SHORT COMMUNICATION

USE OF ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODELS AS A CLINICAL TOOL IN REHABILITATION MEDICINE: A COMPARATIVE TEST CASE

Liang ZHANG, MD, PhD1*, Syoichi TASHIRO, MD, PhD1,2*, Masahiko MUKAINO, MD, PhD3 and Shin YAMADA, MD, PhD1

From the 1Department of Rehabilitation Medicine, Kyorin University School of Medicine, Tokyo, 2Department of Rehabilitation Medicine, Keio University School of Medicine, Tokyo, and 3Department of Rehabilitation Medicine, Hokkaido University Hospital, Tokyo, Japan

*These authors contributed equally to this paper.

Objective: To explore the potential use of artificial intelligence language models in formulating rehabilitation prescriptions and International Classification of Functioning, Disability and Health (ICF) codes.

Design: Comparative study based on a single case report compared to standard answers from a textbook.

Subjects: A stroke case from textbook.

Methods: Chat Generative Pre-Trained Transformer-4 (ChatGPT-4) was used to generate comprehensive medical and rehabilitation prescription information and ICF codes pertaining to the stroke case. This information was compared with standard answers from textbook, and 2 licensed Physical Medicine and Rehabilitation (PMR) clinicians reviewed the artificial intelligence recommendations for further discussion.

Results: ChatGPT-4 effectively formulated rehabilitation prescriptions and ICF codes for a typical stroke case, together with a rationale to support its recommendations. This information was generated in seconds. Compared with standard answers, the large language model generated broader and more general prescriptions in terms of medical problems and management plans, rehabilitation problems and management plans, as well as rehabilitation goals. It also demonstrated the ability to propose specified approaches for each rehabilitation therapy. The language model made an error regarding the ICF category for the stroke case, but no mistakes were identified in the ICF codes assigned.

Conclusion: This test case suggests that artificial intelligence language models have potential use in facilitating clinical practice and education in the field of rehabilitation medicine.

Key words: artificial intelligence; large language models; rehabilitation prescriptions; International Classification of Functioning, Disability and Health codes.

Accepted Jul 5, 2023; Published Sep 11, 2023
DOI: 10.2340/jrm.v55.13373

Correspondence address: Shin Yamada, Department of Rehabilitation Medicine, Kyorin University School of Medicine, 6-20-2, Shinkawa, Mitaka-shi, Tokyo 181-8611, Japan. E-mail: yamada-shin@ks.kyorin-u.ac.jp

Artificial intelligence (AI) is a rapidly evolving and dynamic field with the potential for widespread use in healthcare. ChatGPT, an AI tool developed by OpenAI in November 2022 (1), uses advanced algorithms and machine learning techniques and has potential for use in many applications within the medical field (2–4). Numerous exploratory efforts have also assessed its use in evidence-based medicine (5) and obstetrics and gynaecology (6), identifying both advantages and limitations in its use in these fields.

The potential application of ChatGPT in the field of Physical Medicine and Rehabilitation (PMR) should be explored. PMR is a specialized field of medical practice that focuses on improving impaired function and preventing secondary disorders. PMR requires multidisciplinary collaboration and communication through a teamwork model. As coordinators, PMR clinicians must possess extensive knowledge and experience to manage and organize information from various specialties, including physical therapy, occupational therapy, speech-language pathology, prosthetics and orthotics, nutrition, and psychology. This equips them to integrate information from these specialties and develop comprehensive rehabilitation prescriptions. However, it is not known whether the integration of large language models (LLMs) into this multidisciplinary team management model could improve productivity in the context of rehabilitation clinical practice. Therefore, the aim of this single case report study is to investigate...
the potential of the new AI LLM in the field of rehabilitation medicine, focusing specifically on its ability to facilitate the creation of rehabilitation prescriptions and International Classification of Functioning, Disability and Health (ICF) codes.

### METHODS

A stroke case from textbook (7) was utilized for this qualitative evaluation. All responses were generated by ChatGPT-4, the most advanced LLM from OpenAI, released on 15 March 2023 (1). Prompts were provided to ChatGPT-4 to generate information based on this stroke case, including medical problems, medical management plans, rehabilitation problems, rehabilitation management plans, risk management plans, and rehabilitation goals. ChatGPT-4 was also instructed to generate physical therapy prescriptions, and occupational prescriptions, and advice was sought on speech-language therapy and assistive device usage. In addition, recommendations were requested regarding psychology, sociology, interpersonal relationships, and other relevant factors requiring intervention. Finally, ChatGPT was instructed to generate ICF codes for the patient. A standard approach from the textbook (7) was used for comparison. 2 licensed PMR clinicians reviewed the recommendations generated by ChatGPT and provided comments regarding each item, in order to facilitate further discussion.

**Case report**

A 62-year-old man presenting with sudden onset of difficulty moving his left hand presented to our hospital on foot. He was diagnosed with cerebral infarction in the right precentral gyrus. Prior to onset he was independent in his activities of daily living (ADL) and gait, and lived with his wife and daughter, and worked in a bank. His usual work included object transferring with both hands, less-frequent document writing, and commuting 60 min by bus and train. He was taking medication for hypertension and diabetes mellitus.

Two days post-onset, antiplatelet treatment was initiated for probable thrombosis. The inpatient treatment and follow-up were estimated to take 2–3 weeks. When rehabilitation was initiated, he had a blood pressure of 141/71 mmHg, a heart rate of 82 bpm, with no arrhythmia; motor and sensory disturbances: left hemiparesis with stroke Impairment Assessment Set (SIAS) motor scores (8): knee-mouth test 4, finger Function test 4, hip-flexion test 5, knee-extension test 5, foot-plant test 5; grip strength: right 32 kg, left 19 kg; range of motion of joints: no restriction; abnormal reflex: clonus –, Babinski reflex –, tendon reflexes of the patella and Achilles tendon, attenuated; sensory disturbance: thermal sensation, mildly impaired; touch sensation, mildly impaired; and proprioception, intact. His motor functionality was independent, and gait abnormality was not observed. His balance function was intact. His ADL were independent with self-correction (he felt difficulty when lifting an object with his left hand). The patient wanted to return to work as early as possible.

### RESULTS

Prompts were provided to GPT-4 to generate information based on this stroke case, including medical problems, medical management plans, rehabilitation problems, rehabilitation management plans, risk management plans, and rehabilitation goals (Table I, see details in Appendix S1). A thorough review of the standard clinical ICF code assigned by 2 PMR clinicians was then conducted, comparing it with the table produced by the GPT-4 model (Table II). The 3-digit codes generated by the LLM were accurate, aptly representing the aspects of functioning, disability, and health. No discrepancies were found in the ICF code concerning the inclusion criteria. Furthermore, the personal factors, which were not delineated in the ICF code, were included in the description. However, an error was found when reviewing the case record in the body structures category (s730). The patient had had a stroke, and the original impairment should have been classified as affecting the right precentral gyrus (s110.1), as outlined in the case section. Instead, the table displayed the damage as being in “the upper extremity, left hand.”

### DISCUSSION

In this case study, ChatGPT pinpointed crucial rehabilitation problems, such as motor/sensory impairment and ADL difficulty, and formulated reasonable, yet general, rehabilitation prescriptions (Table I), providing rationale to support its recommendations (Appendix S1). In addition, the LLM automatically generated ICF codes (Table II) within seconds, based on the key problems identified from the patient.

Whereas application of the ICF has the potential to bring ground-breaking changes to the field of rehabilitation, it is time-consuming to apply, which hinders its widespread adoption in clinics (9). To tackle this issue, researchers have been working on developing a supportive tool (10). We believe that the advancement of LLM technology, exemplified by ChatGPT, has the potential to make application of the ICF more practical.

In this comparative study PMR clinicians identified errors and inaccuracies (Table II and Appendix S1) in
Use of artificial intelligence large language models

Appendix S1. Clinicians must construct prompts that provide specific and appropriate approaches based on prognosis prediction. However, the majority of ChatGPT’s responses aligned with recent rehabilitation management requirements for this stroke case. These findings indicate the potential of LLMs, such as ChatGPT, to be integrated into workflow systems or education in the field of rehabilitation medicine as a tool for use by human clinicians.

It is important to note that the stroke case mentioned is merely a straightforward textbook example, whereas real-world clinical medicine and rehabilitation involve more complex and uncertain circumstances, both medically in the acute phase and holistically in the chronic phase. Further evaluation is required to assess the ability of ChatGPT to handle intricate scenarios in a clinical rehabilitation setting. In addition, it is crucial to emphasize the significance of using specific and appropriate prompts for generating accurate outcomes (Appendix S1). Clinicians must construct prompts that are understandable to ChatGPT when composing their inputs (2). Furthermore, caution must be exercised as language models such as ChatGPT may occasionally generate “hallucinations”, resulting in nonsensical
or misleading responses (2). Therefore, LLMs are not suitable for replacing human decision-making in rehabilitation medical practice.

Despite the potential of pre-trained models for various applications, LLMs lack the ability to update their knowledge, as they are trained on large datasets up to 2021, and their learning process halts once training is completed (5,6). Moreover, ChatGPT, being grounded in general knowledge, lacks specificity in rehabilitation medicine, limiting its application in this area. To enhance its applicability, specific training in rehabilitation medicine may be necessary for this pre-trained LLM.

Although the application of ChatGPT or other LLMs in rehabilitation medicine is currently at an early stage, we conclude that the potential benefits are promising as AI technology continues to advance.

In conclusion, ChatGPT shows potential for applications in both clinical practice and education within rehabilitation medicine. However, limitations such as lack of specificity in rehabilitation medicine, exist. Currently, it should not be used to replace the expertise of medical professionals or independently judge medical issues. Future developments, such as specialized models, may enhance its applicability.

**ACKNOWLEDGEMENTS**

This work was supported by The Education/Research Fund of Kyorin University School of Medicine to ZL, ST and SY, JSPS KAKENHI (grant number JP20K16467), The Nakatomi Foundation, The General Insurance Association of Japan, The Daiwa Securities Health Foundation to ST.

**REFERENCES**

7. Yamada S. Tips in assessment of rehabilitation learn from 100 cases; 2013.