Vol.4, Issue 01, April 2024
Journal of Artificial Intelligence General Science JAIGS

https://ojs.boulibrary.com/index.php/JAIGS

The Impact of Large Language Models on Medical Education: Preparing for a Revolutionary Shift in Doctor Training

Dr. Sreeram Mullankandy

MBBS, MBA, Boston University, USA.

ARTICLEINFO
Article History:
Received: 01.05.2024
Accepted: 15.05.2024
Online: 22.05.2024
Keyword: artificial intelligence; large language models; medical education

ABSTRACT

Artificial intelligence holds immense potential to transform healthcare, though its widespread implementation has yet to be realized. This lag is partly because efforts have traditionally focused on easily predicted rather than easily actionable problems. Large language models (LLMs) represent a paradigm shift in our approach to artificial intelligence due to their accessibility and the fact that frontline clinicians are already testing them and identifying potential applications. LLMs in healthcare could significantly reduce clerical burdens, enhance patient education, and more. As we enter this new era of healthcare delivery, LLMs will bring both opportunities and challenges to medical education.[1-5]

Future models should be designed to help trainees develop clinical reasoning skills, promote evidence-based medicine, and provide case-based training opportunities. LLMs may also necessitate changes in how clinical documentation is taught. Additionally, trainees can contribute to training and refining the next generation of LLMs as we explore the best ways to integrate these models into medical education.

Whether we are ready or not, LLMs will soon be integrated into various aspects of clinical practice. We must collaborate closely with students and educators to ensure these models are developed with trainees in mind, guiding medical education responsibly into the next era.[21]

Introduction:

While artificial intelligence (AI) has begun to revolutionize various sectors, healthcare has significantly lagged behind other industries[6-8] . Despite the development of numerous machine learning models in healthcare, only a small fraction has been successfully implemented . This gap between model creation and practical application arises partly because current AI efforts often target the wrong use cases and are not always designed with ethical considerations in mind.

Many AI models focus on problems that are easiest to predict rather than areas where AI could drive substantial improvements. Models with good predictive performance do not always translate into actionable insights. For instance, an algorithm that predicts hospital readmissions might perform well but may lack clinical utility if it fails to identify preventable readmissions or highlight modifiable risk factors . This focus often neglects valuable use cases like customer support or back-office process optimization, which are the most common areas for AI adoption in other industries .

Even predictive models developed by major electronic health record (EHR) vendors with the goal of operationalization have frequently fallen short, delivering solutions that lack robustness and ethical integrity. Some industry-developed models have shown poor performance, limited external validity, and serious biases against vulnerable patient populations.[21]

A Transformative Paradigm Shift

Large language models (LLMs) are semisupervised, generative transformer models trained on massive amounts of text data. They contextualize the sequential nature of words in a sentence to predict the most plausible response, allowing them to interpret user prompts and provide conversational responses. Examples include the recently popularized ChatGPT by OpenAI, which also performs tasks such as summarization and translation.

While we are only beginning to explore the potential use cases in healthcare, these tools are speculated to have transformative potential. Their wide availability to the public enables a paradigm shift in identifying AI applications in medicine. Testing use cases with simple prompt engineering requires far fewer data science resources than building a model from scratch. Clinicians are already experimenting with LLMs for tasks such as drafting insurance appeals, writing preauthorization letters, summarizing patient instructions at different reading levels, generating differential diagnoses, and more. If deployed in a trustworthy and secure manner, many of these use cases could significantly improve daily medical work.[10-15]

Possibilities And Pitfalls

If implemented thoughtfully and with a keen awareness of their limitations, large language models (LLMs) could significantly enhance patient care, reduce clerical workloads, and improve satisfaction for both patients and providers. Some potential use cases for LLMs in healthcare are summarized below.

Table 1. Sample use cases for large language models in health care

Use Case	Possibilities
Task automation	• Drafting prior authorization requests, insurance appeals letters, work letters, disability paperwork, medical leave requests, and more • Triaging patient messages (in-basket) and drafting a skeleton response for provider review
Patient education	• Translating medical jargon into clinic visit instructions or discharge instructions at a patient-friendly reading level • Drafting lifestyle counseling recommendations such as diet, exercise, basic physical therapy etc. • Creating medication tables with clear instructions for use and side effects
Clinical documentation	• Using prior clinic visit notes, diagnoses, laboratory test results, and imaging results to prepare the outlines of a note • Summarizing long hospital course to prepare an outline of a discharge summary • Pulling out key information from a note to identify billing level
Diagnostic assistance	• Offering a basic differential diagnosis based on a problem representation or the available clinical data in the EHR • Identifying case reports of patients with similar problems or rare diseases
Literature review	• Summarizing key studies and their findings in a particular field • Identifying and summarizing the articles relevant to a new research project • Copyediting academic manuscripts to help bridge equity gaps for trainees and providers for whom English is a second language

Definition of abbreviation: EHR = electronic health record.

Reducing clerical work could greatly enhance physicians' daily workflow. By automating tasks or assisting with clinical documentation, LLMs could allow physicians to spend less time on screens and more time with patients. With physician burnout on the rise due to increasing documentation and administrative burdens, LLMs offer a promising solution to alleviate this load. Future iterations of these models, trained on medical record data, could draft prior authorizations or family medical leave letters tailored to individual patients. They could also prepare outlines of clinical notes before patient encounters, including relevant past visits, lab results, and imaging, effectively "prerounding" for doctors.[22-23]

Eventually, these models could work in conjunction with ambient, automatic audio recordings of clinic visits to function as electronic scribes, preparing structured notes by the end of the visit for clinicians to review and sign.

While previous digital scribes have faced issues with accuracy and formatting, future LLMs trained with reinforcement learning from human feedback could better understand and translate human conversation into clinically useful notes. Additionally, these models might help with billing by interpreting and coding notes based on existing documentation. Such advancements could significantly reduce clerical work for physicians, allowing them to focus more on providing high-quality patient care.[24]

Bridging Gaps in Patient Education with LLMs

Large language models (LLMs) could play a crucial role in enhancing patient education. Effective communication and education are vital for safe patient care, yet barriers remain, especially for patients with limited English proficiency. LLMs have the potential to act as translators, converting complex medical jargon into simpler terms for patients at varying reading levels and in different languages. This capability could be applied to clinic visit summaries, discharge instructions, lifestyle counseling, medication counseling, and more, thereby improving health equity in communication for vulnerable patients.

When trained on medical-specific or electronic health record (EHR) data, these models could generate first drafts of patient instructions based on simple prompts like "blood pressure monitoring instructions," and provide summaries tailored to the patient's reading level or native language. Such interventions would not only enhance provider efficiency but also significantly improve patients' understanding of their health.[25]

However, the use of LLMs in healthcare requires careful consideration of their limitations. The U.S. Department of Health and Human Services has outlined six key principles of trustworthy AI, which are essential for evaluating LLMs before implementation:

1.Robust and Reliable: LLMs must be accurate and may require additional training on healthcare-specific data to minimize incorrect information. Regular retraining is necessary to incorporate new studies, guidelines, and recommendations. Given these limitations, even OpenAI has restricted the use of its models for diagnostic or therapeutic services as of March 23, 2023.[17-20]

2. Fair and Impartial: LLMs must be trained to avoid explicit or implicit biases in their training data, which could perpetuate socioeconomic disparities in healthcare. Ongoing efforts to mitigate these biases include developing libraries to measure and describe model bias.

3.Transparent and Explainable: The output of LLMs must be understandable so that clinicians can reasonably interpret the responses. Currently, LLMs cannot reference their source materials to justify responses or provide a certainty level for any given answer, features that are needed in future iterations.

4.Responsible and Accountable: Strict review and oversight protocols by clinicians are necessary to ensure that LLMs provide safe and effective patient care.

5. Safe and Secure: LLMs trained on patient data must protect that data. Future models incorporating EHR data will require safeguards to ensure data security.

6. Effective and Trustworthy: Following these principles ensures that LLMs can be responsibly integrated into healthcare, improving patient education and care outcomes.

Trainees And Llms

As we enter a new era of healthcare delivery where large language models (LLMs) are readily available to trainees, we foresee both significant opportunities and challenges in how these models will impact the next generation of physicians and the ways we teach them.

Coaching for Synthesizing Patient Data into an Assessment and Plan

One of the most challenging transitions for medical trainees is moving from pre-clerkship to clerkship education. During this phase, students must learn to consolidate vast amounts of medical knowledge into relevant clinical insights for the patients they encounter. They must process real-world data to produce problem representations, differential diagnoses, and plausible care plans. Traditionally, early learners develop these skills through close interaction with supervising residents or attending physicians. However, these high-stakes interactions may not suit all trainees.

LLMs could serve as a low-stakes alternative to support self-directed learning. Imagine a future LLM trained on clinical note text, discrete data, PubMed, and other medical references that can review draft clinical notes from trainees, provide recommendations for differential diagnoses or care plans, and explain these suggestions with helpful references. While LLMs cannot replace clinical expertise, they can offer valuable guidance. However, it is crucial to establish best practices for how learners interact with these models. An AI model that reviews a student's work and provides feedback can aid in consolidating knowledge, but one that does the thinking for them could be detrimental. Ensuring the validity of LLM responses for various use cases is critical before exposing learners to potentially erroneous outputs.

Facilitating Effective Application of Evidence-Based Practice

In the era of evidence-based medicine, trainees are instructed to actively integrate the practice of seeking and evaluating medical literature into patient care. This process involves defining the problem, identifying relevant information, searching for pertinent studies, interpreting their findings, and applying them to patient care. While evaluating studies and determining their applicability remains crucial in training, the search for evidence is often overlooked.

While current large language models (LLMs) may not cite sources within their training data, future iterations could be trained for this purpose using a reliable corpus of articles like PubMed, such as BioBert or BioMedLM. These advanced LLMs could efficiently highlight high-quality studies relevant to a given problem, allowing trainees to focus on appraisal. Additionally, trainees could utilize these models to understand the evidence behind established practice patterns, such as managing patients with heart failure with reduced ejection fraction.

Table 2. Principles of trustworthy artificial intelligence applied to large language models

Principles of Trustworthy AI	Specific Examples for LLMs
Robust/reliable	• Current LLMs are at risk of "hallucination," or providing plausible-sounding but incorrect information. For example, some current LLMs will cite studies that sound realistic but do not exist. This is an evolving problem, and LLMs themselves may be able to help identify hallucination (44). • Some LLMs may have knowledge cutoffs, or date limits on the most recent data that have been used to train the model. Some periodic retraining of the model would be necessary to incorporate new studies, guidelines, and recommendations as they arise.
Fair/impartial	• LLMs can incorporate biases found in their training data, which could inadvertently perpetuate harmful racial, sex, and other biases (27, 28). For example, this could lead to bias in generating a work letter for male versus female-identifying patients or in drafting a response to a patient portal message about pain for White versus Black patients.
Transparent/explainable	• Current publicly available LLMs are not able to provide much explainability in their responses in the form of either references to the source materials that they are using to formulate a response or providing an assessment of certainty in the accuracy of a response. This limits a user's ability to accurately interpret a model's responses.
Responsible/accountable	• Because these tools are still in their infancy, there will need to be strict supervision and oversight from the physicians who use LLMs to make sure that information conveyed by these models is not inaccurate or incomplete. For example, although LLMs may be able to draft patient portal message responses, a physician still needs to review and read this message before sending it.
Safe/secure	• Current LLMs have not been built using any specific patient data, but already they have had issues with the leak of conversations between users (30), which is why both physicians and healthcare organizations must be cautious with their use. Physicians should not use protected health information through unsecured online LLMs, and healthcare organizations should create systems for secure computing and business associate agreements to ensure the safety of these data if partnering with organizations that build LLMs.

Generating Synthetic Training Cases for Interactive Learning

Case-based training is a cornerstone of medical education, involving authored cases with focused questions to stimulate clinical reasoning. If trained on electronic health record (EHR) data, an LLM could potentially generate synthetic patient cases and interact with trainees to reveal information gradually. Each case would be unique, challenging learners to manage diagnostic uncertainty in a simulated environment. Future iterations could even assess trainees as a novel form of examination.

However, there are significant limitations to consider regarding LLMs in medical education. Concerns about accuracy are valid, as recent studies show that while LLMs trained on medical-specific data can pass the United States Medical Licensing Examination, their accuracy ranges from 60% to 68%. Furthermore, these models may provide plausible but inaccurate responses, posing challenges for trainees in discerning fact from fiction. Additionally, current LLMs lack the ability to justify or provide references for their responses.

The implementation of LLMs into clinical workflows may necessitate a reevaluation of teaching priorities for trainees. As ambient "scribe" LLMs become more prevalent, traditional teaching of history taking and documentation structures may become less essential compared to communication skills with patients. Educators will need to adapt their teaching approach, focusing on conceptual frameworks and clinical reasoning skills rather than rote memorization. As LLMs alleviate nonclinical burdens, educators may have more time to model these skills with trainees and patients alike.

Conclusions

Whether we are prepared or not, large language models (LLMs) like ChatGPT, Med-PaLM, and others will soon become integral to various facets of clinical practice. We anticipate that their integration will revolutionize both patient and provider experiences, as well as reshape medical education for our trainees. As academic medical centers are pioneers in clinical innovation and education, it is imperative for us to engage with stakeholders to thoughtfully design and implement these tools.

Developing key partnerships with technical leaders within our organizations and external companies specializing in LLMs is essential. We must carefully consider the privacy, safety, and security implications, particularly when utilizing patient data. Collaborating with frontline clinicians is crucial to comprehending their successes and challenges with LLMs, ensuring usability in our implementations.

Moreover, collaboration with educators and students is vital to ensure that these models are designed with trainees' needs in mind, guiding medical education into the next era with ethics and responsibility at the forefront.

References List:

[1]. Prakash, S., Malaiyappan, J. N. A., Thirunavukkarasu, K., & Devan, M. (2024). Achieving Regulatory Compliance in Cloud Computing through ML. AIJMR-Advanced International Journal of Multidisciplinary Research, 2(2).

[2]. Malaiyappan, J. N. A., Prakash, S., Bayani, S. V., & Devan, M. (2024). Enhancing Cloud Compliance: A Machine Learning Approach. AIJMR-Advanced International Journal of Multidisciplinary Research, 2(2).

[3]. Devan, M., Prakash, S., & Jangoan, S. (2023). Predictive Maintenance in Banking: Leveraging AI for Real-Time Data Analytics. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(2), 483-490.

[4]. Eswaran, P. K., Prakash, S., Ferguson, D. D., & Naasz, K. (2003). Leveraging Ip For Business Success. International Journal of Information Technology & Decision Making, 2(04), 641-650.

[5]. Prakash, S., Malaiyappan, J. N. A., Thirunavukkarasu, K., & Devan, M. (2024). Achieving Regulatory Compliance in Cloud Computing through ML. AIJMR-Advanced International Journal of Multidisciplinary Research, 2(2).

[6]. Malaiyappan, J. N. A., Prakash, S., Bayani, S. V., & Devan, M. (2024). Enhancing Cloud Compliance: A Machine Learning Approach. AIJMR-Advanced International Journal of Multidisciplinary Research, 2(2).

[7]. Biswas, A. (2019). Media Insights Engine for Advanced Media Analysis: A Case Study of a Computer Vision Innovation for Pet Health Diagnosis. International Journal of Applied Health Care Analytics, 4(8), 1-10.

[8] Chopra, B., & Raja, V. (2024). Toward Enhanced Privacy in Digital Marketing: An Integrated Approach to User Modeling Utilizing Deep Learning on a Data Monetization Platform. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 1(1), 91-105.

[9]. Raja, V. (2024). Fostering Privacy in Collaborative Data Sharing via Auto-encoder Latent Space Embedding. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 4(1), 152-162.

[10]. Raja, V. ., & chopra, B. . (2024). Exploring Challenges and Solutions in Cloud Computing: A Review of Data Security and Privacy Concerns. Journal of Artificial Intelligence General Science (JAIGS) ISSN:3006-4023, 4(1), 121–144. https://doi.org/10.60087/jaigs.v4i1.86

[11]. SARIOGUZ, O., & MISER, E. (2024). Data-Driven Decision-Making: Revolutionizing Management in the Information Era. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 4(1), 179-194.

[12]. Raja, V. (2024). Exploring Challenges and Solutions in Cloud Computing: A Review of Data Security and Privacy Concerns. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 4(1), 121-144.

[13]. Biswas, A. (2019). Media Insights Engine for Advanced Media Analysis: A Case Study of a Computer Vision Innovation for Pet Health Diagnosis. International Journal of Applied Health Care Analytics, 4(8), 1-10.

[14]. Chennupati, A. (2024). The evolution of AI: What does the future hold in the next two years.

[15]. Chennupati, A. (2024). Addressing the climate crisis: The synergy of AI and electric vehicles in combatting global warming. World Journal of Advanced Engineering Technology and Sciences, 12(1), 041-046.

[16]. Chennupati, A. (2024). The threat of artificial intelligence to elections worldwide: A review of the 2024 landscape. World Journal of Advanced Engineering Technology and Sciences, 12(1), 029-034.

[17]. Talati, D. (2023). Telemedicine and AI in Remote Patient Monitoring. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), 254-255.

[18]. Talati, D. (2023). Artificial Intelligence (Ai) In Mental Health Diagnosis and Treatment. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), 251-253.

[19]. Talati, D. (2023). AI in healthcare domain. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), 256-262.

[20]. Talati, D. (2024). AI (Artificial Intelligence) in Daily Life. Authorea Preprints.

[21]. Fan, K. (2024). Implications of Large Language Models in Medical Education. https://doi.org/10.62594/brmo4385

[22]. Warnes, M. (2024). The Efficacy of NVivo in Conducting Literature Reviews: A Case Study on Defining "Teaching Excellence." https://doi.org/10.62594/vraa3705

[23]. Ferdinand, J. (2024). Recognising Clinical Signs and Symptoms on Black, Asian and Minority Ethnic (BAME) Skin Types. https://doi.org/10.62594/ovfn5405

[24]. Chaudhary, G., Yadav, S., Bastola, P., & Limbu, S. (2024). Challenges and Opportunities in Applying Transformative Learning Theory (A Critical Reflection): A Collaborative Autoethnography. https://doi.org/10.62594/qwfc9551