Empowering Socratic Teaching Through Large Language Models: New Opportunities and Challenges in Intelligent Tutoring Systems

Yueru Lang 1*, Siyang Liu 1, Xiangen Hu 2*, Shaoying Gong 1
1 School of Psychology, Central China Normal University, Wuhan, China
2 Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong
*langyueru_michelle@163.com; xiangen.hu@polyu.edu.hk

Abstract: *This paper explores the necessity and potential of empowering Socratic teaching by large language models in intelligent tutoring systems to enable adaptive conversations and facilitate deeper engagement. Using Socratic Playground for Learning as an example, this paper presents the system's architecture and functionality of two components: lesson creation and Socratic interaction. A pilot study was conducted with 30 participants over five days, revealing that SPL significantly improved their performance and was well-accepted by users. In addition, the paper highlights the key opportunities and challenges that need to be addressed for effective implementation.*
Keywords: Socratic teaching, intelligent tutoring system, large language model, opportunity, challenge

1. Introduction

Socratic teaching has long been valued for its educational benefits in cultivating deeper thinking (Elkowitz, 2021). However, applying face-to-face Socratic teaching has been challenging due to the high skill requirement for instructors and the difficulty sustaining meaningful interactions in classroom teaching (Dalim et al., 2022; Hsu et al., 2022). Recent advancements in large language models (LLMs) present transformative opportunities to enhance the abilities of intelligent tutoring systems (ITSs) to conduct Socratic teaching beyond traditional classes.

LLMs can enhance the Socratic teaching by allowing for more dynamic and personalized interactions with lower development costs. However, relying solely on LLMs carries the risk of over-dependence and may hinder critical thinking (Premkumar et al., 2024). Therefore, incorporating domain-specific knowledge models, learner models, and pedagogical strategies within ITSs can refine LLM's responses to ensure they are contextually appropriate and align with learners' needs and learning goals. Based on Socratic principles of encouraging inquiry rather than providing direct answers, large language model-based intelligent tutoring systems (LLM-ITSs) can be steered towards probing questions, challenging assumptions, and clarifying ideas, thereby enhancing cognitive engagement and active learning.

2. LLM-ITSs: Exemplified by Socratic Playground for Learning

Inspired by the potential of delivering Socratic teaching by LLM-ITSs, this study introduces Socratic Playground for Learning (SPL) — an ITS powered by ChatGPT. SPL is structured into two main components: lesson creation and Socratic interaction. These two components work together to create a highly personalized, customized, and adaptive learning scenario for learners. Figure 1 shows the system architecture of SPL.

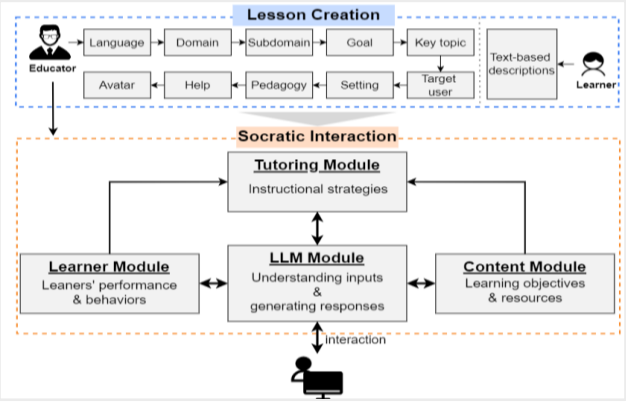

Fig.1 The system architecture of SPL

SPL empowers educators and teachers to customize learning scenarios through its lesson creation component. Educators can set key parameters from a tree-structured framework. Based on the selected parameters, ChatGPT generates a preliminary interactive interface for the lesson. At this stage, educators have the opportunity to further fine-tune the configuration. Besides, SPL goes beyond traditional teacher-driven lesson creation by allowing learners to create a lesson through text-based descriptions.

The Socratic interaction component is where learners actively engage with SPL. SPL initiates the interaction by posing a provoking Socratic question to prompt critical thinking. After receiving learners' inputs, SPL adapts the dialogue that mirrors traditional Socratic teaching to support ongoing engagement and deeper exploration of the topic. SPL can also summarize learners' responses, provide feedback on strengths and weaknesses, give hints when the learner struggles, offer encouragement to avoid frustration and disengagement, and propose follow-up questions that foster a reflective and iterative learning process. If learners need additional support, they can request on-demand help in challenging tasks without feeling overwhelmed.

3. A Pilot Study: Examining the capabilities of SPL

This study used a single-group pretest-posttest design to investigate the effects of SPL on learning performance and explore user acceptance toward SPL. A total of 30 Chinese college students were included in the pilot study. All participants were 18-23 years old (M = 20.43, SD = 0.43).

This study selected 5 typical flaws in experiments from the ARIES program as the learning content (Myers, 2021). The knowledge pretest and posttest were homogeneous transfer tests, consisting of 15 multiple-choice items (15 points) and two case studies (15 points), respectively. The Unified Theory of Acceptance and Use of Technology (UTAUT2) scale developed by Venkatesh et al. (2012) and revised by Zeng (2019) was used to assess the technology acceptance of SPL. Learners were required to score 23 items on the 7-point Likert scale ranging from "strongly disagree" to "strongly agree".

Prior to the learning phase, participants were recruited, signed the informed consent, and completed the prior knowledge test. The learning phase lasted for five days. Participants focused on one specific flaw each day. On average, participants spent 170 minutes learning in SPL. On the day following the learning phase, participants sequentially completed the knowledge posttest and the UTAUT2 scale.

Results of the paired-sample t-test indicated that the knowledge posttest test score was higher than the knowledge pretest score ($t = 7.57$, $p < 0.001$). The proportional learning gain ($= \frac{posttest\ ratio - pretest\ ratio}{1 - pretest\ ratio}$) was calculated at 29.23%. These findings indicate a substantial improvement in learners' knowledge after interacting with SPL.

Results of user acceptance showed that learners generally had favorable attitudes toward SPL, with moderate to high levels of all dimensions of the UTAUT2 scale ($M > 4.50$). Key factors, including performance expectancy ($r = 0.59$), effort expectancy ($r = 0.52$), social influence ($r = 0.75$), hedonic motivation ($r = 0.77$), and habit ($r = 0.81$), were

all significantly and positively associated with intention to use SPL. Learners not only perceived the SPL as beneficial for their learning but also found that using SPL was free of effort, and they were influenced by social environments to adopt this new tool. Learners could derive enjoyment from using SPL, and the significant correlation with habit also supports the potential that SPL could become an ingrained part of lifelong learning. These positive perceptions and strong acceptance highlight the potential of SPL as a valuable tool with sustained engagement.

4. Opportunities and Challenges of LLM-ITSs

LLM-ITSs can foster the scalability and accessibility of personalized Socratic teaching, have the ability to maintain learning engagement and motivation, enable learning as a service and lifelong learning, promote cross-cultural understanding, and support cultural heritage preservation.

While LLM-ITSs present considerable opportunities to enhance education, their implementation also faces challenges that must be noticed and addressed. Key challenges include technological limitations, user readiness and acceptance, assessment transparency and rationality, ethical concerns, and adaptability to different cultures. In addition, Future research should employ experimental designs with rigorous controls to provide a clearer picture of how LLM-ST influences learning.

References

Dalim, S. F., Ishak, A. S., & Hamzah, L. M. (2022). Promoting students' critical thinking through Socratic method: Views and challenges. Asian Journal of University Education, 18(4), 1034-1047. https://doi.org/10.24191/ajue.v18i4.20012

Elkowitz, D. E. (2021). Socratic questioning to engage learners. In A. Fornari & A. Poznanski (Eds.), *How-to guide for active learning* (pp. 107–119). Springer.

Hsu, F., Lin, I., Yeh, H., & Chen, N. (2022). Effect of Socratic Reflection Prompts via video-based learning system on elementary school students' critical thinking skills. *Computers & Education, 183*, Article 104497. https://doi.org/10.1016/j.compedu.2022.104497

Myers, M. H. (2021). Automatic Detection of a Student's Affective States for Intelligent Teaching Systems. *Brain Sciences, 11*, 331. https://doi.org/10.3390/brainsci11030331

Premkumar, P. P., Yatigammana, M. R. K. N., & Kannangara, S. (2024). Impact of Generative AI on Critical Thinking Skills in Undergraduates: A Systematic Review. *Journal of Desk Research Review and Analysis*, *2*(1), 199–215. https://doi.org/10.4038/jdrra.v2i1.55

Venkatesh, T., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and Use of information technology: Extending the unified theory of acceptance and Use of technology. *MIS Quarterly, 36*(1), 157. https://doi.org/10.2307/41410412

Zeng, J. (2019). A study on user acceptance of mobile-assisted language learning based on UTAUT2 [Chinese]. *Technology Enhanced Foreign Language Education,* (6), 16-24. https://doi.org/10.20139/j.issn.1001-5795.2019.06.003