

Developing Effective Educational Chatbots with ChatGPT prompts: Insights from Preliminary Tests in a Case Study on Social Media Literacy

Cansu KOYUTURK^{a*}, Mona YAVARI^{a*}, Emily THEOPHILOU^b, Sathya BURSIC^a, Gregor DONABAUER^{a,c}, Alessia TELARI^a, Alessia TESTA^a, Raffaele BOIANO^d, Alessandro GABBIADINI^a, Davinia HERNANDEZ-LEO^{b*}, Martin RUSKOV^{e**} & Dimitri OGNIBENE^{a,***}

^aUniversità of Milano Bicocca, Milan, Italy ^bUniversitat Pompeu Fabra, Barcelona, Spain

^cUniversity of Regensburg, Regensburg, Germany ^dPolitecnico di Milano, Milan, Italy

^eUniversità of Milano, Milan, Italy ^fUniversity of Essex, Colchester, United Kingdom

*davinia.hernandez-leo@upf.edu, **martin.ruskov@unimi.it, ***Lead PI dimitri.ognibene@unimib.it
+Same contribution

Abstract: Educational chatbots come with a promise of interactive and personalized learning experiences, yet their development has been limited by the restricted free interaction capabilities of available platforms and the difficulty of encoding knowledge in a suitable format. Recent advances in language learning models with zero-shot learning capabilities, such as ChatGPT, suggest a new possibility for developing educational chatbots using a prompt-based approach. We present a case study with a simple system that enables mixed-turn interactions and discuss the insights and preliminary guidelines obtained from initial tests. We examine ChatGPT's ability to pursue natural educational conversations, adapt the educational activity to users' characteristics, such as culture, age, and level of education, and its ability to use diverse educational strategies and conversational styles. Although the results are encouraging, challenges are posed by the highly structured form of responses by ChatGPT, as well as their variability, which can lead to an unexpected switch of the chatbot's role from a teacher to a therapist. We provide some initial guidelines to address these issues and to facilitate the development of effective educational chatbots.

Keywords: ChatGPT, Educational chatbots, Educational prompts

1. Introduction

Educational chatbots have gained increasing attention in recent years for their potential to provide interactive and personalized educational activities with reduced teacher intervention. However, their development has been mostly focused on chatbot-driven conversation flow due to the complexity of mixed-initiative and user-driven approaches, the restricted interaction capabilities of available platforms, the lack of adequate training sets (Pérez et al., 2020), and the substantial efforts required to encode the necessary knowledge in a suitable format, notwithstanding the wide availability of topic-specific information.

Recent advances in Large Language Model (LLM) systems with their zero-shot learning capabilities (Wei et al., 2021), also named prompt-based approaches (Xu et al., 2023), to perform diverse tasks in different domains without training or fine-tuning suggest a new way to develop educational chatbots. Chatbots based on Reinforcement Learning from Human Feedback (RLHF), such as ChatGPT, have been extended to achieve more reliable and safe interactions (Ouyang et al., 2022)

While showing big promises, LLM-based chatbots and their potential are not yet fully understood (Mahowald et al., 2023; Borji, 2023) and present different sources of vulnerability (Xu et al. 2023). The skills necessary for natural educational conversations, such as playing a specific role, maintaining a conversation style (Lu et al., 2023), and assessing students' conditions (Ullman, 2023), are still challenging even for state-of-the-art systems.

We propose an investigation framework to study ChatGPT's ability to produce mixed-turn educational conversations where the students can steer the conversation toward desired topics as well as cover multiple related learning objectives, tailor educational activities to user characteristics such as culture and age and use different educational strategies and conversational styles. Building on Jeon and Lee (2023), in this paper, we provide suggestions for possible teacher orchestration and interrogation strategies and study whether they lead to successful interlocution and content provision. We explore different prompting strategies to test LLM capabilities in educational conversations with a particular focus on reliability, interactivity, and coverage. We report results from tests performed by the authors.

2. Methodology

To frame interactive educational chats with ChatGPT, we focused on the "space of learning" (Marton and Tsui, 2004). We explored the educational exchange during lessons, especially the challenge of imperfect communication between teachers and students. Adding the chatbot as a teacher mediator (Fig. 1), we interpreted the learning interaction using an adaptation of Laurillard's Conversational Framework (2009), considering the distinctions between conceptions and expressions for each role (teacher, student, and chatbot-mediator).

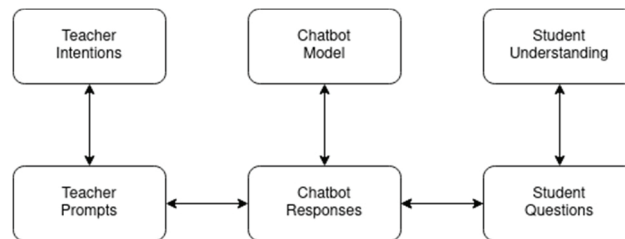


Figure 1. The educational framework: The teacher's intention to convey learning objectives is expressed through generated prompts in the technical framework. The chatbot's stochastic model determines the communication directed to the student. As the student interprets the exchanged messages, their understanding takes shape. Consequently, the actual experience of the messages differs from the teacher's intention, the chatbot's representation, and the student's comprehension.

To establish a learning space and facilitate a directed educational exchange we have formulated six blocks of educational objectives to be integrated into the prompt. Each block targets a critical area of social media literacy (Ognibene et al., 2023; Theophilou et al., 2023). The assessment of the interaction quality between ChatGPT and users was evaluated based on four dimensions: correctness, consistency, conversation fluency, and semantic coverage.

3. Results and Conclusions

We report: I. the insights that emerged during our search for prompts that resulted in effective educational conversations, and II. a classification of the issues we encountered. For details see Koyuturk et al. (2023). Our study showed that ChatGPT with the correct prompt could play the role of the teacher in short interactive educational conversations covering multiple learning objectives related to social media literacy. We found that it is useful and functional to explain in the prompt every foreseeable aspect of the interaction one by one. We aimed to clearly define the structure of interaction and the role of the chatbot. By testing with eight personas with different cultural backgrounds and ages, we verified that ChatGPT could also adapt to students' characteristics. However, relatively small variations of the prompt lead to a range of issues (see table 1).

Table 1. *Classification of observed issues in failed natural educational interactions.*

1. Unnatural Response Format 1. Bullet style 2. Essay Style 2. Interactivity 1. No questions 2. Survey style question presentation 3. Fixed sequence 3. Limited coverage	4. Unreliable behaviors 1. Behavior switching 1. Therapist Eliza style 2. Encyclopedic Q/A 3. Programmer assistant 2. Personal questions 3. Introduction of code-like variable placeholder 4. Imaginary students' answers
---	---

ChatGPT appears to be fine-tuned for question-answering, resulting in responses that may not align with the desired style or exhibit essential conversational elements. Initial prompt attempts resulted in unnatural response formats. Despite instructions to adopt a conversational style and avoid bullet points or essay formats, the chatbot frequently deviated. Turn-taking and questions were either lacking or excessive and the ability to cover multiple learning objectives and related topics was limited. Conversations often continued only when users asked relevant questions about previously unaddressed topics.

Interesting erratic and undesirable behaviors were observed. Requests to make the educational conversation interactive sometimes resulted in the failure to effectively maintain the teacher role. One of the most surprising and frequent behaviors was the tendency to autonomously cover both sides of the educational conversation. Our work shows that studying and extending the LLM's capabilities to consistently play roles is an essential line of research.

Acknowledgements

Spanish AEI (PID2020-112584RB-C33, MDM-2015-0502), Volkswagen S. (ref. 95 566 and 9B145) and Catalan SGR (00930). DHL (Serra Hünter) acknowledges the support of ICREA Academia.

References

- Borji, A. (2023). A categorical archive of ChatGPT failures. arXiv:2302.03494.
- Jeon, J., Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. <https://doi.org/10.1007/s10639-023-11834-1>
- Koyuturk, C., ... & Ognibene, D. (2023). Developing Effective Educational Chatbots with ChatGPT prompts: Insights from Preliminary Tests in a Case Study on Social Media Literacy (with appendix). arXiv:2306.10645
- Laurillard, D. (2009). The pedagogical challenges to collaborative technologies. *Int. J. of CSCL*, 4, 5-20.
- Lu, A., Zhang, H., Zhang, Y., Wang, X., & Yang, D. (2023). Bounding the Capabilities of Large Language Models in Open Text Generation with Prompt Constraints. arXiv:2302.09185.
- Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2023). Dissociating language and thought in large language models: A cognitive perspective. arXiv:2301.06627.
- Marton, F., Tsui, A. B., Chik, P. P., Ko, P. Y., & Lo, M. L. (2004). *Classroom discourse and the space of learning*. Routledge.
- Ognibene, D., ... & Eimler, S. (2023). Challenging social media threats using collective well-being-aware recommendation algorithms and an educational virtual companion. *Frontiers in AI*, 5.
- Ouyang, L., ... Schulman, J. (2022). Training language models to follow instructions with human feedback. In *Adv. in Neural Inf. Proc. Sys.* 35, 27730-27744.
- Pérez, J.Q., Daradoumis, T., & Puig, J.M.M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Comp. Appl. in Eng. Edu.*, 28(6), 1549–1565.
- Theophilou, E., ... & Eimler, S. (2023). Empirically Investigating Virtual Learning Companions to Enhance Social Media Literacy. 345–360. *HelMETO* https://doi.org/10.1007/978-3-031-29800-4_27
- Ullman, T (2023). Large language models fail on trivial alteration to theory-of-mind task. arXiv:2302.08399
- Wei, J., ... & Le, Q. V. (2021). Finetuned language models are zero-shot learners. arXiv:2109.01652.
- Xu, L., Chen, Y., Cui, G., Gao, H., & Liu, Z. (2022). Exploring the universal vulnerability of prompt-based learning paradigm. arXiv:2204.05239

CSCl