However, there was no statistically significant difference between the obtained indicators ($p<0.05$).

Among evaluation criteria, the Clinical accuracy received the highest overall mean score (4.59±0.59, $p<0.05$) surpassing all other criteria. The difference was statistically significant. Clinical accuracy was also the highest for plans developed for the rehabilitation of patients with the respiratory and nervous systems pathology (4.75±0.44 and, 4.70±0.47 respectively).

The overall mean score of Safety criterion was somewhat lower (4.41±0.71). However, the average score obtained by this criterion was statistically significantly higher compared to the Progressive design, Feasibility and accessibility, and Focus on result criteria ($p<0.05$). In terms of pathology classes, the Safety criterion for plans for diseases of the musculoskeletal system was rated the highest, and the digestive system was the lowest (4.60±0.61 and 3.80±0.70, respectively).

The overall scores for the criteria Individualization, Progressive design, Feasibility and accessibility, and Focus on result were also quite high, but there was no statistically significant difference between them ($p>0.05$). It is worth noting that the Feasibility and accessibility criterion had the lowest overall mean score (4.09±0.75), but was rated the highest by experts regarding rehabilitation plans for digestive system pathology.

According to the study design the experts also were prompted to make a conclusion on the suitability of physical rehabilitation plans for use. In cases of disagreement between experts' opinions, the final distribution of frequencies took into account a more conservative conclusion (Fig. 1).

The general analysis of the conclusions obtained in this way showed that in most cases (72.00%) these plans were classified as "Suitable for use with corrections". 80.00% of the plans generated for cases of nervous, cardiovascular and digestive systems pathology received this rating. In only 14 (28.00%) cases, both experts independently identified the proposed rehabilitation plans as "Completely suitable for use". Most often (50.00% of cases) such a high rating occurred among rehabilitation plans for patients with respiratory system pathology. It is important to note that none of the plans was classified as "Completely unsuitable for use", indicating a baseline level of acceptability for all plans created by the studied LLM.

The level of agreement between the assessments provided by the experts was high across all evaluation

Table 3. Inter-Rater Reliability Metrics for Expert Evaluations

Evaluation Criteria	Agreement, %	ICC, 95% CI
Individualization	84.00	0.83, 0.71-0.90
Clinical accuracy	86.00	0.80, 0.65-0.89
Safety	86.00	0.86, 0.77-0.92
Progressive design	86.00	0.86, 0.77-0.92
Feasibility and accessibility	88.00	0.84, 0.74-0.91
Focus on Results	84.00	0.85, 0.76-0.91
Overall Suitability for use	90.00	0.78, 0.64-0.87

Note. ICC – intraclass correlation coefficient, CI – confidence interval

criteria. The agreement percentage ranged from 84.00% to 90.00% (Table 3).

In terms of evaluation criteria, the highest percentage of agreement was observed for the Feasibility and accessibility criterion (88.00%). ICC values ranged from 0.80 (CI: 0.65-0.89) to 0.86 (CI: 0.77-0.92). k_w values for the overall suitability of rehabilitation plans across pathology classes ranged from 0.58 to 1.00 with the overall k_w of 0.77 (Fig.1).

DISCUSSION

Upon its release in open access, ChatGPT has been the subject of numerous scientific studies aimed at exploring its performance in a wide variety of medical tasks [14, 15]. The findings obtained in our study highlight the potential of LLMs in creating effective rehabilitation plans even for patients with various comorbidities. The rehabilitation plans proposed by ChatGPT-4o for patients with various pathologies have been highly rated by experts across various evaluation criteria which in turn underscores the sufficient quality of these plans. The highest mean scores of Clinical accuracy and Safety criteria indicate that the model has been trained on a sufficient amount of modern medical data in this field and can effectively use it. On the other hand, it is able to anticipate and mitigate potential risks associated with rehabilitation activities. These findings are extremely important to ensure the quality of the proposed interventions. The high overall score obtained by ChatGPT-4o in this study suggests that it can be used effectively in clinical settings, especially in the case of limited human resources. The studied LLM showed equally high efficiency in the rehabilitation plans development for patients with pathologies of various systems. This suggests a certain level of versatility of its application and the possibility of use in different clinical scenarios. Such high performance, as found in our research, is supported by other studies with similar objectives but different designs and versions of this LLM [16-18].

Despite ChatGPT-4o has demonstrated a strong ability to generate individualized, accurate, and safe rehabilitation plans, there remains room for improvement in ensuring consistent progressive design and aligning rehabilitation plans with patient resources and environments.

There is no doubt that evaluating rehabilitation plans against pre-defined criteria is important and reflects a more formalized approach. However, it is equally important to take into account the expert's general impression of the analyzed rehabilitation plan about its suitability for use as criterion referenced assessment cannot fully reflect all the nuances. On the other hand, creating a generalized conclusion requires the expert not only to use certain knowledge, but also professional experience, which is extremely important. As noted earlier, a conservative approach was used in calculating the final frequency distribution of usability judgments in cases of disagreement between experts. Although this approach may slightly underestimate the final result, this decision was made deliberately, giving priority to patient safety and the quality of the plan. Under these circumstances, the significant number of plans categorized as "Suitable for use with corrections" suggests that although the LLM can create relevant and potentially beneficial rehabilitation plans, there remains a considerable necessity for professional evaluation and adjustment to ensure these plans meet the specific needs of individual patients. The higher rate of completely suitable plans for respiratory conditions (50%) could be due to the more standardized nature of pulmonary rehabilitation practices or a larger volume of used training data in this domain.

Inter-rater reliability indicators play an important role in the analysis of the results of studies with raters involvement. A high percentage of agreement between experts' assessments obtained in our research was revealed, both on separate criteria and on the indicator of the general suitability of rehabilitation plans for use. These results were confirmed by fairly high ICC values, indicating good reliability of the results. The relatively narrow 95% confidence intervals for these ICCs further confirm the robustness of the agreement. These results suggest that the peer review process was reliable, providing a solid basis for evaluating the quality of rehabilitation plans produced by ChatGPT-4o.

The observed variability in inter-rater agreement across pathology classes, as evidenced by the weighted Cohen's kappa values ranging from moderate (k_w = 0.58) to almost perfect agreement (k_w = 1.00), warrants careful interpretation. This heterogeneity in agreement levels may be attributed to several factors such as small sample size within each pathology class, complexity of pathology-specific rehabilitation, expertise bias or quality variability in LLM output and underscore the needs of further research in this field. Despite this, the overall k_w of 0.77 demonstrates substantial agreement and suggests that the experts generally concurred in their evaluations of the rehabilitation plans generated by ChatGPT-4o, despite the variability across different pathology classes. However, the discrepancies within specific classes highlight the necessity for continued refinement in both the LLMs' output and the evaluation criteria, particularly in addressing the nuances of complex cases.

CONCLUSIONS

This study provides valuable insights into the potential of ChatGPT-4o in generating physical rehabilitation plans across various pathology classes. The overall performance of LLM according to the evaluation criteria was 4.30±0.28 points out of 5.00. Rehabilitation plans generated for patients with respiratory system diseases received the highest rating (4.37±0.36). Clinical accuracy and Safety showed the best results among all criteria (4.59±0.59 and 4.41±0.71, respectively).

None of the created rehabilitation plans was rated by experts as "Completely unsuitable for use". At the same time, 72% were categorized as "Suitable for use with corrections" which underscores the continued necessity of expert oversight. Across different pathology classes, respiratory and musculoskeletal rehabilitation plans received the highest rate of overall suitability for use.

The inter-rater reliability analysis showed strong agreement between experts in their assessments, as confirmed by high values of agreement percentage, ICC and k_w.

The findings of our study demonstrate both the promise and the current limitations of LLMs usage in the field of physical rehabilitation and aligns with the current understanding that AI-generated medical content should not replace but rather augment clinical expertise.

REFERENCES

1. Large language model. Cambridge english dictionary. https://dictionary.cambridge.org/dictionary/english/large-language-model [Accessed 20 December 2024]
2. Chu Z, Ni S, Wang Z et al. History, development, and principles of large language models-an introductory survey. ArXiv, abs/2402.06853. 2024. doi: 10.48550/arXiv.2402.06853. DOI
3. Naveed H, Khan AU, Qiu S et al. A comprehensive overview of large language models. ArXiv, abs/2307.06435. 2023. doi: 10.48550/arXiv.2307.06435. DOI
4. Peringanji D. The impact of large language models (llms) on everyday applications: opportunities, challenges, and considerations. International Research Journal of Engineering and Technology. 2024;11(5):1150-1158.
5. Hadi MU, Al-Tashi Q, Qureshi R et al. Large Language Models: A comprehensive survey of applications,challenges, datasets, limitations, and future prospects. Techrxiv. 2024. Preprint. doi:10.36227/techrxiv.23589741.v6. DOI
6. Bedi S, Liu Y, Orr-Ewing L et al. A systematic review of testing and evaluation of healthcare applications of large language models (LLMs). medRxiv. 2024. doi: 10.1101/2024.04.15.24305869. DOI
7. Mumtaz U, Ahmed A, Mumtaz S. LLMs-healthcare: current applications and challenges of large language models in various medical specialties. Artif Intell Health. 2024;1(2):16–28. doi:10.36922/aih.2558. DOI
8. Nazi ZA, Peng W. Large language models in healthcare and medical domain: A review. Informatics. 2024;11(3):57. doi: 10.13140/RG.2.2.18450.43209. DOI
9. Thirunavukarasu AJ, Ting DSJ, Elangovan K et al. Large language models in medicine. Nat Med. 2023;29(8):1930-1940. doi:10.1038/s41591-023-02448-8. DOI
10. Aderonmu JA. Emerging challenges in meeting physiotherapy needs during COVID-19 through telerehabilitation. Bulletin of Faculty of Physical Therapy. 2020;25(1):16. doi:10.1186/s43161-020-00018-4. DOI
11. Naqvi WM, Shaikh SZ, Mishra GV. Large language models in physical therapy: time to adapt and adept. Front Public Health. 2024. doi:10.3389/fpubh.2024.1364660. DOI
12. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012. DOI

13. Gwet K. Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters. Oxford: Advanced Analytics, LLC, Gaithersburg. 2014, pp. 62-65.
14. Viswanathan VK, Jain VK, Vaish A et al. Chatbots and their applications in medical fields: current status and future trends: a scoping review. Apollo Med. 2024. doi:10.1177/09760016241259851. DOI
15. Wu J, Ma Y, Wang J et al. The application of chatgpt in medicine: a scoping review and bibliometric analysis. J Multidiscip Healthc. 2024;17:1681-1692. doi:10.2147/JMDH.S463128. DOI
16. Menekşeoğlu AK, İş EE. Comparative performance of artificial ıntelligence models in physical medicine and rehabilitation board-level questions. Rev Assoc Med Bras (1992). 2024;70(7):e20240241. doi:10.1590/1806-9282.20240241. DOI
17. Rutkowski SM, Galán-Mercant A. Artificial intelligence in rehabilitation: comparing OpenAI ChatGPT recommendations to evidence-based meta-analyses. Physiother Rev. 2023;27(4):5-16. doi:10.5114/phr.2023.133714. DOI
18. Mittal K, Dhar M. Use of ChatGPT by physicians to build rehabilitation plans for the elderly: A mini-review of case studies. J Indian Acad Geriatr. 2023;19(2):86-93. doi:10.4103/jiag.jiag_27_23. DOI

CONFLICT OF INTEREST
The Authors declare no conflict of interest

CORRESPONDING AUTHOR
Yaroslav Mykhalko
Uzhhorod National University
3 Narodna Sq., 88000 Uzhhorod, Ukraine
e-mail: yaroslav.myhalko@uzhnu.edu.ua

ORCID AND CONTRIBUTIONSHIP
Yaroslav Mykhalko: 0000-0002-9890-6665 A B D F
Svitlana Dyditska: 0000-0001-5015-2465 B C D
Larisa Balatska: 0000-0002-7963-2726 A C D
Felix Filak: 0000-0001-7595-5416 B E
Yelyzaveta Rubtsova: 0000-0001-9395-1822 B E

A – Work concept and design, B – Data collection and analysis, C – Responsibility for statistical analysis, D – Writing the article, E – Critical review, F – Final approval of the article

RECEIVED: 21.12.2024
ACCEPTED: 28.03.2025