Voice

LLMs as Promising Personalized Teaching Assistants: How Do They Ease Teaching Work?

Jiayi Liu (刘嘉怡**)** ⓘD
East China Normal University

Bo Jiang (江波**)**
East China Normal University

Yu'ang Wei (魏雨昂**)** ⓘD
East China Normal University

Keywords

Human–AI collaboration, large language model, personalized teaching

A variety of Large Language Models (LLMs)-based tools have been developed to automate teaching material generation, student assessment, and feedback provision. This demonstrates LLMs' great potential in making large-scale personalized teaching a reality. Personalized teaching at scale demands tailored learning materials, personalized instruction, and meaningful feedback. LLMs, such as ChatGPT, are promising tools for facilitating personalized instruction. With extensive text data, advanced algorithms, and computing power, LLMs possess the capability to assist in instruction. This essay reviews recent advancements in utilizing LLMs for instruction planning and assessment. Then, we conclude the approaches for LLMs to assist teaching work, as shown in Figure 1. Additionally, we discuss the complementary roles of teachers and LLMs and offer advice for human–AI collaboration.

Corresponding author:
Jiayi Liu, Department of Educational Information Technology, East China Normal University, 3663 North Zhongshan Road, Shanghai 200062, China.
Email: 2966755926@qq.com

Two ways for LLMs to support personalized teaching

LLMs help automate the generation of tailored teaching materials

Personalized instruction materials lay the foundation for effective pedagogical design. These materials include a wide range of educational resources, such as books, teaching aids, multimedia materials, and more. Across various subjects, the complexity and abundance of teaching materials necessitate machine assistance to alleviate the burden on educators. Consequently, recent research has focused on leveraging LLMs to create personalized teaching content, offering teachers well-structured resources tailored to their specific needs.

LLMs revolutionize the process of personalized teaching materials generation across diverse learning domains. In language learning, Lu et al. (2023) introduce a pioneering human–AI collaborative system for efficiently generating reading quiz questions. The system leverages several transformer-based NLP models to offer process-oriented support to instructors during reading quiz design. This includes an abstractive summarization model to compress long paragraphs, a paraphrase model to rewrite sentences, and a negation model to formulate distractive options. After evaluation, instructors found the resulting questions comparable to those designed by humans. The system simplifies the creation of reading quizzes, showcasing the integration of human expertise and AI capabilities. Similarly, in programming education, Sarsa et al. (2022) harness Codex, a specialized LLM developed by OpenAI, to automate the formulation of programming exercises and code explanations. By incorporating structured explanations aligned with SOLO taxonomies, this approach demonstrates the potential for LLMs to streamline the generation of high-quality exercises suitable for pedagogical use.
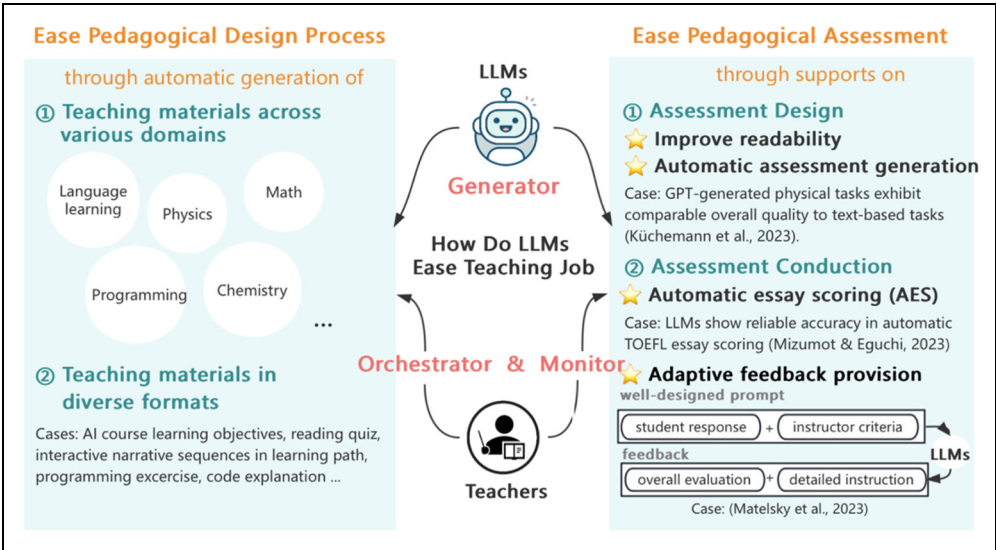

Figure 1. Applications and roles of LLMs in teaching work.

Moreover, the automation of pedagogical design is further exemplified by the LLMs' capacity to generate learning objectives. Sridhar et al. (2023) delve into GPT-4's ability to automatically generate learning objectives aligned with Bloom's taxonomy, offering an efficient approach to pedagogical planning. Additionally, generative AI is employed to create narrative fragments, which can be interspersed into learning paths to improve learner engagement (Diwan et al., 2023), showcasing LLMs' versatility in automating pedagogical design processes. These endeavors underscore the transformative potential of LLMs in automating the generation of diverse types of teaching materials, paving the way for more efficient and personalized educational experiences.

While LLM-generated content may require supervision and adjustment, it significantly alleviates teachers' workload during pedagogical preparation. As generative machine learning models continue to evolve, we can anticipate the emergence of sophisticated, ready-to-use teaching materials across various subjects in the near future.

Automated assessment and feedback provision via LLMs

In education, assessment and evaluation are crucial components for understanding student progress and forming instructional decisions. However, personalized assessments and targeted feedback provision for each student can be extremely time-consuming and energy-consuming for educators, making it hard to implement at scale.

Several researchers have taken trials to incorporate LLMs in the processes of assessment design and conduction, which achieved positive results. In terms of assessment design, a recent research by Küchemann et al. (2023) developed physics curriculum tasks using GPT-3.5, and compared the quality of GPT-generated tasks against that derived from traditional textbooks. Surprisingly, GPT-generated tasks exhibit comparable correctness, adequate difficulty levels, and overall quality to text-based tasks, showcasing LLMs' proficiency in synthesizing domain knowledge for assessment formulation. However, shortcomings such as a lack of specificity and context in GPT-formulated tasks were noted. In terms of automated evaluation, recent research found LLMs' capacity in automatic scoring in the language learning domain. For instance, some researchers (Mizumoto & Eguchi, 2023) found GPT-3's reliable accuracy in automatic scoring essays from TOEFL, presenting LLMs' potential to supplement human evaluations.

Besides, LLMs can assist teachers by providing tailored feedback during the pedagogical process. By feeding well-organized prompts to LLMs, teachers can receive tailored feedback for students. For example, Matelsky et al. (2023) introduced a framework for open-ended questions and transformed student responses along with instructor criteria to an LLM. They're surprised to find that the LLM managed to give overall and detailed feedback, offering both feedback on the entire answers as a whole and feedback on specific parts that may be erroneous (Matelsky et al., 2023).

However, despite the promise of LLM-generated feedback, its impact on student performance remains an open question. Notably, students' ability to comprehend and utilize feedback can be

varied by their learning ability (Zhu et al., 2020). The provision of personalized feedback doesn't directly lead to improved learning performance. In the future, more empirical research is needed to investigate ways of efficiently integrating Generative Artificial Intelligence (GAI)-generated feedback into pedagogical practice. Problems surrounding learners' capacity and willingness to comprehend and utilize GAI-generated feedback can be further explored.

Advice for teacher–LLM collaboration

Teachers take the role of orchestrator and LLMs as the material generator

In the teacher–LLM collaboration, teachers can play the role of orchestrators, formulating diverse and high-quality pedagogic designs. By allocating time-consuming tasks to LLMs, teachers are allowed to focus on learning activities refinement. Despite various teaching materials generated by LLMs, they wouldn't be helpful to students without adaption to learners' needs. Thus, teachers need to take a deeper dive into students' needs and appropriately organize various teaching resources generated by LLMs, ensuring seamless integration of these resources into instruction practice.

Teachers monitor and supervise the performance of LLMs

In the evaluation process, teachers can take the roles of monitor and assessment designer. The ability of LLMs to generate thorough and complicated evaluations remains limited, making it necessary to be supervised before being introduced into the classroom. Meanwhile, though some researches show LLMs' potential in automated evaluations, current studies mainly focus on specific domains and lack an integrated perspective. Teachers need to incorporate a variety of assessment methods to conduct a well-rounded evaluation of students' abilities. This may include introducing formative and summative assessment into evaluation and organizing group projects, hands-on activities, and oral presentations. Despite AI assistance, teachers still need to actively participate in the assessment process to guarantee the validity of evaluations.

Contributorship

Jiayi Liu was responsible for essay writing and contributed by reviewing the literature to determine the approaches for LLMs to assist teachers in personalized teaching. Bo Jiang determined the essay's topic and overall structure and was responsible for supervising the essay's progress. Yu'ang Wei contributed by reviewing the literature to provide an overview of the methods to incorporate LLMs in personalized teaching.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iDs

Jiayi Liu 🆔 https://orcid.org/0009-0007-9927-2653
Yu'ang Wei 🆔 https://orcid.org/0000-0002-8187-4011

References

Diwan, C., Srinivasa, S., Suri, G., Agarwal, S., & Ram, P. (2023). AI-based learning content generation and learning pathway augmentation to increase learner engagement. *Computers and Education: Artificial Intelligence*, *4*, 100–110. https://doi.org/10.1016/j.caeai.2022.100110

Küchemann, S., Steinert, S., Revenga, N., Schweinberger, M., Dinc, Y., Avila, K. E., & Kuhn, J. (2023). *Physics task development of prospective physics teachers using ChatGPT*. ArXiv Preprint ArXiv:2304.10014. https://doi.org/10.48550/arXiv.2304.10014

Lu, X., Fan, S., Houghton, J., Wang, L., & Wang, X. (2023, April 19). ReadingQuizMaker: A human-NLP collaborative system that supports instructors to design high-quality reading quiz questions. CHI'23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Article No. 454, pp. 1–18. https://doi.org/10.1145/3544548.3580957

Matelsky, J. K., Parodi, F., Liu, T., Lange, R. D., & Kording, K. P. (2023). *A large language model-assisted education tool to provide feedback on open-ended responses*. ArXiv Preprint ArXiv:2308.02439. https://doi.org/10.48550/arXiv.2308.02439

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, *2*(2), Article 100050. https://doi.org/10.1016/j.rmal.2023.100050

Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022). Automatic generation of programming exercises and code explanations using large language models. ICER'22: Proceedings of the 2022 ACM Conference on International Computing Education Research—Volume 1, pp. 27–43. https://doi.org/10.1145/3501385.35439

Sridhar, P., Doyle, A., Agarwal, A., Bogart, C., Savelka, J., & Sakr, M. (2023). *Harnessing LLMs in curricular design: Using GPT-4 to support authoring of learning objectives*. ArXiv Preprint ArXiv:2201.11227. https://doi.org/10.48550/arXiv.2306.17459

Zhu, M., Liu, O. L., & Lee, H.-S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education*, *143*, Article 103668. https://doi.org/10.1016/j.compedu.2019.103668