Be My Mate: Simulating Virtual Students for collaboration using Large Language Models

Sergi Solera-Monforte
Pablo Arnau-González
Miguel Arevalillo-Herráez
Universitat de València
{sergi.solera, pablo.arnau, miguel.arevalillo}@uv.es

Abstract

Advancements in machine learning, particularly Large Language Models (LLMs), offer new opportunities for enhancing education through personalized assistance. We introduce "Be My Mate," an agent that leverages LLMs to simulate virtual peer students in online collaborative education. The system includes a subscription module for real-time updates and a conversational module for generating supportive interactions. Key challenges include creating temporally realistic interactions and credible error generation. The initial demonstration shows promise in enhancing student engagement and learning outcomes.

1 Introduction

In recent years, advancements in machine learning have impacted the educational landscape, particularly with the emergence of Large Language Models (LLMs). These technologies offer new opportunities to enhance the learning experience and provide personalised assistance. However, the possibility of using LLMs to create a collaborative learning environment has not been explored. Currently, LLMs have passed the Turing test and it is difficult to tell LLM-generated text content apart from human-generated text. This, along with the emergent abilities in LLMs (Chang et al., 2024), shows the possibility of having LLM impersonate a peer student.

In the context of online collaborative education, it is often difficult to pair students with peers, or obtain a large enough pool of students to group individuals. For this reason, we propose the use of LLMs disguised as students to generate interaction with other students as their peers. This would enable benefits associated with collaborative learning, without the difficulties of pairing students.

However, the implementation of these systems is not straightforward. We offer a first demo of Be My Mate[1], an initial attempt to simulate virtual students for collaboration using large language models. By leveraging the capabilities of LLMs, we aim to create virtual entities that can interact with human students, participate in group discussions, and contribute to collaborative projects.

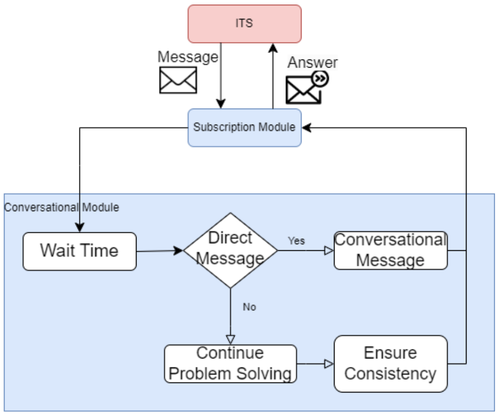

Figure 1: Processing Message Schema

2 Agent Implementation

We start from (Arevalillo-Herráez and Arnau, 2013), an existing system that supports the collaborative resolution of math word problems, while providing active feedback. The system utilises a WebSocket-based queue notification protocol, STOMP, for publishing updates about the state of the solution and the conversation of a given solution instance (room).

The proposed agent must subscribe to these updates to receive the current status and produce relevant contributions to the problem-solving. In order to do so, we split the system in two main modules, as shown in Figure 1:

Subscription module: Establishes the connection with the Intelligent Tutoring System (ITS) and subscribes for updates on the generated rooms. Every time a new room is created, the module will

[1] https://github.com/SPAM-research/BeMyMate

Proceedings of the 17th International Natural Language Generation Conference: System Demonstrations, pages 1–3
September 23–27, 2024.

listen to the user's inputs and modifications on the solution status (i.e. notes or equations) to propose new solutions or a conversational message to the student.

Conversational module: Its main goal is to provide relevant messages for supporting the student. The module has to decide whether to send a message to the student or propose a new letter, or equation. Consistency with the current context is ensured with two components that prepare, or discard, the output of the model to conform with the expected outputs in our system. In this direction, we check if the user has directly spoken to his peer (the agent), and therefore will be expecting a response, or, on the contrary, the system can freely propose new steps, such as letter assignation or equation definition. Alternatively, in case of a message to advance the resolution status, we must ensure the consistency of such message in terms of the expected message by the ITS. These validations aim to ensure the messages are credible. For example, ensuring that the provided equations are correctly structured for a mathematical context or that the letters used in the proposed equations by the agent have been previously defined. This validation must not prevent the agent from failing, but from making mistakes that may not be credible.

3 Technical and scientific challenges

During the analysis phase, we have identified temporisation as a main factor that must be addressed to convince alumni that they are interacting with peers. In the context of our ITS, humans tend to take more than a couple of seconds to write a chat message or propose a new definition. The time taken to write a message is influenced by a variety of factors, such as how advanced the problem is (at the beginning of a problem, humans try to understand it before writing the first message), familiarity with the system, age, or others. This time should be modelled and depends on the specifics of the system to be integrated in. At this stage and after manually inspecting a variety of values, we decided to include a uniform random wait time between 2 and 6 seconds for generating a message to the user, and between 5 and 10 seconds for generating the next step.

Another relevant factor that has been found relevant, is the type of mistakes the LLM can produce. In our context (an arithmetic, and algebra ITS) human mistakes are often related to the identification of relations between the problem quantities. The impersonated student should commit mistakes as if it were a regular student. This would create a perception of interacting with a peer. However, the errors should be credible. Moreover, although our system is capable of sending direct messages to the students, the agent should be able to decide to directly talk to the student to explain a concept or ask questions, as a peer would do.

Finally, regarding data privacy and collection, taking into account the sensitivity of educational-related data, authors should only capture the required samples to evaluate the system while preserving the student's anonymity. To this end, a new challenge arises for deciding if a student utterance is safe or not (directly leaks any personal data), correct it if possible, or directly reject it. Moreover, and appropiate storage scheme must be implemented to ensure that only individuals with research access can access the confidential data.

4 Conclusions

The demonstrated work[2] is a proof of concept of what we are aiming for in the Be My Mate agent. Although some aspects of our demo require further refinement, it demonstrates that LLM-based agents are a feasible option for providing collaborative learning environments. While significant progress must be made before agents can truly emulate humans and pass the Turing Test -such as timing the agent's interventions to enable a realistic conversational flow-, this demo highlights the potential of using LLMs to exploit the advantages of collaborative learning without the need for multiple students to be connected simultaneously. This first stone opens many research opportunities such as is addressing the bias these models might contain in the educational context. Moreover, the usefulness of these models, although theoretically sound must be empirically validated, either by studying the agent-student interaction by experts, in terms of learning compared to a Control Group, or by directly surveying the students with relevant tests.

In addition, the use of LLMs replacing real students can provide further benefits in this type of setting. In particular, it allows the system to rely on a single student model to make decisions, such as selecting the next activity, determining the optimal level of scaffolding, and tailoring the content and

[2]Video Available at: https://mmedia.uv.es/fbs?cmd=view&name=be_my_pal_demo.mp4

system behavior to the learner's needs and preferences.

Acknowledgments

Research supported by project CIGE/2023/63 funded by Generalitat Valenciana; TED2021-129485B-C42, funded by MCIN/AEI/10.13039/501100011033 and the European Union "NextGenerationEU"/PRTR; and Generatitat Valenciana through grants CIAPOS/2022/163; ACIF/2021/439.

References

Miguel Arevalillo-Herráez and David Arnau. 2013. A hypergraph based framework for intelligent tutoring of algebraic reasoning. In *Artificial Intelligence in Education*, pages 512–521, Berlin, Heidelberg. Springer Berlin Heidelberg.

Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. 2024. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3):1–45.