



The impending impacts of large language models on medical education

Sangzin Ahn^{1,2}

¹Department of Pharmacology and Pharmacogenomics Research Center and ²Center for Personalized Precision Medicine of Tuberculosis, College of Medicine, Inje University, Busan, Korea

The advent of large language models

Large language models (LLMs) are machine learning models that are trained on extremely large datasets of text and are capable of multiple natural language processing tasks, such as translation, summarization, and grammar correction. By learning which word (or token) is most probable to appear after a sequence of preceding words in a self-supervised (no labeling required) manner, the LLM is able to predict the next single word, and is therefore described as generative. A prompt, a chunk of text that usually describes the objective, is provided to the model, and by iteratively predicting the next word to follow the prompt, the LLM can generate a long sequence of coherent and grammatically correct text. At first glance, it may seem that LLMs can only perform an autocomplete function, but by carefully crafting the prompt (also known as prompt engineering), LLMs can perform a variety of tasks. For example, if the LLM is prompted with “Translate

this into French: what rooms do you have available?” the model will respond “Quels sont les chambres que vous avez disponibles?” Additionally, if the prompt “Correct this to standard English: she no went to the market.” is provided, the response will be “She did not go to the market [1].”

It is well known that scaling up language models (amount of computation, number of model parameters, and training dataset size) results in better performance in downstream tasks. Often, the effect of scaling increases predictably, but some emergent abilities are observed not in smaller models but in larger models. Some examples include arithmetic, transliterating from the international phonetic alphabet, recovering a word from its scrambled letters, and question answering [2]. The advancement in prompt engineering and deeper studies in the scaling law of language models are leading to an era where LLMs are becoming increasingly versatile in a variety of tasks, which were not considered possible a few years ago. In this article, the author will discuss four abilities of LLMs and their potential impacts on medical education.

Received: December 30, 2022 • Revised: February 2, 2023 • Accepted: February 7, 2023
Corresponding Author: Sangzin Ahn (<https://orcid.org/0000-0003-2749-0014>)
Department of Pharmacology, College of Medicine, Inje University, 75 Bokji-ro, Busanjin-gu, Busan 47392, Korea
Tel: +82.51.890.5909 Fax: +82.50.4043.1326 email: sangzinahn@inje.ac.kr

Korean J Med Educ 2023 Mar; 35(1): 103–107
<https://doi.org/10.3946/kjme.2023.253>
eISSN: 2005-7288

© The Korean Society of Medical Education. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ability to retrieve information: self-learning with dynamic text

LLMs not only have information about language, but also implicitly contain general information embedded within their parameters. A recent study showed that an LLM with instruction prompt tuning can perform medical question answering and reasoning, and also showed an accuracy of 67.6% on MedQA (US Medical License Exam) questions [3]. In the near future, medical students will be able to access highly sophisticated LLM-based medical knowledge bases that allow for the creation of dynamic learning materials tailored to their specific needs and questions. This approach differs significantly from traditional methods of education that rely on static texts, which are written in advance by an author who assumes the needs of the reader. The use of dynamic text enables a more personalized and effective learning experience, as it provides students with highly accurate and timely information that is contextually relevant to their individual needs. This approach to learning also allows for a more efficient acquisition of knowledge, as students can quickly obtain answers to their cascading questions and delve more deeply into topics of particular interest.

Ability to generate essays and articles: transformation of evaluation methods

LLMs have demonstrated the ability to summarize documents, rewrite a given paragraph, and even write a whole essay using a list of keywords [4]. This has the potential to significantly impact written evaluations and potentially render assignments in the form of essays that are based on general information obsolete. Assignments

should be designed in a way that challenges students to apply critical thinking, despite the use of LLM tools, by providing materials that require comprehension or demand the application of personal experiences or unique contexts. Instant assessments can be utilized to limit the influence of LLMs and increased use of formative assessments can help to accurately assess students' academic achievements. Evaluations may also progressively shift towards more oral forms. Utilizing speech-to-text software, LLMs can provide immediate feedback to students and facilitate discussions with instructors. LLMs can also summarize and assess these discussions, allowing for the accumulation of formative assessments over time. This shift towards oral evaluations can be beneficial for both students and teaching staff. For students, it promotes active participation and listening in an engaging environment, enhancing the learning experience. For teaching staff, it allows for more efficient progress assessment and teaching, as evaluations and assessments can be conducted concurrently.

Ability to generate human-like speech: interacting with realistic patient chatbots

LLMs have demonstrated the ability to generate human-like speech through the iterative injection of prompts with previous dialogues. This capability allows LLMs to generate realistic conversations that are coherent in context, leading some to consider the possibility of LLM-powered chatbots exhibiting consciousness [5]. The use of simulated patient chatbots powered by LLMs can assist medical students in improving their clinician-patient communication, clinical information retrieval, and problem-solving skills. Through simulated conversations with the highly accessible chatbot, medical students can practice

medical interviewing, diagnostic reasoning, and patient explanation of treatment options. The incorporation of a chatbot in an exam setting can also facilitate more accurate student assessment, as the computer can provide objective feedback on performance. Additionally, such chatbots enable students to gain experience interacting with a diverse range of virtual patients, including those with disabilities or rare medical conditions, which may not be feasible to be performed by standardized patient actors. This exposure to a wide range of medical conversations can help medical students become better prepared for their future medical practice.

Ability to reason in a form of chain of thought: learn clinical reasoning from LLMs

LLMs have the ability to analyze and understand the relationships between words and concepts in a text or dataset in order to perform reasoning. This process, called chain of thought reasoning, allows LLMs to logically follow a sequence of ideas and make informed decisions based on their analysis [6]. LLMs are given specific prompts to guide them in producing step-by-step explanations that lead to a conclusion. Recent research has demonstrated that LLMs are also capable of performing reasoning in the medical field and are able to answer medical questions with a rather high level of accuracy [3,7]. Clinical reasoning is a crucial skill that medical education aims to cultivate in students. In the near future, medical students will be able to use LLM-based systems to ask questions and receive explanations in the form of a chain of thought. Novice learners can benefit from the LLM-based system by learning about the causes and consequences of diseases through explanations of pathophysiological and biological processes. For medical

students in the clinical phase, the LLM-based system can help in the development of reasoning skills by showing generation of tentative hypotheses and deducing or refuting them.

Current limitations of LLMs for medical education

Hallucination, the generation of false or logically incorrect text that appears plausible and grammatically correct, is a known issue in LLMs. This can lead to confusion or misinformation for learners and poses a challenge for the use of LLMs as a learning system. To address this problem, prompt chaining or fact-checking methods are being explored [8]. A combination of a foundation language model and a knowledge base for querying factual information may be used, which enables providing appropriate contextual information and also staying up-to-date with the latest medical information. Another significant concern with LLMs is inconsistency, as small changes in the prompt can result in divergent responses, undermining the reliability of the model. Research is ongoing in methods to improve the consistency of LLMs, such as careful prompt design, adjustments to training parameters, and correcting incorrect beliefs [9]. Additionally, task-specific and domain-specific LLMs are being developed to improve natural language processing in specific fields, such as the biomedical domain. These efforts include training on high-quality text, using domain-specific tokenizers or vectorizers, and fine-tuning after training [10]. While LLMs currently have limitations, it is hoped that these issues will be addressed in the near future, enabling the potential use cases discussed in this article.

Recommendations and conclusion

Curiosity is a cognitive trait that motivates individuals to seek out new knowledge and experiences, which is crucial for learning and personal development. In the field of medicine, it is especially relevant, as the pursuit of knowledge is vital for providing effective patient care. Despite its significance, curiosity is often not adequately fostered in medical education [11]. In South Korea, a trend has emerged towards the implementation of criterion-referenced assessment in medical schools. This type of assessment evaluates a student's performance against predefined standards rather than comparing them to their peers. This allows curricula to be designed to cover the minimum required information, freeing up time for students to engage in self-directed learning and explore their own interests. Lectures and other teacher-centered instructional methods should be designed to be engaging and stimulate curiosity, rather than simply conveying information. Students can utilize LLM-based learning systems to delve into their own questions and interests developed during class through inquiry-based self-learning.

As the field of information technology and artificial intelligence (AI) continues to advance at a rapid rate, there is an increasing demand for T-shaped professionals who have the ability to cross disciplinary boundaries and possess a diverse range of skills. AI literacy is a crucial component of this, as it enables individuals to comprehend and utilize AI systems to their full potential. This is similar to how LLMs can serve as versatile multi-purpose machines when given proper prompts by the user. Medical students today will encounter the opportunities and challenges associated with the integration of AI into medicine throughout their careers as doctors. To effectively teach these students, teachers must also be

proficient in the use of novel technology and provide an AI-integrated learning environment that is advanced and efficient.

In summary, LLMs have the potential to transform medical education and assessment through dynamic learning materials, oral evaluations, simulated patient interactions, and the ability to support reasoning processes. To effectively teach and promote AI literacy among medical students, teachers should adopt engaging, inquiry-based teaching methods and become proficient in the use of novel technology, providing an AI-integrated learning environment. By doing so, they can help foster curiosity in their students and prepare them for the integration of AI into the field of medicine.

ORCID:

Sangzin Ahn: <https://orcid.org/0000-0003-2749-0014>

Acknowledgements: The author would like to thank Jong Tae Lee from the Department of Preventive Medicine, College of Medicine, Inje University for the constructive discussions regarding future medical education.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (no., 2018R1A5A2021242).

Conflicts of interest: No potential conflict of interest relevant to this article was reported.

Author contributions: All work was done by Sangzin Ahn.

References

1. Prompt examples from OpenAI API documentation. <https://beta.openai.com/examples>. Accessed December 30, 2022.

2. Wei J, Tay Y, Bommasani R, et al. Emergent abilities of large language models. arXiv [Preprint]. 2022 Oct 26. <https://doi.org/10.48550/arXiv.2206.07682>
3. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. arXiv [Preprint]. 2022 Dec 26. <https://doi.org/10.48550/arXiv.2212.13138>
4. Hutson M. Could AI help you to write your next paper? *Nature*. 2022;611(7934):192-193.
5. Arcas BA. Do large language models understand us? *Daedalus*. 2022;151(2):183-197.
6. Wei J, Wang X, Schuurmans D, et al. Chain of thought prompting elicits reasoning in large language models. arXiv [Preprint]. 2022 Jan 28. <https://doi.org/10.48550/arXiv.2201.11903>
7. Liévin V, Hother CE, Winther O. Can large language models reason about medical questions? arXiv [Preprint]. 2022 Jul 17. <https://doi.org/10.48550/arXiv.2207.08143>
8. Wu T, Terry M, Cai CJ. AI chains: transparent and controllable human-ai interaction by chaining large language model prompts. Paper presented at: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; April 29, 2022; New Orleans, USA. <https://doi.org/10.1145/3491102.3517582>
9. Hase P, Diab M, Celikyilmaz A, et al. Do language models have beliefs? methods for detecting, updating, and visualizing model beliefs. arXiv [Preprint]. 2021 Nov 26. <https://doi.org/10.48550/arXiv.2111.13654>
10. Wang B, Xie Q, Pei J, Tiwari P, Li Z. Pre-trained language models in biomedical domain: a systematic survey. arXiv [Preprint]. 2021 Oct 11. <https://doi.org/10.48550/arXiv.2110.05006>
11. Sternszus R, Saroyan A, Steinert Y. Describing medical student curiosity across a four year curriculum: an exploratory study. *Med Teach*. 2017;39(4):377-382.