

Fine-Tuning the Llama2 Large Language Model Using Books on the Diagnosis and Treatment of Musculoskeletal System in Physical Therapy

Jun-hee Kim, PT, Ph.D

Department of Physical Therapy, College of Software and Digital Healthcare Convergence, Yonsei University, Wonju, South Korea

Background In physical therapy, accurate information on musculoskeletal diagnosis and treatment is essential for patient care. Generative language models use machine learning to perform tasks like text generation and question answering, mimicking human language understanding.

Purpose This study aimed to fine-tune the Llama2 language model using text data from books on the diagnosis and treatment of the musculoskeletal system in physical therapy, and to compare it to the base model to evaluate its usability in medical fields.

Study design Technical evaluation study

Methods The Llama2-13B Chat model was fine-tuned using text from books on musculoskeletal diagnosis and treatment in physical therapy and trained on 1.2 million tokens. Responses were compared to the base model based on relevance and specialization to clinical questions.

Results Compared to the base model, the fine-tuned model consistently generated answers specific to the musculoskeletal system diagnosis and treatment, demonstrating improved understanding of the specialized domain.

Conclusions The model fine-tuned for musculoskeletal diagnosis and treatment books provided more detailed information related to musculoskeletal topics, and the use of this fine-tuned model could be helpful in medical education and the acquisition of specialized knowledge.

Key words Generative language model; Llama; Musculoskeletal system; Natural language processing; Physical therapy.

**J Musculoskelet
Sci Technol**

2024; 8(2): 65-73

Published Online

Dec 31, 2024

pISSN 2635-8573

eISSN 2635-8581



Article History

Received 5 Aug 2024

Revised 27 Aug 2024

(1st)

Accepted 30 Aug 2024

CONTACT

move@yonsei.ac.kr

Jun-hee Kim,

Department of Physical
Therapy, College of
Software and Digital
Healthcare Convergence,
Yonsei University,
Wonju, South Korea

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

In November 2022, OpenAI launched the ChatGPT service based on a generative language model (GLM) called GPT.¹ This is an important milestone in the development of artificial intelligence (AI) technology, which has brought about major changes in natural language processing and computer science.^{1,2} GLM-based services such as ChatGPT use machine learning algorithms to mimic and generate the language that people understand and use.^{1,3} These GLM-based services can perform a variety of tasks, including text generation, question answering, and sentence completion.^{1,3}

GLMs, such as OpenAI's GPT, Google's BERT, and Meta's Llama, are currently being actively developed and have influenced many fields such as economy, industry, and medicine.⁴⁻⁶ It is used in information search, customer service, and education, and its impact continues to expand.^{7,8} These GLMs can perform a variety of tasks, including generating text, answering questions, and completing sentences.⁹ The diversity and efficiency of tasks these models can perform depend largely on the amount of information that the models can process and generate based on the data used for training.¹⁰

However, GLMs are not perfect and still need to be

improved. GLMs in service currently cover general knowledge but lack knowledge of specialized domains.^{11,12} This means that the model may have difficulty answering specialized questions accurately.^{11,12} Additionally, these models often fail to provide complete and accurate answers to users' questions.^{13,14} There is a representative problem called 'hallucination', which refers to a phenomenon in which the model generates information that is not in the training data or provides incorrect information.^{15,16} These issues reduce model reliability and can provide incorrect information to users.^{15,16}

To solve these problems, there is a need for GLMs tuned to specialized domains. Especially in the medical field, since accurate and reliable information is important, the need for GLMs that include medical information is becoming more and more prominent.^{4,12,17,18} These models should be able to provide medical knowledge to people other than medical experts.^{4,19} These models can also play an important role in efficiently supporting the training of healthcare professionals.^{4,20,21} In particular, it can help students in medical professions understand complex medical terms, basis

for medical guidance, and determine appropriate treatment plans for various patient situations.²⁰⁻²²

The purpose of this study was to provide empirical evidence on the benefits of domain-specific fine-tuning of language models, demonstrating that such models not only improve accuracy in the specialized domain of musculoskeletal diagnosis and treatment but also highlight the potential for these models to serve as educational tools and decision-support systems in clinical settings. Therefore, this study compared the answers generated when questions related to the specialized musculoskeletal domain of physical therapy were applied to fine-tuned and untuned models, respectively, to evaluate and confirm the possibility of whether language models fine-tuned to the professional domain can correctly answer professional knowledge.

METHODS

Data collection and pre-processing

The process of collecting, processing, and fine-tuning the data to the language model is shown in Figure 1. For this

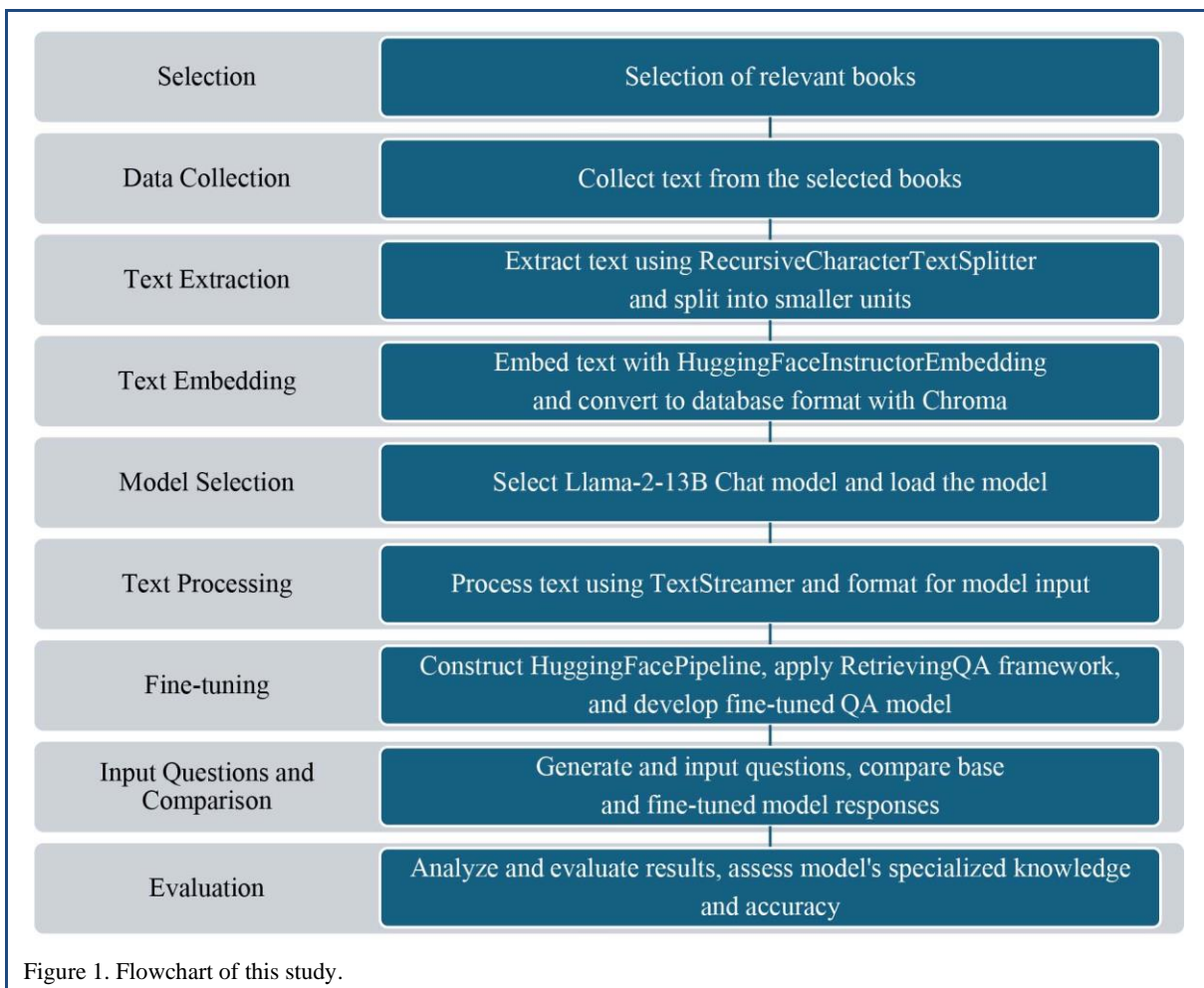


Figure 1. Flowchart of this study.

study, books related to the diagnosis and treatment of musculoskeletal disorders in the field of physical therapy were selected. First, the book ‘Functional Anatomy’, which is considered the most basic area in diagnosing or treating musculoskeletal disorders, was selected.²³ And a book called ‘Kinesiology’, which describes the movement of the musculoskeletal system in the human body, was selected.²⁴ In addition, ‘Diagnosis and Treatment of Movement System Impairment Syndromes’, ‘Muscle Imbalance’, and ‘Kinetic Control’ books, in which diagnostic theories using movement or musculoskeletal systems were written in relation to musculoskeletal disorders, were selected.^{25–27} Finally, a book called “Therapeutic Exercise,” which describes exercise therapy, one of the options for treating musculoskeletal disorders, was selected.²⁸ Document text was extracted using the RecursiveCharacterTextSplitter, and large chapters of the document were divided into small units to increase processing efficiency. The extracted text was embedded by HuggingFaceInstructorEmbedding. The embedded document was then converted into a database format using Chroma to enable saving and retrieval.

Model selection and fine-tuning

The Llama-2-13B Chat model was selected and fine-tuned using a dataset comprising approximately 1.2 million tokens extracted from the chosen textbooks. The fine-tuning process involved training for 1 epoch, with a learning rate of $2e-4$ and a batch size of 4. These hyperparameters were optimized based on preliminary experiments to ensure the model’s performance in specialized tasks. This model is specifically designed for interactive AI and question-answering scenarios. The Llama-2-13B model was loaded using the AutoGPTQForCausalLM library, and the settings required for model fine-tuning were performed. TextStreamer, a component that processes input text, was used to process text data effectively. This step was important in preparing and formatting the input text before feeding it to the model. The process of creating a response using the Llama-2-13B

model by constructing the HuggingFacePipeline has been simplified. This pipeline has been applied to generate consistent and contextually relevant responses by managing various aspects such as tokenization and decoding during the creation process. Finally, the RetrievingQA framework was used to develop a QA model with added search functionality. Through this process, a Fine-tuned QA model was established, which is expected to provide more accurate and contextual responses to various musculoskeletal disease diagnosis and treatment-related questions.

Input questions and base comparison of models

Questions were selected based on key topics and frequent inquiries encountered in physical therapy practice, as identified from the selected books. The selection process involved consultations with experts in the field to ensure that the questions reflect real-world clinical scenarios. This approach aimed to test the model’s ability to respond accurately to common and critical issues within the musculoskeletal domain. The questions created were entered into the base and fine-tuned models, respectively. The questions created and entered according to the selected books were as shown in Table 1.

RESULTS

The answers generated for each question of the base and the fine-tuned models are presented in Table 2. The base model often failed to understand or provide relevant information because the questions were unclear or key terms were not recognized. In contrast, the fine-tuned model consistently produced answers specific to the musculoskeletal context, demonstrating improved understanding of specialized topics. In particular, the fine-tuned model did not show any lack of information for any question and showed improved performance in processing musculoskeletal inquiries based on fine-tuned text content. Regarding the serratus anterior muscle, the base model provided a general over-

Table 1. Questions related diagnosis and treatment of musculoskeletal system

List of questions
1) Tell me about the action of serratus anterior muscle
2) Tell me about the scapulo-humeral rhythm
3) Tell me about the scapular downward rotation movement impairment syndrome
4) Tell me about the upper crossed syndrome
5) Tell me about the kinetic medial rotation test
6) Tell me management for glenohumeral joint after surgery

Table 2. Questions and answers from models before and after fine-tuning

Question	Tell me about the action of serratus anterior muscle
Base model	The serratus anterior muscle is a key player in shoulder movement and positioning. It is a fan-shaped muscle located on the side of the ribcage, and it helps to rotate the scapula (shoulder blade) and move the arm forward and backward. The serratus anterior muscle also helps to stabilize the shoulder joint and maintain proper alignment of the shoulder blades.
Fine-tuned model	The serratus anterior muscle produces upward rotation of the scapula during arm elevation, which exceeds the downward rotation produced by the active middle deltoid and supraspinatus. The serratus anterior muscle also produces a subtle posterior tilting and external rotation torque to the upwardly rotating scapula. The secondary actions of the serratus anterior muscle become evident when there is paralysis of the muscle, resulting in a downwardly rotated position of the scapula, along with slight anterior tilting and internal rotation.
Question	Tell me about the scapulo-humeral rhythm
Base model	Based on the provided context, I cannot provide information about the scapulo-humeral rhythm as there is no relevant information given. The term “scapulo-humeral” does not appear in the context provided, and therefore, I am unable to provide any information or answer related to it.
Fine-tuned model	The scapulo-humeral rhythm refers to the synchronized movement of the scapula and humerus during shoulder flexion. According to Inman and Saunders, there is a “setting phase” during the first 60 degrees of shoulder flexion and 30 degrees of abduction, where the movement of the scapula is highly variable. After the setting phase, the humerus and scapula move in a constant ratio, with 2 degrees of glenohumeral motion for every 1 degree of scapulothoracic motion, resulting in 120 degrees of glenohumeral joint motion and 60 degrees of scapular motion at the completion of shoulder flexion. However, more recent studies have reported some variability in the exact timing of that motion.
Question	Tell me about the scapular downward rotation movement impairment syndrome
Base model	I don't know about a “scapular downward rotation movement impairment syndrome.” Based on my training data, there is no widely recognized condition or syndrome with this name. It is possible that you may be referring to a specific movement pattern or dysfunction, but without more information, I cannot provide further clarification. If you have any additional details or context, I would be happy to help as best I can.
Fine-tuned model	The primary movement impairment in this syndrome is insufficient scapular upward rotation. Muscle impairments include dominance, shortness, or stiffness of the downward rotator muscles such as rhomboids, levator scapulae, latissimus, pectoralis minor and major, and insufficient activity of the upward rotator muscles like serratus anterior and trapezius. This syndrome can cause symptoms such as pain, problems, and associated diagnoses. Treatment options are available, including confirming tests, movement pattern impairments, and relative flexibility and stiffness impairments.
Question	Tell me about the upper crossed syndrome
Base model	I don't know about the “upper crossed syndrome”. I'm not familiar with that term or concept. It's possible it's a specific condition or injury, but without more information, I can't provide any further insight or explanation. If you have any additional details or context, I may be able to help better.
Fine-tuned model	The upper crossed syndrome is also known as the proximal or shoulder girdle crossed syndrome. It is characterized by tightness of the upper trapezius and levator scapula on the dorsal side, which crosses with tightness of the pectoralis major and minor. Additionally, there is weakness of the deep cervical flexors ventrally, which crosses with weakness of the abdominal muscles. This syndrome is described as having a specific pattern of muscle weakness and tightness that crosses between the dorsal and ventral sides of the body.

Table 2. Continued

Question	Tell me about the kinetic medial rotation test
Base model	I don't know about a "kinetic medial rotation test." Based on the phrase "kinetic medial," I might guess that it has something to do with the movement or rotation of the knee joint, but without more information, I can't provide any specific details or information about such a test.
Fine-tuned model	The kinetic medial rotation test is a part of the assessment and rating of low threshold recruitment efficiency in chapter 8 of the shoulder girdle. It is also known as KMR T. The test involves evaluating the patient's ability to perform medial rotation movements with their arm. Specifically, it assesses for downward rotation, elevation, and anterior translation of the glenohumeral joint. The results of this test can help determine the efficiency of the patient's low threshold recruitment of the rotator cuff muscles.
Question	Tell me management for glenohumeral joint after surgery
Base model	<p>I can provide information on post-surgical management for a Glenohumeral joint based on current medical knowledge and research. However, please note that individualized advice should always be sought from qualified healthcare professionals who have access to your specific medical history and current condition. After surgery to repair or replace the Glenohumeral joint (e.g., rotator cuff repair, shoulder replacement), the following general principles may apply:</p> <ol style="list-style-type: none"> 1. Immobilization: The arm is usually immobilized in a sling or brace for several weeks to allow the joint to heal properly. This helps prevent further injury, reduces pain, and promotes proper alignment during the healing process. 2. Pain management: Your healthcare provider will likely prescribe medication to manage any discomfort or pain you experience during the recovery period. It is essential to follow their instructions carefully and not exceed the recommended dosage. 3. Physical therapy: Once the initial inflammation has subsided, physical therapy exercises are crucial to regaining strength, flexibility, and range of motion in the affected shoulder. A trained physical therapist will design a personalized program tailored to your needs and progress. 4. Wound care: If you had an open procedure, your surgeon will provide guidelines for wound care and dressing changes to promote healing and minimize the risk of complications. 5. Follow-up appointments: Schedule regular follow-up appointments with your surgeon to monitor your progress, address any concerns or issues, and adjust your treatment plan as needed. 6. Lifestyle modifications: Depending on the severity of your condition and the type of surgery performed, you may need to modify certain activities or behaviors to avoid exacerbating the injury or compromising the healing process. For example, you may need to avoid heavy lifting, bending, or overhead movements for several months. 7. Monitoring for complications: Your healthcare team will keep a close eye on you for potential complications such as infection, nerve damage, or blood clots. Promptly report any unusual symptoms or redness to your healthcare provider. 8. Gradual return to activities: As you recover, gradually resume normal activities, starting with low-impact exercises like swimming or cycling. Avoid heavy lif.
Fine-tuned model	After surgery, postoperative management is crucial for optimal outcomes. The text highlights the importance of early mobilization and exercises to prevent stiffness, strengthen the muscles, and improve range of motion. Immobilization for too long can lead to contracture, loss of motion, and weakness. The text also suggests that young, active patients who have sustained an acute, traumatic, anterior dislocation may benefit from surgical stabilization, as there is a high risk of recurrence after nonoperative management. However, it does not provide specific information on the management after surgery.

view of its role in shoulder movement and stabilization. In contrast, the fine-tuned model provided a more detailed description, emphasizing specific movements during arm lifting and outcomes related to muscle paralysis. For the scapulo-humeral rhythm, the base model was insufficient to

provide information due to lack of context, but the fine-tuned model provided a comprehensive definition, including details on the "set-up phase" and the constant rate of humerus and scapula movement during shoulder flexion. Regarding scapular downward rotation syndrome, the base

model was unfamiliar and additional information was requested. In contrast, the fine-tuned model not only defined the major movement disorders but also outlined the associated muscle dysfunction, symptoms, and available treatment options. In the case of upper crossed syndrome, the base model acknowledged ignorance of the term and suggested that more context was needed. Conversely, the fine-tuned model identified specific muscle tension and weakness patterns associated with the syndrome, providing a detailed description. Regarding the kinetic medial rotation test, the base model lacked information about the test and provided speculative guesses based on the phrases provided. In contrast, the fine-tuned model describes the components of the test in more detail and explains how the results contribute to assessing rotator cuff muscles. Regarding postoperative management of the glenohumeral joint, the base model provides general postoperative management principles, emphasizing the need for individualized advice. On the other hand, fine-tuned models represent a more nuanced understanding, emphasizing the importance of early mobilization, exercise, and potentially surgical stabilization for specific cases.

DISCUSSION

The base model generally provided accurate information, but it lacked detailed information about certain terms or concepts. However, general information was provided on postoperative management of the glenohumeral joint. The fine-tuned model provided more detailed and specialized information on specific terms and concepts. It may provide more contextual information, but there may be differences in the quality of the response. It was suggested that fine-tuned models tend to provide more detailed and context-specific answers, especially in areas related to specific muscle activities, syndromes, and tests. For serratus anterior, scapula-humeral rhythm, scapular inferior rotation syndrome, superior crossed syndrome, and kinetic medial rotation test, fine-tuned model provided in-depth insights, whereas base model often provided uninformed or speculative guesses. Overall, the fine-tuned model demonstrated superior performance, demonstrating its ability to provide more detailed and contextual information across a variety of musculoskeletal topics compared to the base model, which provides general information.

These results were consistent with various studies showing that model tuning enables effective and expert responses to questions by equipping models with specific knowledge of specialized domains.^{6,11,14} Several studies have already attempted to fine-tune large-scale language models or build

medical language models in the medical field.^{19,29,30} Yang et al.²⁹ introduced a large-scale clinical language model called GatorTron, developed with more than 9 billion words of text, and the generative model using this language model led to the improvement of medical question answers.²⁹ Additionally, Lu et al.³⁰ introduced the ClinicalT5 model and said that this model can be applied to various tasks in the medical field because it can understand and process medical terminology, context, and special language structures.³⁰ Nova¹⁹ suggested that using a GLM learned from text data such as medical records or from medical personnel's voice data could simplify and provide medical information that was difficult for patients to understand.¹⁹

Development of a language model in a specific field such as the above or improvement of the model through fine-tuning is also used for education in that field.^{8,11,21} Karabacak et al.²¹ suggested that AI and GLMs will provide significant opportunities to improve medical education through realistic simulations, digital patients, personalized feedback, assessment methods, and removal of language barriers.²¹ Unlike previous studies that used medical records, this study used text data from books containing knowledge about the musculoskeletal system in the field of physical therapy to fine-tune the model, but it is expected to be helpful in education as well. This is an important advancement that could provide a new dimension in expanding medical education and expertise. This approach is expected to contain specific knowledge about the field and provide more detailed information, allowing medical professionals and educators to provide learners more effectively with an in-depth understanding of specific topics. The expertise gained through fine-tuning the model not only increases its applicability in clinical situations, but also provides learners with rich experience of real-world situations in the field. This realistic training environment can improve the practical skills of healthcare professionals and increase the effectiveness of research-based training. Additionally, the utilization of GLM can overcome language barriers and improve the quality of global medical education. Able to understand and handle multiple languages and cultures, GLM can foster international educational collaboration and improve information exchange between healthcare professionals.

However, several problems are being raised in content creation using AI and GLM. Among them are content quality, bias, and ethical and legal issues.²¹ In addition, it is required to develop guidelines and policies in evaluating the accuracy of AI-generated content.^{20,21} To this end, it is said that reliability and reliability can be increased by sharing information on data and evaluation methods used in educa-

tion.²¹ Continuous research and academic cooperation are needed to realize the maximum potential of AI and GLM in medical education and to respond to potential risks and challenges at the same time.²¹ It is said that through the cooperation of medical experts, these technologies will be integrated in an effective and responsible way, leading to improved medical care between learning experiences and patients.^{20,21}

There are several limitations to this study. Firstly, the study focused exclusively on text data from books related to musculoskeletal diagnosis and treatment, which may limit the model's exposure to a wide range of medical conditions and patient cases. It could result in the model providing less generalizable or comprehensive answers outside the specific domain of physical therapy. Secondly, this study did not use quantitative metrics such as accuracy and precision to evaluate the performance of the fine-tuned model. While the qualitative analysis indicates that the fine-tuned model offers more contextually relevant and specialized answers, the lack of statistical evaluation limits the ability to objectively measure and compare its performance against the base model. Future studies should include these quantitative assessments to provide a more rigorous and objective evaluation of model improvements. Additionally, the selected textbooks may introduce bias based on their specific medical approaches, potentially affecting the model's responses. Lastly, relying solely on textbook information for fine-tuning may not capture the most up-to-date or empirically validated knowledge, potentially limiting the model's effectiveness in reflecting current clinical practices. To address these limitations, future studies should include a more diverse dataset that encompasses a broader range of medical conditions and patient cases to improve the model's generalizability. Furthermore, in future studies, it will be essential to validate the performance using quantitative metrics to evaluate and compare the performance of fine-tuned and underlying models. Additionally, future research should explore the inclusion of real-world clinical data and recent research articles in the fine-tuning process to ensure that the model remains current and reflective of empirical knowledge. Finally, investigating the model's performance on non-specialized and broader medical questions will provide insights into its versatility and potential applications beyond the initial scope of this study.

CONCLUSION

The model fine-tuned with books on musculoskeletal diagnosis and treatment showed superior performance in providing various information related to the topic compared

to the basic model. This demonstrates the effectiveness of model tuning for a specific knowledge domain and suggests that model enhancement using textual data on the musculoskeletal system may help in medical education and expertise acquisition. However, there will be a need for guidelines and policies to evaluate reliable AI-generated content, and there will be a need to increase reliability by sharing information on data and evaluation methods used in education.

Key Points

Question Can a fine-tuned language model provide accurate information on musculoskeletal diagnosis and treatment?

Findings The fine-tuned model demonstrated superior performance in providing detailed and specialized information.

Meaning Model fine-tuning can enhance the applicability of language models in specialized medical education.

Article information

Conflict of Interest Disclosures: The author certifies that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria, educational grants, participation in speakers' bureaus, membership, employment, consultancies, stock ownership, or other equity interest, expert testimony, or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge, or beliefs) in the subject matter or materials discussed in this manuscript. The author reports the following details of affiliation or involvement in an organization or entity with a financial or non-financial interest in the subject matter or materials discussed in this manuscript.

Funding/Support: I would like to disclose that this research was conducted without the receipt of any external funding. No specific grants, financial support, or sponsorships were utilized in the design, data collection, analysis, interpretation, or composition of this study. The author acknowledges that the study is entirely self-funded, and the absence of external funding has not influenced the independence or objectivity of our research findings.

Acknowledgment: None.

Ethic Approval: None.

Author contributions

Conceptualization: JH Kim.

Data acquisition: JH Kim.
 Design of the work: JH Kim.
 Data analysis: JH Kim.
 Project administration: JH Kim.
 Interpretation of data: JH Kim.
 Writing – original draft: JH Kim.
 Funding acquisition: JH Kim.
 Writing–review&editing: JH Kim.

REFERENCES

- Roumeliotis KI, Tselikas ND. ChatGPT and Open-AI models: a preliminary review. *Futur Internet*. 2023; 15(6):192.
- Zhang M, Li J. A commentary of GPT-3 in MIT technology review 2021. *Fundam Res*. 2021;1(6):831-833.
- Fui-Hoon Nah F, Zheng R, Cai J, et al. Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. *J Inf Technol Case Appl Res*. 2023;25(3): 277-304.
- Thirunavukarasu AJ, Ting DSJ, Elangovan K, et al. Large language models in medicine. *Nat Med*. 2023; 29(8):1930-1940.
- Hadi MU, Qureshi R, Shah A, et al. *Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects*. Published online 2023.
- Luk M. *Generative AI: overview, economic impact, and applications in asset management. economic impact, and applications in asset management* (September 18, 2023). Published online 2023.
- Ray PP. *ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems*. Published online 2023.
- Ilieva G, Yankova T, Klisarova-Belcheva S, et al. Effects of generative chatbots in higher education. *Inf*. 2023;14(9):492.
- Min B, Ross H, Sulem E, et al. Recent advances in natural language processing via large pre-trained language models: a survey. *ACM Comput Surv*. 2023; 56(2):1-40.
- Radford A, Narasimhan K, Salimans T, et al. *Improving language understanding by generative pre-training*. Published online 2018.
- Baidoo-Anu D, Ansah LO. Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*. 2023;7(1):52-62.
- Zhang Y, Pei H, Zhen S, et al. Chat generative pre-trained transformer (ChatGPT) usage in healthcare. *Gastroenterol Endosc*. 2023;1(3):139-143.
- Sobieszek A, Price T. Playing games with AIs: the limits of GPT-3 and similar large language models. *Minds Mach*. 2022;32(2):341-364.
- Zhang C, Bauer S, Bennett P, et al. *Understanding causality with large language models: Feasibility and opportunities*. arXiv preprint arXiv:230405524. Published online 2023.
- Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*. 2023;15(2).
- Ji Z, Lee N, Frieske R, et al. Survey of hallucination in natural language generation. *ACM Comput Surv*. 2023; 55(12):1-38.
- Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit Med*. 2023;6(1):120.
- Yu P, Xu H, Hu X, et al. Leveraging generative AI and large language models: a comprehensive roadmap for healthcare integration. In: *Healthcare*. Vol 11. MDPI; 2023:2776.
- Nova K. Generative AI in healthcare: advancements in electronic health records, facilitating medical languages, and personalized patient care. *J Adv Anal Healthc Manag*. 2023;7(1):115-131.
- Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. *JMIR Med Educ*. 2023;9(1):e46885.
- Karabacak M, Ozkara BB, Margetis K, et al. The advent of generative language models in medical education. *JMIR Med Educ*. 2023;9:e48163.
- Javaid M, Haleem A, Singh RP. ChatGPT for healthcare services: an emerging stage for an innovative perspective. *BenchCouncil Trans Benchm Stand Eval*. 2023; 3(1):100105.
- Cael C. *Functional Anatomy: musculoskeletal anatomy, kinesiology, and palpation for manual therapists, enhanced edition: musculoskeletal anatomy, kinesiology, and palpation for manual therapists, enhanced edition*. Burlington, Massachusetts: Jones & Bartlett Learning; 2020.
- Neumann DA, Kelly ER, Kiefer CL, et al. *Kinesiology of the musculoskeletal system: foundations for rehabilitation*. London, England: Elsevier; 2017.
- Sahrmann S. *Diagnosis and treatment of movement impairment syndromes*. London, England: Elsevier Health Sciences; 2001.

26. Page P, Frank C, Lardner R. *Assessment and treatment of muscle imbalance: the janda approach*. Champaign, Illinois: Human Kinetics; 2009.
27. Comerford M, Mottram S. *Kinetic control: the management of uncontrolled movement*. Sydney, New South Wales: Elsevier Australia; 2012.
28. Kisner C, Colby LA, Borstad J. *Therapeutic exercise: foundations and techniques*. Philadelphia, Pennsylvania: F.A. Davis Company; 2017.
29. Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records. *NPJ Digit Med*. 2022;5(1):194.
30. Lu Q, Dou D, Nguyen T. ClinicalT5: a generative language model for clinical text. In: *Findings of the Association for Computational Linguistics: EMNLP*; 2022:5436-5443.