J Rehabil Med 2023; 55: jrm13373

SHORT COMMUNICATION

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

Liang ZHANG, MD, PhD[1#], Syoichi TASHIRO, MD, PhD[1,2#], Masahiko MUKAINO, MD, PhD[3] and Shin YAMADA, MD, PhD[1]
From the [1]Department of Rehabilitation Medicine, Kyorin University School of Medicine, Tokyo, [2]Department of Rehabilitation Medicine, Keio University School of Medicine, Tokyo, and [3]Department 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.**

LAY ABSTRACT

This comparative single case report study aims to explore the potential application of ChatGPT, a novel artificial intelligence language model, in the clinical practice of rehabilitation medicine. ChatGPT-4 was used to generate rehabilitation prescriptions and International Classification of Functioning, Disability and Health codes for a representative stroke case. The outcomes were evaluated by 2 licensed PMR clinicians. ChatGPT-4 rapidly formulated rehabilitation prescriptions and International Classification of Functioning, Disability and Health codes for the stroke case, and provided logical explanations for its recommendations. The majority of the rehabilitation treatment plan and codes were accurately generated, although there were a few discrepancies. These findings highlight the potential use of artificial intelligence large language models as tools in clinical rehabilitation practice.

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

J Rehabil Med 2023: jrm13373.

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

Published by Medical Journals Sweden, on behalf of the Foundation for Rehabilitation Information. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (https://creativecommons.org/licenses/by-nc/4.0/)

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM

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-pat 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

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM

Table I. Comparison of textbook description and ChatGPT4.0 answer

Standard textbook answer	ChatGPT4.0 answer
Medical problems:	*Medical problems:*
1. Cerebral infarction	1. Cerebral infarction
2. Hypertension	2. Hypertension
	3. Diabetes mellitus
Medical management plan:	*Medical management plan:*
1. Diagnosis of stroke type to rule out exacerbating pathogenesis	1. Continue antiplatelet treatment
2. Antiplatelet treatment	2. Blood pressure management
3. To find out the origin of thrombus and start the specific treatment	3. Blood glucose management
4. Better control for comorbidities	4. Medication review and management
	5. Routine follow-up appointments
	6. Education and counselling
	7. Multidisciplinary team approach
Rehabilitation problems:	*Rehabilitation problems:*
1. Left hemiparesis	1. Left hemiparesis (upper limb)
2. ADL difficulty	2. Impaired motor function in the left hand
3. Decreased endurance in using paretic hand	3. Mildly impaired sensory function
4. To return to work	4. Challenges in performing work-related tasks with the left hand, such as object transfer and document handling
	5. The need to adapt to potential changes in daily and work-related activities
Rehabilitation management plans:	*Rehabilitation management plans:*
1. Occupational therapy: assessment of upper limb function, instrumental ADL training and vocational rehabilitation.	1. Physical therapy: left hand motor functional trainings.
2. Physical therapy: functional training for a range of motion, dexterity, reduction of synergy pattern, endurance, bimanual movement, instruction of self-training, and applied gait training outside.	2. Occupational therapy: work-related task analysis, adaptive equipment training, energy conservation strategies, and home safety and ADL assessments.
3. Multi-disciplinary team approach to return to work	3. Speech-language therapy: if speech, language, or cognitive deficits are identified, targeted therapy and communication strategies may be provided.
	4. Patient education: about their condition, its management, and potential complications.
	5. Goal setting: establish realistic, achievable rehabilitation goals in collaboration with the patient.
	6. Regular progress monitoring and adjustments.
	7. Multidisciplinary team approach: to create and implement a comprehensive care plan addressing the patient's medical and rehabilitation needs.
Rehabilitation goals:	*Rehabilitation goals:*
1. To achieve independence in his ADL and gait outside.	1. Improve left hand motor function
2. Return to work, possibly with a slight modification of work style.	2. Improve left hand sensory function
3. Short-term recovery in sensorimotor impairment could be expected. Two weeks will be enough for him to get back to work.	3. Restore independence in activities of daily living (ADL)
	4. Facilitate a safe and timely return to work
	5. Promote patient self-management
	6. Optimize quality of life

ADL: activities of daily living.

the output of the LLM, which focused primarily on expert judgement to optimize rehabilitative intervention. Specifically, ChatGPT demonstrated weakness in determining the chronicity of the disease and selecting 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

Table II. ChatGPT-4-generated International Classification of Functioning, Disability and Health (ICF) code based on this stroke case

ChatGPT-4

Category	ICF Code	Information for the patient
Body Functions	b760	Control of voluntary movement, left hand
	b730	Muscle power functions, left hand
	b140	Memory functions (if cognitive issues arise)
Body Structures	s730	Structure of upper extremity, left hand
Activities and Participation	d440	Fine hand use, including object manipulation
	d450	Walking and moving independently
	d850	Remunerative employment (office work)
	d760	Family relationships
Environmental Factors	e310	Support from immediate family (wife and daughter)
	e580	Health services, systems, and policies
Personal Factors (not coded in ICF)	N/A	Age: 62 years
	N/A	Occupation: office worker at a bank
	N/A	Medical history: hypertension and diabetes mellitus

N/A: not available.

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.

The authors have no conflicts of interest to declare.

REFERENCES

1. OpenAI. OpenAI homepage. 2023. Available from: https://openai.com/product/gpt-4
2. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023; 388: 1233–1239. DOI: 10.1056/NEJMsr2214184.
3. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature 2023; 616: 259–265. DOI: 10.1038/s41586-023-05881-4.
4. Patel SB, Lam K. ChatGPT: the future of discharge summaries? Lancet Digit Health 2023; 5: e107–e108. DOI: 10.1016/S2589-7500(23)00021-3.
5. Zhou Z, Wang X, Li X, Liao L. Is ChatGPT an evidence-based doctor? Eur Urol 2023; 10.1016/j.eururo.2023.03.037. DOI: 10.1016/j.eururo.2023.03.037.
6. Grunebaum A, Chervenak J, Pollet SL, Katz A, Chervenak FA. The exciting potential for ChatGPT in obstetrics and gynecology. Am J Obstet Gynecol 2023; 10.1016/j.ajog.2023.03.009. DOI: 10.1016/j.ajog.2023.03.009.
7. Yamada S. Tips in assessment of rehabilitation learn from 100 cases; 2013.
8. Tsuji T, Liu M, Sonoda S, Domen K, Chino N. The stroke impairment assessment set: its internal consistency and predictive validity. Arch Phys Med Rehabil 2000; 81: 863–868. DOI: 10.1053/apmr.2000.6275.
9. Mukaino M, Prodinger B, Yamada S, Senju Y, Izumi SI, Sonoda S, et al. Supporting the clinical use of the ICF in Japan – development of the Japanese version of the simple, intuitive descriptions for the ICF Generic-30 set, its operationalization through a rating reference guide, and interrater reliability study. BMC Health Serv Res 2020; 20: 66. DOI: 10.1186/s12913-020-4911-6.
10. Senju Y, Mukaino M, Prodinger B, Selb M, Okouchi Y, Mizutani K, et al. Development of a clinical tool for rating the body function categories of the ICF generic-30/rehabilitation set in Japanese rehabilitation practice and examination of its interrater reliability. BMC Med Res Methodol 2021; 21: 121. DOI: 10.1186/s12874-021-01302-0.