Chatlang: A Two-Window Approach to Chatbots for Language Learning

Sheng Yu[1], Jared Coleman[1], and Bhaskar Krishnamachari[1]

[1]Viterbi School of Engineering, University of Southern California,
centaurusyu2022@gmail.com, {jaredcol,bkrishna}@usc.edu

September 21, 2023

Abstract

Role-playing is an effective tool in language education, allowing learners to practice a target language in realistic scenarios and receive immediate feedback from teachers, native speakers, and fellow learners. This method is less viable, however, outside of the classroom, where access to speakers is more limited. Recent advancements in large language models (LLMs) have the potential to change this by enhancing the quality and accessibility of chatbot-based role-playing systems. This paper introduces "Chatlang," a two-threaded LLM-powered chatbot-based role-playing system. Users interact with the system using two chat windows displayed side-by-side in a web interface. The first window is for role-playing, where the user and chatbot are expected to never break character and only use the target language. In the second window, the user can use their native language to ask questions about the conversation or target language and get context-relevant feedback on their mistakes. This design enables uninterrupted role-playing alongside instant feedback and support. We present the design of Chatlang and discuss potential avenues for future research.

1 Introduction

Role-playing has long been recognized as an effective method in language education, creating immersive environments that enable learners to practice their skills within context. This method can be used to more naturally reinforce language concepts that students are likely to need for interactions in realistic scenarios (i.e. going to the store, ordering food at a restaurant, small-talk with strangers, etc.). Empirical studies have consistently illustrated the efficacy of role-playing in enhancing student engagement and improving test scores [9]. Outside of the classroom, role-playing opportunities for students is constrained by individual access to native speakers, teachers, and fellow learners. Large language models like ChatGPT [5], which exhibit near-native fluency in multiple languages and demonstrate effectiveness as translators [2], grammar Correctors [1], and simulated Role-Players [6], may pave the way for on-demand, internet-based practice through role-play, accessible to students globally.

We present a two-threaded approach to language learning through role-playing with large language models. The first thread represents the role-playing conversation, where the user engages in a conversation with an AI chatbot in the target language. The second thread represents the user's conversation with a *Tutor* AI chatbot, which provides feedback on the user's language usage and answers user questions about the conversation in the first thread (e.g. vocabulary, translations, etc.) as well as the language in general. The two threads are displayed side-by-side as chat windows in a web application, allowing the user to seamlessly switch between them. The motivation for this design is to allow the user to engage in a role-playing conversation without interruption while also providing a dedicated space for the user to ask questions and receive feedback. Chatlang is open-source and available on GitHub [1]. We also run a bring-your-own-keys (OpenAI) deployment at chatlang.net[2].

[1] https://github.com/ANRGUSC/chatlang
[2] https://chatlang.net

The rest of this document is organized as follows. In Section 2, we discuss related work in the area of language learning and role-playing with large language models. Then in Section 3, we present the design of our application, including the user interface and the underlying architecture. We conclude in Section 4 with a discussion of future work and promising research directions.

2 Related Work

The integration of role-playing in educational settings, especially for language learning, has been the subject of extensive research. One domain that has explored this integration is the realm of Massively Multiplayer Online Role-Playing Games (MMORPGs) [7]. These platforms allow users to immerse themselves in scenarios where they can practice language skills in real-time with other players. Such games, however, are not tailored specifically for language learning.

With the advent of large language models, the potential for role-playing in a controlled, educational context has expanded. One study explored the possibility of role-playing with such models, although not with a primary focus on language learning [10]. The power of these models in simulating conversational partners opens up avenues for more immersive learning experiences. Li et al. [3] recently proposed a language-learning chatbot solution with features like user login, corrective feedback, and more. Their system lacks the unique tutor/teacher feature present in our solution, however, and does not allow user customization of the role-playing scenario. ChatBack is another related tool that was used to study the preferences of learners regarding corrective feedback [4]. The authors hypothesized that learners prefer feedback from a separate role rather than the conversation partner. This finding aligns with the motivation behind our two-threaded approach. Interestingly, they utilized grammar correction models to pinpoint errors, which might be less comprehensive (but possibly more focused and/or accurate) than the LLM-only approach we use. Another intriguing direction is the work on user-adaptive language learning chatbots that come with a curriculum [8]. These chatbots are restricted to using certain words and phrases to create a more guided learning experience.

In summary, while several studies and tools explore the integration of role-playing and chatbots in language learning, our solution, Chatlang, stands out with its two-threaded approach. By allowing users to engage in uninterrupted immersive role-playing while having access to immediate feedback and support, Chatlang presents a unique and promising direction in the field of language education.

3 Design

The user interface is a web app with two main pages. On the first page (Figure 1), the user configures the role-playing scenario and the language learning settings. In the current implementation, the user can

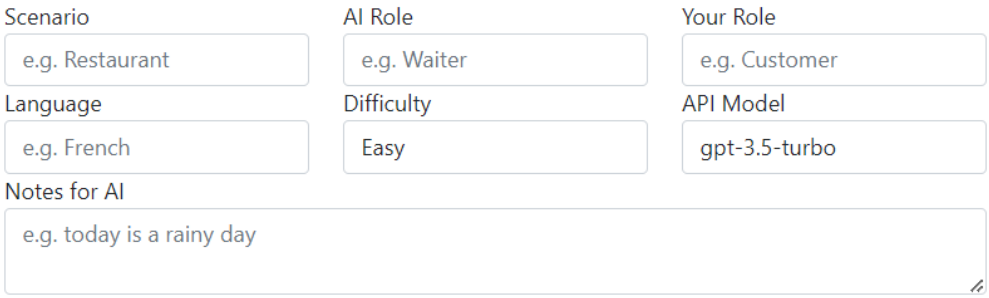

Figure 1: Main Page of the Web App with User Configuration Settings.

configure the settings shown in Table 1.

The second page (Figure 2) is the main chat page. On this page, there are two chat windows: the role-playing chat window (Thread 1) and the Tutor chat window (Thread 2). In the role-playing chat window, the user engages in a conversation with the AI chatbot in the target language (and ideally, should *only* use the target language). In the Tutor chat window, the user will recieve feedback on their language usage (in Thread 1) and can ask questions about the conversation (in Thread 1) or the language in general.

Table 1: User Configuration Settings for the Web App

Setting	Description
Scenario	The scenario the user wants to role-play (e.g., restaurant, hotel, etc.)
AI Role	The role the AI chatbot will play in the role-playing conversation (e.g., waiter, customer, etc.)
Your Role	The role the user will play in the role-playing conversation (e.g., waiter, customer, etc.)
Language	The language the user wants to learn. The user can enter anything in this field but should consider which languages the AI chatbot is capable of speaking.
Difficulty	The level of difficulty (of vocabulary, grammar, etc.) the user wants to practice. The AI chatbot will adjust its responses to match the user's difficulty level.
Model	The language model the AI chatbot will use to generate responses. The user can choose from a list of models that are available on the server.
Notes	Additional notes for the AI chatbot to consider when generating responses.

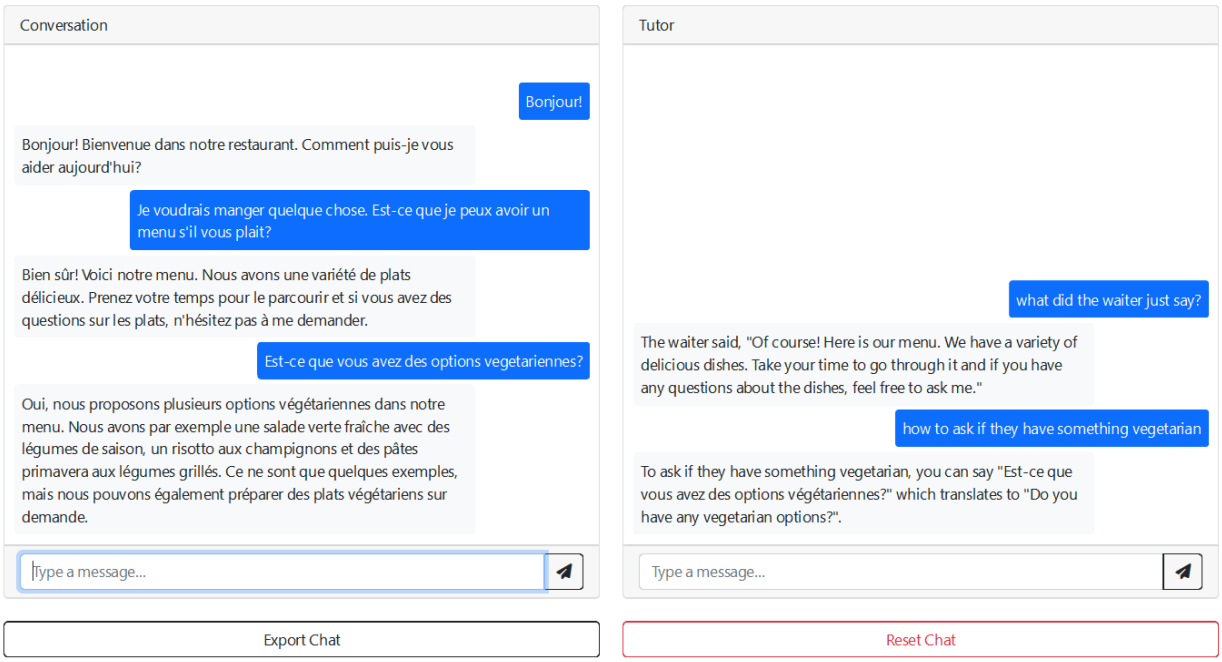

Figure 2: Main Chat Page of the Web App with Two Chat Windows.

From a technical perspective, there are three AI chatbots involved in the application: the Role-Player, the Tutor, and the Corrector. The role of each of these chatbots is depicted in Figure 3. The Role-Player chatbot responds to the user's messages in the role-playing chat window (Thread 1). It is seeded with a system prompt that includes the configuration settings from the main page (scenario, roles, language, difficulty, model, and notes) and prompts the AI chatbot to respond in the target language assuming the role specified by the user. For example, the system prompt for the conversation in Figure 2 is:

> You are an AI chatbot that will role-play with the user for them to practice their language skills. Your role is Waiter and the user's role is Customer. The scenario is Restaurant. The target

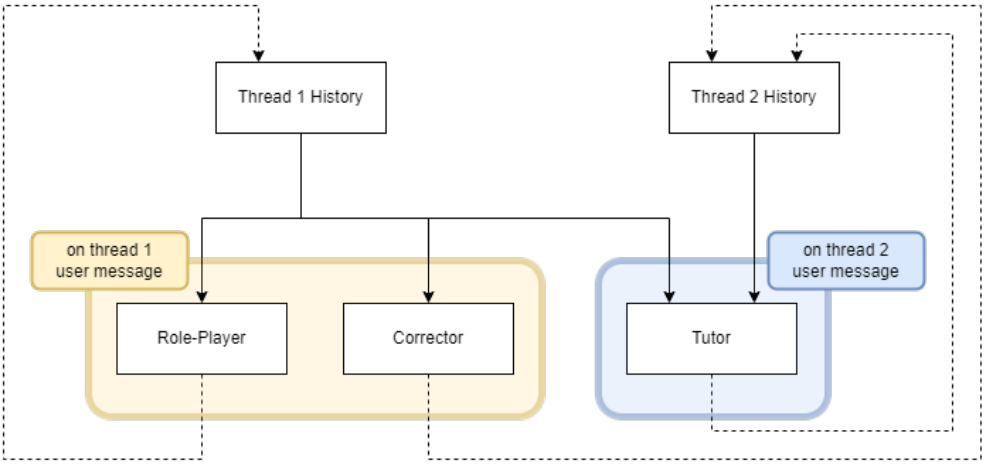

Figure 3: Architecture of Chatlang

> language is French. Do not use any other languages and do not break character. Use easy level language: *very* simple words/grammar and short one-sentence responses.

The entire prompt to the Role-Player chatbot is the concatenation of the system prompt and the user's messages in the role-playing chat window (truncated, if necessary, to fit within the maximum token length of the model).

The Tutor chatbot responds to the user's messages in the Tutor chat window (Thread 2). It is seeded with a system prompt that prompts the AI chatbot to respond in the user's native language, assuming the role of a Tutor monitoring the role-playing conversation. For example, the system prompt for the conversation in Figure 2 is:

> The user is role-playing with an AI chatbot to practice their French language skills. The user is playing the role of Customer and the AI chatbot is playing the role of Waiter. The scenario is Restaurant. You are a Tutor that is monitoring the AI chatbot and the user. When the user asks you a question, you should answer it in their native language English. The user may ask you questions about the conversation (i.e. what words/phrases mean), how to say something in the target language, etc.

The entire prompt to the Tutor chatbot is the concatenation of the system prompt and the user's messages in the *both* chat windows (truncated, if necessary, to fit within the maximum token length of the model).

The Corrector chatbot responds to the user's messages from the role-playing chat window (Thread 1) in the Tutor chat window (Thread 2). It is seeded with a system prompt that prompts the AI chatbot to respond in the user's native language, assuming the role of a Corrector monitoring the role-playing conversation. Also unlike the Role-Player and Tutor chatbots, the Corrector chatbot does not respond directly to user messages. Rather, on every user message sent in Thread 1, the Corrector chatbot is given the history (possibly truncated) of the user's messages in the role-playing chat window (Thread 1) and the Tutor chat window (Thread 2) and then asked to identify and correct any errors in last user message. For example, the entire prompt for the conversation in Figure 2 would be:

> The user is role-playing with an AI chatbot to practice their French language skills. The user is playing the role of Customer and the AI chatbot is playing the role of Waiter. The scenario is Restaurant. You are a Tutor that is monitoring the AI chatbot and the user.
>
> Tutor/User Conversation: [Thread 1 History]
>
> User conversation with AI assistant: [Thread 2 History]

If the user made a mistake in their last message, correct their mistake and give them advice (in their native language English) on how to improve. If they made no mistake, give no advice (empty string).

4 Conclusion

We have presented Chatlang, a two-threaded approach to language learning through role-playing with large language models. We have discussed the design of the application, including the user interface and the underlying architecture.

The Chatlang two-threaded approach to language learning through role-playing with large language models opens up several interesting research directions. A few research questions that merit further investigation are:

- **RQ1**: How does the two-threaded approach compare to a single-threaded approach in terms of language learning outcomes (e.g. proficiency, fluency, etc. as measured by a standardized test)?
- **RQ2**: How does the two-threaded approach compare to a single-threaded approach in terms of engagement (e.g. time spent in the application, length of conversation, user satisfaction, etc.)?
- **RQ3**: How does the two-threaded approach compare to a human role-playing partner/teacher in terms of language learning outcomes and engagement?
- **RQ4**: How does proactive error correction in Thread 1 (i.e., the role-player correcting the user's mistakes) compare to reactive error correction in Thread 2 (i.e., the tutor correcting the user's mistakes) in terms of language learning outcomes and engagement?

Finally, we note that the two-window approach of doing role-play dialog in one window and having access to a tutor in another window that we have presented could be useful for learning in other (non-language-related) settings as well. For example, such a system could be used to help someone role-play how to negotiate a salary increase with their employer or to practice peaceful communication and conflict resolution. Likewise, the tool could also be used by businesses for customer service training, to practice sales pitches, and much more. The ability to practice and ask questions and get feedback on one's own part in the dialog is thus potentially of great benefit for many inter-personal communication settings across family, social and business environments. We hope that the source code we provide will make it easy for practitioners to build such tools for other contexts.

References

[1] Tao Fang, Shu Yang, Kaixin Lan, Derek F. Wong, Jinpeng Hu, Lidia S. Chao, and Yue Zhang. Is chatgpt a highly fluent grammatical error correction system? A comprehensive evaluation. *CoRR*, abs/2304.01746, 2023. doi: 10.48550/arXiv.2304.01746. URL https://doi.org/10.48550/arXiv.2304.01746.

[2] Wenxiang Jiao, Wenxuan Wang, Jen-tse Huang, Xing Wang, and Zhaopeng Tu. Is chatgpt A good translator? A preliminary study. *CoRR*, abs/2301.08745, 2023. doi: 10.48550/arXiv.2301.08745. URL https://doi.org/10.48550/arXiv.2301.08745.

[3] Yu Li, Chun-Yen Chen, Dian Yu, Sam Davidson, Ryan Hou, Xun Yuan, Yinghua Tan, Derek Pham, and Zhou Yu. Using chatbots to teach languages. In René F. Kizilcec, Katie Davis, and Xavier Ochoa, editors, *L@S'22: Ninth ACM Conference on Learning @ Scale, New York City, NY, USA, June 1 - 3, 2022*, pages 451–455. ACM, 2022. doi: 10.1145/3491140.3528329. URL https://doi.org/10.1145/3491140.3528329.

[4] Kai-Hui Liang, Sam Davidson, Xun Yuan, Shehan Panditharatne, Chun-Yen Chen, Ryan Shea, Derek Pham, Yinghua Tan, Erik Voss, and Luke Fryer. ChatBack: Investigating methods of providing grammatical error feedback in a GUI-based language learning chatbot. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 83–99, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.bea-1.7. URL https://aclanthology.org/2023.bea-1.7.

[5] OpenAI. Chatgpt [large language model]. https://chat.openai.com/chat, 2023. Accessed: insert-date-here.

[6] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. *CoRR*, abs/2304.03442, 2023. doi: 10.48550/arXiv.2304.03442. URL https://doi.org/10.48550/arXiv.2304.03442.

[7] Mark Peterson. Massively multiplayer online role-playing games as arenas for second language learning. *Computer Assisted Language Learning*, 23(5):429–439, 2010. doi: 10.1080/09588221.2010.520673. URL https://doi.org/10.1080/09588221.2010.520673.

[8] Kun Qian, Ryan Shea, Yu Li, Luke Kutszik Fryer, and Zhou Yu. User adaptive language learning chatbots with a curriculum. In Ning Wang, Genaro Rebolledo-Mendez, Vania Dimitrova, Noboru Matsuda, and Olga C. Santos, editors, *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky - 24th International Conference, AIED 2023, Tokyo, Japan, July 3-7, 2023, Proceedings*, volume 1831 of *Communications in Computer and Information Science*, pages 308–313. Springer, 2023. doi: 10.1007/978-3-031-36336-8_48. URL https://doi.org/10.1007/978-3-031-36336-8_48.

[9] Hana Raz. Role-play in foreign language learning. *System*, 13(3):225–229, 1985. ISSN 0346-251X. doi: https://doi.org/10.1016/0346-251X(85)90037-5. URL https://www.sciencedirect.com/science/article/pii/0346251X85900375.

[10] Murray Shanahan, Kyle McDonell, and Laria Reynolds. Role-play with large language models. *CoRR*, abs/2305.16367, 2023. doi: 10.48550/arXiv.2305.16367. URL https://doi.org/10.48550/arXiv.2305.16367.