frontiers | Frontiers in Public Health

TYPE Technology and Code
PUBLISHED 24 May 2024
DOI 10.3389/fpubh.2024.1364660

Check for updates

OPEN ACCESS

EDITED BY
Mousa Al-kfairy,
Zayed University, United Arab Emirates

REVIEWED BY
Shailesh Tripathi,
University of Applied Sciences Upper Austria, Austria
Fady Alnajjar,
United Arab Emirates University, United Arab Emirates

*CORRESPONDENCE
Summaiya Zareen Shaikh
✉ drzareensummaiya@gmail.com

RECEIVED 02 January 2024
ACCEPTED 10 May 2024
PUBLISHED 24 May 2024

CITATION
Naqvi WM, Shaikh SZ and Mishra GV (2024) Large language models in physical therapy: time to adapt and adept.
Front. Public Health 12:1364660.
doi: 10.3389/fpubh.2024.1364660

COPYRIGHT
© 2024 Naqvi, Shaikh and Mishra. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Large language models in physical therapy: time to adapt and adept

Waqar M. Naqvi[1,2,3], Summaiya Zareen Shaikh[4]* and Gaurav V. Mishra[5]

[1]Department of Interdisciplinary Sciences, Datta Meghe Institute of Higher Education and Research, Wardha, India, [2]Department of Physiotherapy, College of Health Sciences, Gulf Medical University, Ajman, United Arab Emirates, [3]NKP Salve Institute of Medical Sciences and Research Center, Nagpur, India, [4]Department of Neuro-Physiotherapy, The SIA College of Health Sciences, College of Physiotherapy, Thane, India, [5]Department of Radiodiagnosis, Datta Meghe Institute of Higher Education and Research, Wardha, India

Healthcare is experiencing a transformative phase, with artificial intelligence (AI) and machine learning (ML). Physical therapists (PTs) stand on the brink of a paradigm shift in education, practice, and research. Rather than visualizing AI as a threat, it presents an opportunity to revolutionize. This paper examines how large language models (LLMs), such as ChatGPT and BioMedLM, driven by deep ML can offer human-like performance but face challenges in accuracy due to vast data in PT and rehabilitation practice. PTs can benefit by developing and training an LLM specifically for streamlining administrative tasks, connecting globally, and customizing treatments using LLMs. However, human touch and creativity remain invaluable. This paper urges PTs to engage in learning and shaping AI models by highlighting the need for ethical use and human supervision to address potential biases. Embracing AI as a contributor, and not just a user, is crucial by integrating AI, fostering collaboration for a future in which AI enriches the PT field provided data accuracy, and the challenges associated with feeding the AI model are sensitively addressed.

KEYWORDS

artificial intelligence, BioMedLM, evidence-based practice, large language models, physical therapy, physical therapy education, rehabilitation

Introduction

The healthcare landscape is undergoing a seismic shift driven by the relentless advancement of artificial intelligence (AI) and machine learning (ML) (1). Physiotherapists, like all healthcare professionals, stand at the precipice of this transformative era in professional education, clinical practice, and research. While some may view AI as a threat, it presents a unique opportunity for physiotherapists to elevate our practice and revolutionize patient care (2). Computer programs that indirectly enable humans to seem intelligent by doing tasks, providing options, and executing decisions on behalf of humans are called artificial intelligent systems (3) (Figure 1). Chat Generative Pre-trained Transformer (Chat-GPT) has managed to achieve human-like performance using deep ML and deep neural networks (4). A large amount of complex data is collected in the medicinal field, which often requires analysis to break down complex information into simple interpretations. Although OpenAI's ChatGPT is one of the biggest LLM marching ahead in AI dataset development and accuracy, due to the presence of vast data, it gets difficult to have accurate results on ChatGPT (5). Stanford University is among the pioneers in developing BioMed-LLM focused on datasets created from the PubMed database. There are also other LLMs that are more focused on medical databases

10.3389/fpubh.2024.1364660

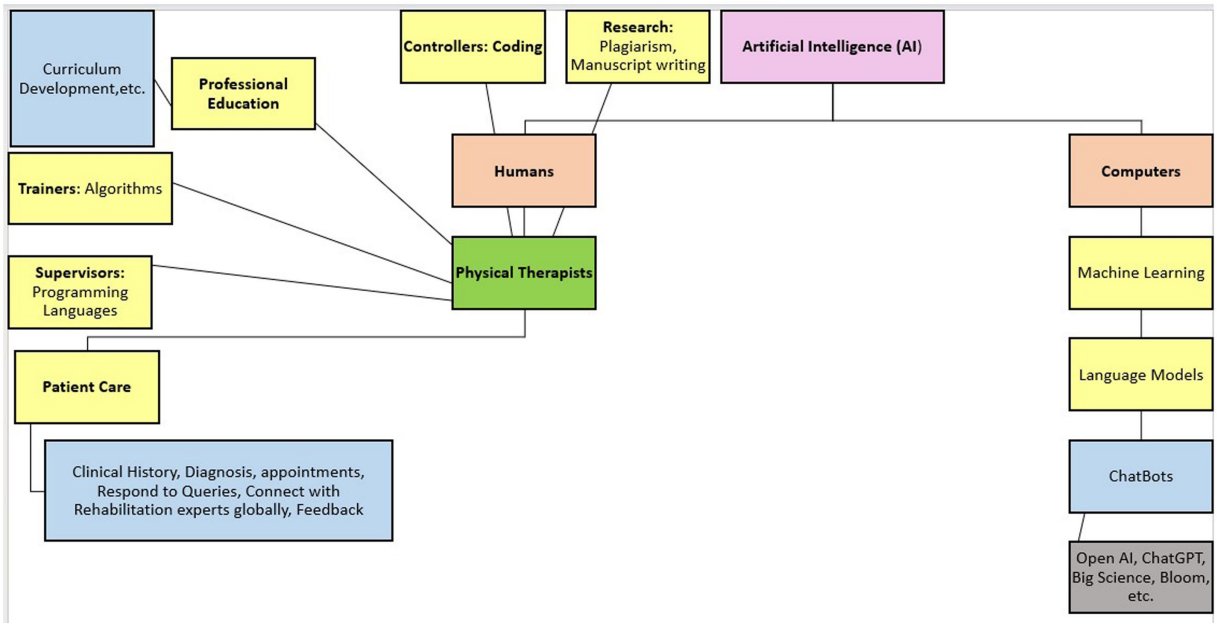

FIGURE 1
Overview of utility of artificial intelligence in physical therapy.

and datasets, such as Microsoft's Bio-GPT and Google's Med-PaLM (6). These datasets are in the stage of fine-tuning to provide more accurate and proficient results. The LLM programs enable the clinician, medical researcher, and educator to make scientifically informed decisions with respect to symptoms, assessment, diagnosis, and further plans of action (7).

Generative AI is an umbrella term used when AI is used to generate ideas, content, images, etc. It is defined as a branch of ML that translates innovative content in the form of text, images, and audio-visual into computer codes (8). Generative AI ranges of standardized assessments in medical education not only underscore their capabilities but also prompt a reconsideration of our existing evaluation methods (9). Generative AI delves into the burgeoning field of generative AI and its applications within the medical domain. Mohammed H. M. Amin et al. in their study, explore the utility of generative adversarial networks (GANs) for brain tumor segmentation, demonstrating its potential in enhancing medical image analysis tasks. Kumar, Sharma, and Tyagi provide an extensive overview of the state-of-the-art and recent developments in GANs, shedding light on their diverse applications and promising advancements. Thakkar, Anand, and Jaiswal discuss the burgeoning applications of GANs in healthcare, highlighting their significance in various medical contexts. Fasakin et al. conducted a comprehensive survey on the utilization of GANs in healthcare, presenting an analysis of trends, techniques, and applications, offering valuable insights into the field's current landscape. Finally, Paul, et al. focus on medical image synthesis using GANs, showcasing their efficacy in generating synthetic medical images, thereby facilitating research, training, and diagnostic tasks in the medical domain. These articles collectively underscore the growing significance of generative AI in revolutionizing medical imaging, diagnostics, and healthcare delivery.

There are other types of AI that group, assign, choose, and decide, and can be semi to fully automated as well. Generative AI systems have been developed for content creation, idea generation, image creation, and audio creation. Overall, generative AI terminology has been utilized depending on the type of usage by consumers. There are, however, certain support tools or systems developed which are related to *language*. The way coding programs are used for the development of application-based technology, similarly, language coding is used as a support tool by researchers and educators for creating academic content for AI chatbot systems (10). The LLM would serve as a tool to assist physical therapists (PTs) in effectively using the capabilities of generative AI. Our paper contributes to a general discussion among research groups worldwide, necessitating the use of Generative AI and LLMs in physical therapy and rehabilitation. To conclude, large language models (LLMs) are programs developed in order to simplify complex content by providing a comprehensive summary (11, 12). However, the question arises: Is the content useful, valid, reliable, and scientific? To answer this question: yes, it can be all of the above, provided humans feed into this system of AI, although it is full of challenges and limitations, and ensuring strict protocols to train the LLM of relevant literature databases for training and utilization. It is also essential that the LLMs be fed with unbiased, factually accurate data. Literature in the past 5 years states that the application of LLMs in the realm of healthcare, including areas such as neurology, oncology, psychiatry, and more, has shown promising results, except for studies directly pertaining to physical therapy, which are limited in the current research data (Table 1).

Ensuring data accuracy and challenges

Data accuracy stands as the cornerstone of reliable and impactful decision-making. However, as AI systems become more sophisticated, several challenges emerge in ensuring the accuracy of the data they process based on four major challenges (18, 19):

TABLE 1 Background review of literature on the current utility of AI in healthcare.

Title	Objective	Methodology	Key findings
Recognizing physiotherapy exercises using machine learning (13)	Developed a system for Physiotherapy exercise assessment using pose detection.	Uses pose detection for exercise assessment with a camera and provides feedback on posture and body angles.	96% accurate for pose detection.
Artificial intelligence and machine learning applications in musculoskeletal physiotherapy (14)	Utility of machine learning (ML) in musculoskeletal medicine.	Review on supervised and unsupervised ML with diagnostic imaging and clinical decision-making assistance.	ML can enhance physiotherapy practice through the automation of tasks involving data analysis, classification, and prediction.
Artificially Intelligent Physiotherapy (15)	Framework using AI and ML for digitalized physiotherapy practice.	Uses OpenPose Library for detecting joint angles and providing exercise suggestions based on performance.	Enables remote physiotherapy sessions with personal feedback, aiming for cost reduction.
Machine Learning Approach for Physiotherapy Assessment (16)	Enhance the accessibility of physiotherapy treatment using a Kinect sensor and inertial measurement unit (IMU) for motion tracking.	Tracks human motion using a Kinect sensor and IMU chip, with a feedback band for movement correctness.	Framework supports remote physiotherapy sessions, reducing treatment costs.
Automated Error Detection in Physiotherapy Training (17)	Develop a system for scoring and feedback in manual skills training using sensor data.	Utilizes IMU and sEMG sensor data for gesture classification, scoring, and feedback	90% accurate in movement pattern execution for manual skills training.

- Model Forgetfulness: It is a tendency of AI models to gradually lose previously acquired knowledge over time, particularly when exposed to evolving datasets, especially when it is supposed to predict or forecast something.
- User Acceptability: It is the willingness of end-users to trust and use AI-driven solutions in their decision-making, depending on the transparency of the LLMs and its features.
- Explainability: The ability of AI systems to provide interpretable explanations for their decisions and predictions, often referred to as "black boxes."
- Hallucination effect: It occurs in generative AI models where the system generates false or unrealistic outputs that resemble real data, which poses risks in applications such as image generation, text synthesis, and data augmentation.

Humans will play the role of creators, and supervisors and their responsibility would be to ensure setting up boundaries for judicious use of these models and programs. The other factor deciding the reliability of content would be the process involving data management, data filtering, and the level of training provided to these programs and users for ethical use of the system. The pros definitely outweigh cons of AI, although the cons can pose a serious risk if not taken into account. The potential benefits of AI and ML in physical therapy are vast and encompass several crucial areas (Table 2). Medical and health sciences practitioners and educators have delved into the use of language models in ophthalmology, radiology, orthopedics, dermatology, surgery, obstetrics, and gynecology for educational and clinical practice development (20–24). Wilhem et al. evaluated the performance of four language models, i.e., GPT-3.5. Turbo, Bloomz, BigScience, among others, created medical content across the above specialties and concluded that LLMs are capable of generating consistent and trustworthy, clinically safe medical advice to specific prompts (25). Studies by Mesko et al. conclude that content creation is consistent and trustworthy, and LLM-generated medical advice can be tailored to their cases based on clinical guidelines and expert opinions. However, limited studies have validated the consistency of responses to similar prompts or case scenarios to assess the robustness of LLMs and their efficacy. The steps ideally include prompting, screening by an expert, review of responses by another set of experts, fine-tuning of errors, and benchmarking them against gold standard datasets to assess user experience and feedback. In this way, LLM responses can be modulated and compared after being fed with accurate data and evidence-based clinical practice guidelines to meet patient needs (26). A recent scoping review by Ullah et al. assessed the challenges and barriers to using LLMs in diagnostic medicine in the field of pathology (27). Many language models, such as Claude, Command, and Bloomz, have been programmed for creating accurate medical advice (28) (Figure 2). Hence, recognizing AI as more than just a tool but as a genuine contributor is essential, which entails integrating AI into physical therapy to improve data accuracy and encourage collaborative efforts. It is crucial to address the challenges associated with training AI models sensitively, particularly handling patient data and upholding ethical standards. Therefore, navigating these obstacles thoughtfully, we can completely leverage AI's potential to revolutionize the practice of physical therapy (Figure 3).

The perspective discussed in this paper is how and why language learning models (LLMs) of the AI universe will be a sweeping game-changer in physical therapy in terms of professional education, clinical practice, and evidence-based practice.

Training of models by physical therapists

- *Why to train:* LLMs are large-sized programs of the AI system that have been fed into datasets and trained by infinite words, bibliography which have been imported from scientific articles, books, and other content which is available on the internet search engines (29). For example, if the LLM system is developed in such a way that only PubMed database articles, books, and other literature is needed from a particular database. This was established by Stanford University in which they developed a chatbot LLM of 2.7B in size extensively trained and imported

from the PubMed database for answering model questions and answers, which was previously known as PubMed GPT in the field of medicine website link (30). On similar lines, in nursing education, Glauberman et al. have concluded that algorithm biases must be addressed at each step of training the LLM with a continuous evaluative process, and data of many different races, conditions, and severity must be incorporated and categorized (31). AI diminishes humans' capacity to learn from our experiences; hence, these experiences must be coded, reviewed continuously, and fed, and re-fed with ever-changing factual data by human experts at multiple levels (32) *How to train:* Deep ML is a method of training the machine or programs with many layers. For instance, processing large amounts of images, videos, and written content across a particular field such as physical therapy practice, education, and research. Data management must be trained by data managers via standard and enriching meta-data according to FAIR guidelines (Findable, Accessible, Interoperable, Reusable) (33). Human interruption or supervision, is essential at every step. It has to be trained by humans efficiently in order to be used by humans, as it augments our work more efficiently. Humans will not be replaced or else the risk of automation is huge (34).

- *What to train:* Data feeding into these systems can be in several formats and types. Most commonly evaluated cases in musculoskeletal rehabilitation, such as degenerative, inflammatory, and so on, anatomy of the involved joints, specific pointers in history taking, differential diagnosis, specific investigations required, patient education, the acute and chronic line of treatment, time required to assess, diagnose, and complete the treatment and much more, can be entered into the LLM with strong, scientific evidence (35). On the other hand, the system can be linked with high-quality databases and journals, which will be qualified based on their journal metrics, type of peer review, and other such factors to enable the facilitation of high-quality content and can be proven as a verified source of information (36). Similarly, this format can be applied to various branches of physical therapy rehabilitation such as neurological, pediatrics, geriatrics, cardiovascular, pulmonary, community-based, hand, and sports (37). All of these can be categorized and classified according to the International Classification of Function (ICF) domains and processed in a uniform system by the ICF codes (38, 39).
- *Whom to train:* Not only the program developers and bio-engineers but also the PTs should start learning coding, algorithms, and types of AI, its chatbots, language models, and

TABLE 2 Advantages of large language models in clinical practice, professional education, and research.

A. Clinical practice:
• Patient assessment and screening.
• Customization of tailor-made rehabilitation protocols.
• Patient counseling and education.
• Accurate diagnosis.
B. Professional education:
• Development of learning resources for students.
• Curriculum development and enhancement.
• Quality assurance.
• Creating examination formats (e.g., short answers, quiz, and multiple choice questions)
C. Research and evidence-based practice:
• Literature review
• Research synthesis
• Plagiarism detection
• Referencing
• Integration of data into manuscript publication.

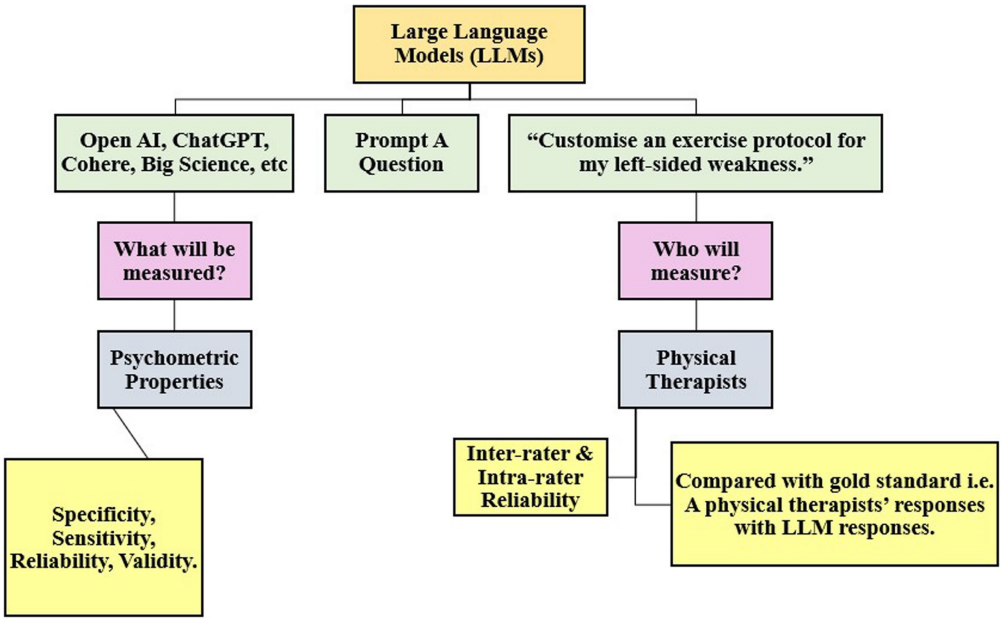

FIGURE 2
Processing steps of a large language model (LLM).

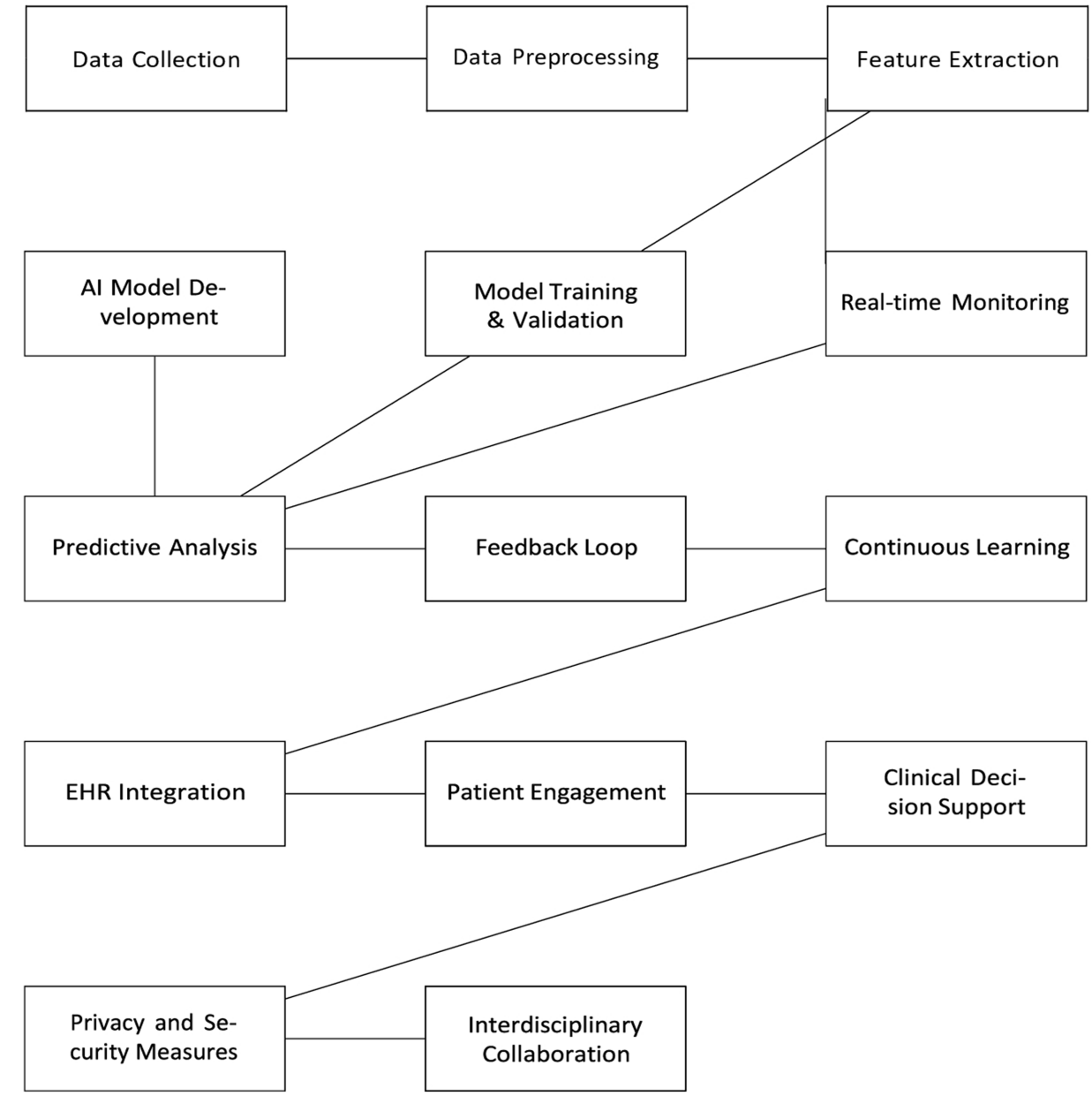

FIGURE 3
Schematic diagram of embracing al in physiotherapy applications. *EHRs: Electronic health records.

its basic functioning. There are also no coding applications, such as visual studio code (VSC), a free support tool application that can be of great help in debugging, refining, and editing web applications. It is the above stakeholders who, in turn, can train the language model systems for developing reliable, valid, sensitive, and specific responses to the prompt or question asked by the client (40).

- *Levels of training:* The training has to be extensive, meaning it should be part of the curriculum where the stakeholders can receive formal training hours under expert guidance in the domains of education, examination, administrative, clinical, and research. Inter-rater and intra-rater reliability will be bi-directional and calculated in two ways. First, there will be human prompts, and second, there will be machine prompts. Similar methods would be applied for machine and human responses. The process of establishing the validity of AI will be similar to validating a new instrument, technique, or method. The gold standard validator will be humans. The responses of the machine must be validated according to face, content, and design with the responses of humans, especially physical therapy experts (41).
- *Will the process be smooth or not?* The smoothness of the process will depend on the type of language models, user interface, user experience, recall value, and most specifically, the feedback of the patient or client. The entire process should be competent enough in terms of patient engagement, patient recall, patient's user

experience, satisfaction, and how truthful, harmless, and hassle-free the processes were. However, feedback from the client at every step becomes imperative (42).

- *Regulation and Recognition by the Physiotherapy organizations globally:* In the meantime, we can predict that AI, its LLMs, its types, and variants will be part of the curriculum recognized by universities, organizations, research institutions as well as physical therapy licensing authorities, associations, and societies for uniform ethical and considerate use worldwide (43, 44).
- *How will it be regulated?* Global accreditation of these LLMs must be regulated by a stringent screening system to ensure malpractices, unethical usage, and plagiarism are avoided at the grassroot level by the above regulating authorities. This will also aid in minimizing plagiarism and unethical misconduct and make stakeholders wary of falling into the trap of predatory influences (45).

TABLE 3 Ways how LLMs can reduce a physical therapist's workload.

Physical therapist-related	Client-related
1. Less time consuming	1. Client engagement and satisfaction
2. Differential diagnosis based on symptoms	2. History taking
3. More clients can be assessed in a given time period.	3. Making notes
4. Workload reduction in documentation	4. Documentation
5. Comprehensive and evidence-based review of literature.	5. Scheduling appointments
6. Uniform system of curricular assessment and evaluation.	6. Suggest experts globally across specialties
	7. Protocol designing
	8. Client follow-up with high recall
	9. Saves effort, money

A call for the physical therapists

The burning question is: will LLMs achieve the desired quality in comparison to human touch? The answer to it is no, as human touch can never be replaced, mimicked, or be imitated. However, it can help a PT significantly by mechanizing other jobs and do it in a much better way than humans (46). Additionally, AI appears to be a current deficit in the medical curriculum, and most students surveyed were supportive of its introduction. These results are consistent with previous surveys conducted internationally (47). Rowe et al. have summarized on how AI can replace PTs by minimizing the extra effort the therapists put and maximizing their output with more objectivity. They also conclude that automation of tasks can be a matter of concern if not guided by ethical principles (48). Supervision and review of PTs become imperative, for example in diagnostic imaging, patient measurement data, and clinical decision support. Tack concludes that the existing literature base can identify cases as a preliminary step where ML is capable of performing equal to or more accurately than human levels in the field of physical therapy (14). Interacting with a patient at first go, engaging the client, taking notes, documenting, taking history, differentially diagnosing the patient's symptoms, patient education and counseling, connecting with the best experts globally without leaving the comfort of your home, scheduling the appointments, analyzing movement, function, restrictions, and customizing a tailor-made treatment protocol in accordance with smart goals can prove to be a boon to PTs saving them time, effort, and money and providing a chance to treat a larger number of patients in a particular time frame (49). Overall, this will engage the client in a much more effective way, maintain communication standards, prevent cumbersome visits to treatment centers, and maximize patient follow-up. Valuable time, money, and energy will be saved and the client can be given the option to make scientifically informed decisions with active participation in the treatment (50) (Table 3).

The physical therapy curriculum differs worldwide in terms of content, subjects, credit hours, nomenclature of the degree, course duration, course outcomes, evaluation methods, basis of evaluation, and types of projects to be completed, even though all PTs treat and assess same conditions and diseases in any corner of the world. Culminating these variations and discrepancies into one is only possible from developing and using a single, concrete type of unified system accessible to all and trained uniformly. The medical education field can hugely benefit from creating common examination patterns globally. The recent success of LLMs, such as BioMedLM, scoring exceptionally high on mock United States Medical Licensing Exams (USMLEs), highlights the immense potential of AI in healthcare (51). When a formally trained licensed PT can treat any client in the world, these variations in training, licensing, and practice must be bonded with a common thread or a language. Therefore, it necessitates the need to develop a common scientific curriculum and assessment protocol for the benefit of the PT's community. These protocols must be error-free, continuously updated, and flexible and competent enough to adapt with continuous feedback assessment (52). Current challenges are listed in Table 4, along with strategies to address challenges associated with implementing generative AI in physical therapy while fostering acceptance and collaboration within the medical community and among patients and users.

Therefore, it becomes necessary for PTs to learn language models that will enable the amalgamation of technology, ML, and the rehabilitation field. Coding must be a part of the curriculum to build and design efficient physical therapy LLMs in clinical, educational, and research settings among health professions education to harness the complete potential of AI and ML requires proactive engagement from physiotherapists (53). We must embrace AI and ML, explore the capabilities of these technologies, and identify opportunities for integrating them into our practice. Start small, experiment, and learn from experience to become familiar with BioMed-LLMs and Biomed Natural Language Processing (NLP) (54). These support tools are crucial for navigating the vast medical literature, extracting valuable insights and mastering Biomed-NLP allows one to process and interpret complex medical information, enhancing one's understanding and decision-making by contributing to dataset *fine-tuning*, our expertise will become invaluable in ensuring that AI and ML models are trained on accurate and scientifically transparent data (55, 56). We, as a community, can also ensure that these language models are optimized for the specific needs of physiotherapy practice in rehabilitation. Collaborating with researchers and developers to create AI and ML tools specifically designed for physiotherapy will directly influence the development of applications and support tools to address unique challenges faced by physiotherapists (57). A

TABLE 4 Current challenges and strategies for future directions in physiotherapy.

Challenges	Strategies
Data availability and quality	1. Collaborate with healthcare institutions for access to diverse and high-quality datasets.
	2. Implement data augmentation techniques to expand dataset diversity.
	3. Ensure data privacy and security measures are robust.
Implementation frameworks	1. Develop standardized protocols for integrating generative AI into physical therapy practices.
	2. Provide training and resources to PT professionals for seamless integration.
	3. Foster interdisciplinary collaborations with AI experts.
Model evaluation	1. Establish comprehensive evaluation criteria, including clinical efficacy and safety.
	2. Utilize cross-validation techniques to assess model performance.
	3. Continuously monitor and update models based on real-world feedback.
Policy and regulatory challenges	1. Advocate for clear guidelines and regulations tailored to AI applications in healthcare.
	2. Collaborate with policymakers and regulatory bodies to ensure compliance.
	3. Stay informed about evolving legal frameworks and adapt accordingly.
Biases	1. Conduct thorough bias assessments during model development.
	2. Implement techniques such as fairness-aware learning to mitigate biases.
	3. Regularly audit AI systems for bias detection and correction.
Patient and community acceptance	1. Educate patients and communities about the benefits and risks of AI in physical therapy.
	2. Solicit feedback and involve stakeholders in decision-making processes.
	3. Maintain transparency regarding AI usage and outcomes.

collaborative environment can be promoted that fosters innovation and the adoption of AI and ML technologies by sharing our experiences, knowledge, and insights, thus supporting our colleagues in their learning journey. Although these models pose challenges such as bias and explainability, they offer a glimpse into the future of medical education and evaluation, where AI will act as a powerful tool to enrich learning and assessment for medical and allied health sciences curriculum and practice (58).

Limitations

Majovsky et al. conducted a study that evaluated the generation of an authentic-looking fabricated scientific paper using ChatGPT and that is possible but human supervision is a mandate (59). Plagiarism detection software and ethical considerations must align with the updated ICMJE recommendations for using AI technology. AI detection tools should be employed to screen and negotiate content, and stringent boundaries must be established. The future trajectory of AI will heavily depend on how these ethical challenges are addressed (60).

This paper is a call for action by physiotherapists to embrace AI and ML by acquiring necessary knowledge and actively shaping their development, we can ensure that our profession remains at the forefront of healthcare in this evolving landscape. We have the opportunity to revolutionize physical therapy and improve patient care. We authors feel this is an apt time and opportunity for PTs to delve into LLMs, not only to utilize it but also to learn to develop and fine-tune the model for a better future. Etymologically speaking, PTs must take charge as experts, i.e., *adept* by being proficient rather than adapting to changing times in the coming decade. Let us consider this opportunity and lead the way to the future of AI-driven healthcare by doing more collaborative studies, creating new beginnings with an admixture of various disciplines. It is time to delve into LLM training, fine-tune the datasets, and provide solutions at par with those of other medical and health sciences professionals for the future. *Language Model* programming in AI is the future of medical and allied health professions, especially physical therapy.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

WN: Conceptualization, Writing – original draft, Writing – review & editing. SS: Writing – original draft, Writing – review & editing. GM: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The manuscript was supported by Datta Meghe Institute of Higher Education and Research, India to help fund this publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? *Am J Med*. (2018) 131:129–33. doi: 10.1016/j.amjmed.2017.10.035

2. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell*. (2023) 6:1169595. doi: 10.3389/frai.2023.1169595

3. Heaton J. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: deep learning. *Genet Program Evolvable Mach*. (2018) 19:305–7. doi: 10.1007/s10710-017-9314-z

4. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan T, Ting DW. Large language models in medicine. *Nat Med*. (2023) 29:1930–40. doi: 10.1038/s41591-023-02448-8

5. Introducing ChatGPT. OpenAI. (2023). Available at: https://openai.com/blog/chatgpt.

6. OpenAI (2023). GPT-4 technical report. ArXiv, abs/2303.08774

7. Ramesh A, Kambhampati C, Monson JR, Drew PJ. Artificial intelligence in medicine. *Ann R Coll Surg Engl*. (2004) 86:334–8. doi: 10.1308/147870804290

8. Preiksaitis C, Rose C. Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. *JMIR Med Educ*. (2023) 9:e48785. doi: 10.2196/48785

9. Pearce J, Chiavaroli N. Rethinking assessment in response to generative artificial intelligence. *Med Educ*. (2023) 57:889–91. doi: 10.1111/medu.15092

10. Shah NH, Entwistle D, Pfeffer MA. Creation and adoption of large language models in medicine. *JAMA*. (2023) 330:866–9. doi: 10.1001/jama.2023.14217

11. ChatGPT Generative Pre-trained TransformerZhavoronkov A. Rapamycin in the context of Pascal's wager: generative pre-trained transformer perspective. *Oncoscience*. (2022) 9:82–4. doi: 10.18632/oncoscience.571

12. Currie G. A conversation with ChatGPT. *J Nucl Med Technol*. (2023) 51:255–60. doi: 10.2967/jnmt.123.265864

13. Kanungo J, Mane D, Mahajan P, Salian S. "Recognizing physiotherapy exercises using machine learning." 2023 Second International Conference on Trends in Electrical, Electronics, and Computer Engineering (TEECCON) 218, (2023).

14. Tack C. Artificial intelligence and machine learning | applications in musculoskeletal physiotherapy. *Musculoskelet Sci Pract*. (2019) 39:164–9. doi: 10.1016/j.msksp.2018.11.012

15. Godse S, Singh S, Khule S, Wakhare SC, Yadav V. Artificially Intelligent. *Physiotherapy*. (2021) 10:77–88. doi: 10.4018/IJBCE.2021010106

16. Durve I, Ghuge S, Patil S, Kalbande D. Machine learning approach for physiotherapy assessment. 2019 International Conference on Advances in Computing, Communication, and Control (ICAC3) (2019). doi: 10.1109/ICAC347590.2019.9036783

17. Jovanovic M, Seiffarth J, Kutafina E, Jonas SM. Automated error detection in physiotherapy training. *Stud Health Technol Inform*. (2018) 248:171. doi: 10.3233/978-1-61499-858-7-164

18. Beierle C, Timm IJ. Intentional forgetting: an emerging field in AI and beyond. *Künstl Intell*. (2019) 33:5–8. doi: 10.1007/s13218-018-00574-x

19. Hatem R, Simmons B, Thornton JE. A call to address AI "hallucinations" and how healthcare professionals can mitigate their risks. *Cureus*. (2023) 15:e44720. doi: 10.7759/cureus.44720

20. Grünebaum A, Chervenak J, Pollet SL, Katz A, Chervenak FA. The exciting potential for ChatGPT in obstetrics and gynecology. *Am J Obstet Gynecol*. (2023) 228:696–05. doi: 10.1016/j.ajog.2023.03.009

21. Makino M, Yoshimoto R, Ono M, Itoko T, Katsuki T, Koseki A, et al. Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. *Sci Rep*. (2019) 9:11862. doi: 10.1038/s41598-019-48263-5

22. Chervenak J, Lieman H, Blanco-Breindel M, Jindal S. The promise and peril of using a large language model to obtain clinical information: ChatGPT performs strongly as a fertility counseling tool with limitations. *Fertil Steril*. (2023) 120:575–83. doi: 10.1016/j.fertnstert.2023.05.151

23. Drum B, Shi J, Peterson B, Lamb S, Hurdle JF, Gradick C. Using natural language processing and machine learning to identify internal medicine-pediatrics residency values in applications. *Acad Med J Assoc Am Med Coll*. (2023) 98:1278–82. doi: 10.1097/ACM.0000000000005352

24. Melvin RL, Broyles MG, Duggan EW, John S, Smith AD, Berkowitz DE. Artificial intelligence in perioperative medicine: a proposed common language with applications to FDA-approved devices. *Front Digl Health*. (2022) 4:872675. doi: 10.3389/fdgth.2022.872675

25. Brin D, Sorin V, Vaid A, Soroush A, Glicksberg BS, Charney AW, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. *Sci Rep*. (2023) 13:16492.

26. Meskó B. The impact of multimodal large language models on health Care's future. *J Med Internet Res*. (2023) 25:e52865. doi: 10.2196/52865

27. Ullah E, Parwani A, Baig MM, Singh R. Challenges and barriers of using large language models (LLM) such as ChatGPT for diagnostic medicine with a focus on digital pathology – a recent scoping review. *Diagn Pathol*. (2024) 19:43. doi: 10.1186/s13000-024-01464-7

28. Wilhelm TI, Roos J, Kaczmarczyk R. Large language models for therapy recommendations across 3 clinical specialties: comparative study. *J Med Internet Res*. (2023) 25:e49324. doi: 10.2196/49324

29. Chen PF, He TL, Lin SC, Chu YC, Kuo CT, Lai F, et al. Training a deep contextualized language model for international classification of diseases, 10th revision classification via federated learning: model development and validation study. *JMIR Med Inform*. (2022) 10:e41342. doi: 10.2196/41342

30. Stanford CRFM (2022). Available at: https://crfm.stanford.edu/2022/12/15/biomedlm.html

31. Glauberman G, Ito-Fujita A, Katz S, Callahan J. Artificial intelligence in nursing education: opportunities and challenges. *Hawaii J Health Soc Welf*. (2023) 82:302–5.

32. Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. *NPJ Digit Med*. (2019) 2:77. doi: 10.1038/s41746-019-0155-4

33. Buchlak QD, Esmaili N, Bennett C, Farrokhi F. Natural language processing applications in the clinical neurosciences: a machine learning augmented systematic review. *Acta Neurochir Suppl*. (2022) 134:277–89. doi: 10.1007/978-3-030-85292-4_32

34. Liu S, Wright AP, Patterson BL, Wanderer JP, Turer RW, Nelson SD, et al. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. *J Am Med Inform Assoc*. (2023) 30:1237–45. doi: 10.1093/jamia/ocad072

35. Temsah O, Khan SA, Chaiah Y, Senjab A, Alhasan K, Jamal A, et al. Overview of early ChatGPT's presence in medical literature. *Insights from a Hybrid Literature Review by ChatGPT and Human Experts Cureus*. (2023) 15:e37281. doi: 10.7759/cureus.37281

36. Salvagno M, Taccone FS, Gerli AG. *Can artificial intelligence help for scientific writing? Critical care*. London, England.

37. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthcare J*. (2021) 8:e188–94. doi: 10.7861/fhj.2021-0095

38. Leonardi M, Lee H, Kostanjsek N, et al. 20 years of ICF-international classification of functioning, disability and health: uses and applications around the world. *Int J Environ Res Public Health*. (2022) 19:11321. doi: 10.3390/ijerph191811321

39. Rauch A, Cieza A, Stucki G. How to apply the international classification of functioning, disability and health (ICF) for rehabilitation management in clinical practice. *Eur J Phys Rehabil Med*. (2008) 44:329–42.

40. Voytovich L, Greenberg C. Natural language processing: practical applications in medicine and investigation of contextual autocomplete. *Acta Neurochir*. (2022) 134:207–14. doi: 10.1007/978-3-030-85292-4_24

41. Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, et al. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the chat-GPT model. *Research Square*. (2023) rs.3:2566942. doi: 10.21203/rs.3.rs-2566942/v1

42. Wen A, Elwazir MY, Moon S, Fan J. Adapting and evaluating a deep learning language model for clinical why-question answering. *JAMIA Open*. (2020) 3:16–20. doi: 10.1093/jamiaopen/ooz072

43. Bhatnagar R, Sardar S, Beheshti M, Podichetty JT. How can natural language processing help model informed drug development? A review. *JAMIA Open*. (2022) 5:ooac043. doi: 10.1093/jamiaopen/ooac043

44. Irvin JA, Pareek A, Long J, Rajpurkar P, Eng DK, Khandwala N, et al. CheXED: comparison of a deep learning model to a clinical decision support system for pneumonia in the emergency department. *J Thorac Imaging*. (2022) 37:162–7. doi: 10.1097/RTI.0000000000000622

45. Stahl CC, Jung SA, Rosser AA, Kraut AS, Schnapp BH, Westergaard M, et al. Natural language processing and entrustable professional activity text feedback in surgery: a machine learning model of resident autonomy. *Am J Surg*. (2021) 221:369–75. doi: 10.1016/j.amjsurg.2020.11.044

46. Pustina D, Coslett HB, Ungar L, Faseyitan OK, Medaglia JD, Avants B, et al. Enhanced estimations of post-stroke aphasia severity using stacked multimodal predictions. *Hum Brain Mapp*. (2017) 38:5603–15. doi: 10.1002/hbm.23752

47. Stewart J, Lu J, Gahungu N, Goudie A, Fegan PG, Bennamoun M, et al. Western Australian medical students' attitudes towards artificial intelligence in healthcare. *PLoS One*. (2023) 18:e0290642. doi: 10.1371/journal.pone.0290642

48. Rowe M, Nicholls DA, Shaw J. How to replace a physiotherapist: artificial intelligence and the redistribution of expertise. *Physiother Theory Pract*. (2022) 38:2275–83. doi: 10.1080/09593985.2021.1934924

49. Kristinsson S, Zhang W, Rorden C, Newman-Norlund R, Basilakos A, Bonilha L, et al. Machine learning-based multimodal prediction of language outcomes in chronic aphasia. *Hum Brain Mapp*. (2021) 42:1682–98. doi: 10.1002/hbm.25321

50. Shahsavar Y, Choudhury A. User intentions to use ChatGPT for self-diagnosis and health-related purposes: cross-sectional survey study. *JMIR Hum Factors*. (2023) 10:e47564. doi: 10.2196/47564

51. Brin D, Sorin V, Vaid A, Soroush A, Glicksberg BS, Charney AW, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. *Sci Rep*. (2023) 13:16492. doi: 10.1038/s41598-023-43436-9

52. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. (2023) 9:e45312. doi: 10.2196/45312

53. Athaluri SA, Manthena SV, Kesapragada VK, Yarlagadda V, Dave T, Duddumpudi RTS. Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus*. (2023) 15:e37432. doi: 10.7759/cureus.37432

54. Choo YJ, Chang MC. Use of machine learning in stroke rehabilitation: a narrative review. *Brain NeuroRehabil*. (2022) 15:e26. doi: 10.12786/bn.2022.15.e26

55. Russo AG, Ciarlo A, Ponticorvo S, Di Salle F, Tedeschi G, Esposito F. Explaining neural activity in human listeners with deep learning via natural language processing of narrative text. *Sci Rep*. (2022) 12:17838. doi: 10.1038/s41598-022-21782-4

56. Sim JA, Huang X, Horan MR, Stewart CM, Robison LL, Hudson MM, et al. Natural language processing with machine learning methods to analyze unstructured patient-reported outcomes derived from electronic health records: a systematic review. *Artif Intell Med*. (2023) 146:102701. doi: 10.1016/j.artmed.2023.102701

57. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. (2022) 28:31–8. doi: 10.1038/s41591-021-01614-0

58. Hazarika I. Artificial intelligence: opportunities and implications for the health workforce. *Int Health*. (2020) 12:241–5. doi: 10.1093/inthealth/ihaa007

59. Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: Pandora's box has been opened. *J Med Internet Res*. (2023) 25:e46924. doi: 10.2196/46924

60. ICMJE News & Editorials (2023). Available at: https://www.icmje.org/news-and-editorials/updated_recommendations_may2023.html.