Research in the Application of AI and LLM in Educational Field

Zhating Wang*

College of Mathematics, Tongji University, Shanghai, China

*Corresponding author: 2254253@tongji.com

Abstract. Recently, AI and large language models, represented by ChatGPT4, have developed rapidly, accomplishing various practical functions such as human-computer interaction and information integration. The large language model provides necessary technical support in the field of education. As a result, an increasing number of researchers are exploring the feasibility and prospects of applying AI and large language models in various fields. This paper analyzes the application of AI and large language models in education, including professional education such as art and mathematics, as well as general education and high school and social education. It analyzes the advantages, disadvantages, and deficiencies of AI in education by analyzing the commonalities of studies from different perspectives and proposes development directions for the future application of AI in education. The main purpose of this article is to analyze the current application of AI and large models in the field of education, while also providing convenience for future researchers in this field.

Keywords: Large language model; Educational application; Artificial intelligence.

1. Introduction

Large language models, a subset of artificial intelligence, are trained on vast datasets comprising billions of words from sources like articles, books, and online content. These models typically rely on neural network structures that utilize deep learning techniques, which have demonstrated remarkable success in fields such as medicine, to map the intricate web of connections between words within textual data. The training regimen for these models can be complex and multi-layered, sometimes requiring human intervention at various stages. As a result, these models acquire an understanding of how words interact within language, enabling them to employ this knowledge to perform a variety of natural language processing tasks.

Natural language processing (NLP) described a wide field of computational research, aiming to achieve the automatic analysis of language by imitating human's ability. The target of the generative AI is to produce models which can give users what they want and make use of the method it learned to solve the problems in natural language and do some prediction in content. -- in other words, 'natural language generation' tasks. With the development of AI, LLMs now can recognize, interpret, and generate text with minimal or no specific fine-tuning required.

ChatGPT (OpenAI) is a kind of large language model (LLM) based on deep learning and NLP, through which computers can understand and generate contents like human-beings. It's feature is that it can have a conversation-like interact with users and have a response according to the users' prompt in forms of answering questions, continuing storytelling, generating articles, providing information, and other forms of textual content. It performs well in understanding contents and information retrieval. So many researchers saw the potential of ChatGPT in the field of education. Their research suggests that given the strong capability in integrating information and powerful 'language skills', disciplines that require vast text libraries, especially in history, art, and language education, have great potential with ChatGPT.

The development of ChatGPT in 4 generations now delivered:

GPT-1: A text generation model based on Transformer, which adopted a pre-training mechanism, combining unsupervised study and supervised fine-tuning to have a better performance in language processing launched in 2018.

GPT-2: Compared to GPT-1, GPT-2 features a significantly larger model size with 1.5 billion parameters, which is ten times more than GPT-1, and it has been trained with more advanced

techniques, including increased use of unsupervised learning, enhancing its ability to understand and generate more nuanced and coherent text.

GPT-3: GPT-3 is a massive language model with over 175 billion parameters, making it significantly larger and more capable than its predecessors. It excels in a variety of NLP tasks, including few-shot and zero-shot learning, due to its vast scale and advanced training techniques.

GPT-4: GPT-4 operates at a much larger scale than GPT-3, with more data, parameters, and context, leading to enhanced knowledge and pattern recognition abilities. Additionally, GPT-4 supports visual inputs, marking a significant advancement in contextual reasoning compared to GPT-3, which is limited to text.

Some researchers combined LLM with subject-oriented education and made unique model to meet the needs of the subject. For example, LLaVA-Docent is a multimodal large language model (MLLM) designed for art appreciation education, combining visual and textual data to provide a personalized conversational partner that enhances students' deep understanding and appreciation of artworks. The model was developed through an iterative enhancement research method, utilizing a virtual dialogue dataset generated by GPT-4, and has shown potential in promoting student engagement and increasing accessibility in art education. However, LLaVA-Docent faces limitations in generating diverse questions and providing natural conversations, necessitating further expansion of the dataset and algorithm optimization to enhance its performance and educational value [1].

The paper will analyze the feasibility, implementation, and potential future improvements of the current application of large language models in the field of education.

2. Application in educational field

2.1. Applications in Specific Field and Professional Education

2.1.1 Art

In the field of integrating Large Language Models (LLMs) with art education, researchers have constructed the entire model using three components, which are the vision encoder g, the LLM $f\phi$, and the projection layer W. Firstly, the vision encoder g is specialized for capturing the visual modality. This algorithm accepts image data Xv as input and outputs Zv. Secondly, a trainable projection layer W is used to bridge the different modalities between images and language. The projection layer W projects the visual features Zv into visual embedding tokens Hv, which have the same dimension as the language embedding tokens Hq. (Given the equation $Hv = W \cdot Zv$) Thirdly, researchers select Vicuna (Chiang et al., 2023) as the LLM $f\varphi(\cdot)$, parameterized by φ. The input data for the LLM is the concatenation of Hv and Hq. The final formula is $Xa = f\,\varphi(Hv \oplus Hq)$, where $\oplus$ denotes concatenation. In the end, it achieved the result of helping students better understand the paintings from various perspectives and could provide personalized suggestions, effectively promoting the development of art education [1].

2.1.2 Math

In Mathematics education, researchers have developed a new LLM-driven process and application called ItemForge, specifically designed for the automatic generation of math assessment items. ItemForge is a Python-based web application that utilizes mathematical concepts, retrieved concept summaries, ideal student knowledge, and instructional descriptions of cognitive process and knowledge dimensions to generate assessment items. In the collection work of the text library, researchers retrieve and integrate mathematical concept information from higher education textbooks and course lectures through the LangChain framework and the Facebook AI Similarity Search (FAISS) library. The primary method involves using vector-based retrieval techniques to search for text pages similar to the selected mathematical concepts from a vector database to form the concepts and relationships.

In the end, researchers successfully developed an effective assessment system for mathematical education outcomes. This system can work with feedback from math experts on the language quality

of generated tasks and solutions, the coverage of target concepts, the clarity of task statements, and the mathematical quality of solutions. Based on this feedback, the system can form assessments that are discriminating in terms of proficiency levels [2].

2.2. Applications in universal education

Research explores the benefits and challenges of developing, deploying, and evaluating a large language model (LLM) chatbot, MoodleBot, in higher education settings. It highlights the potential of integrating LLMs into learning management systems (LMS) like Moodle to support self-regulated learning (SRL) and help-seeking behavior. The study found that LLM-based chatbots like MoodleBot can significantly enhance the teaching and learning process, with an 88% accuracy rate in providing course-related assistance, indicating their suitability for integration into courses to improve personalized learning and reduce teacher administrative burden [3].

At the same time, introducing AI into educational use can also provide assistance to teachers. In the research by Chen and Du on the current shortage of international Chinese teachers, it is pointed out that the Spark Model has a promising future in reducing the burden on both teachers and students. The large model simulates international Chinese teachers, providing a convenient and clear dialogue environment for Chinese learners, while systematically providing teachers with objective data on learners' progress, significantly improving the efficiency of teaching [4,5].

2.3. Applications in high school and social education

At present in Shanghai, China, AI has been implemented in the most important college entrance examination for Chinese students, the college-entrance examination. The English oral test in Shanghai has achieved human-computer interaction and a combination of machine and human scoring. This greatly reduces the workload while obtaining objective scores, and at the same time, it provides students with many convenient ways to practice and learn.

The large models have played a significant role in the following three most important parts of oral tests: Question Design, where they enable personalized, automated, and diverse question setting to improve the efficiency and quality of question formulation; Intelligent Grading and Marking, by supporting intelligent scoring for all subjects and question types, providing in-depth evaluation and multi-dimensional grading quality detection; and Intelligent Examination Management and Services, which enhance the work efficiency of examination affairs and empower the entire process of educational examination services. Collectively, these advancements provide a lot of students with an environment for self-practice and simulation, significantly enhancing their preparation and performance in oral tests [6].

Large language models are also put to good use in library management. By establishing a "Book Assistant," readers can receive precise book information and borrowing guides around the clock, while also achieving cross-language communication functions and reminders for book lending and return deadlines. Even more, ChatGPT-like large language models can become more adaptable conversational agents, and can remember the content of conversations and connect the logic of context, incorporating it into their responses, and engage in more natural and coherent conversations with users over time. Due to training on vast amounts of data, large models have the potential to provide personalized responses based on the context of the given prompt, and can generate responses using different tones and structures according to user preferences and needs. These capabilities will make users feel less like they are talking to a machine, thereby enhancing user engagement and participation [7].

3. Current deficiencies

Current AI is primarily focused on organizing and reasoning based on existing data, and it does not possess a strong capability to draw conclusions from the existing content. Therefore, it not only

requires a very large amount of data, that is, it has a high dependency on data, but also demands a high level of accuracy in the data.

At the same time, the well-known hallucination issue with large models persists. Although the occurrence of hallucinations is low after fine-tuning, there is still a possibility that they may occur. This means that beginners cannot truly detach from the traditional book-based learning and teacher-student system when using AI for learning [8].

On the other hand, the security of students' information is also a concerning issue. The "black box" nature of AI means that the collection and processing of students' data and information are not entirely transparent, which leads to concerns among students and parents. Even with the high efficiency and value that AI offers in helping learning, they still need to carefully consider whether to use AI as a supplementary learning tool [8].

4. Conclusion

AI has shown great performance in single disciplines such as art, mathematics, and language education, and it can also provide excellent assistance in students' daily learning. However, the imprecision of its results and privacy security issues remains factors that cause users to hesitate when considering whether to use AI for educational purposes.

In the future AI education field, reducing dependence on data is an important direction for research. This mainly involves enhancing the AI's ability to judge the authenticity of new content when processing existing information. From the perspective of the model, strengthening AI's reasoning capabilities, enhancing the model's adaptability, improving the model's interpretability, and optimizing the model's generalization abilities can all lead to better applications of AI education. From the perspective of data, improving data quality, including the introduction of domain knowledge, and at the same time strengthening user data privacy protection are also very important.

References

[1] Lee U, Jeon M, Lee Y, Byun G, Sone Y, Shin J, Ko H, & Kim H. LLaVA-docent: Instruction tuning with multimodal large language model to support art appreciation education. Computers and Education: Artificial Intelligence, 2024, 7, 100297.

[2] Meissner R, Pögelt A, Ihsberner K, Grüttmüller M, Tornack S, Thor A, Pengel N, Wollersheim H-W, & Hardt W. LLM-generated competence-based e-assessment items for higher education mathematics: methodology and evaluation. Front. Educ., 2024, 9, 1427502.

[3] Neumann A T, Yin Y, Sowe S, Decker S, & Jarke M. An LLM-Driven Chatbot in Higher Education for Databases and Information Systems. IEEE Transactions on Education, 2024, 1.

[4] Chen X, & Du Y. Adaptation and transformation of international Chinese language teachers in the era of smart education—A discussion based on the application of large language models. Journal of Yangzhou University (Higher Education Research Edition), 2024, 28(5), 41-50.

[5] Cui X L. The dilemma and countermeasures of written language expression for Chinese second language learners. World Chinese Language Teaching, 2024, 38 (04): 435-447.

[6] Wang Z L. Digital Transformation of Educational Examinations with Cognitive Intelligence Models. Journal of China Examinations, 1005-8427, 2023, 08-0011-8.

[7] Wei M, Luo K, Tang M, Zhu W, & Yin K. Research on the application of large language model (LLM) embedded in scenario-based services in smart libraries. Library Theory and Practice. 2024.

[8] Alqahtani T, Badreldin H A, Alrashed M, Alshaya A I, Alghamdi S S, Saleh K B, Alowais S A, Alshayab O A, Rahman I, Al Yami M S, & Albekairy A M. The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. Research in Social and Administrative Pharmacy, 2023, 19, 1236–1242.