

Personalization in English-Teaching Chatbots

Siyan Li

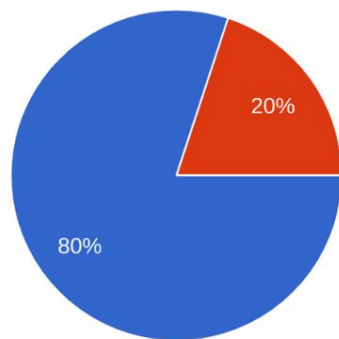
THE NEED FOR PERSONALIZATION

- Surveyed X / Twitter for **Mandarin native speakers** about English learning feedback preferences
- Feedback style, feedback type, & responder demographics
- 456 Responses!!

THE NEED FOR PERSONALIZATION

你喜欢老师给你反馈的时候用什么样的语气？

456 responses



● 口语化的 / colloquial

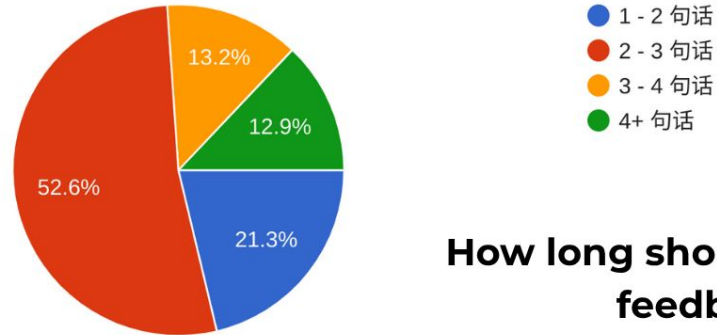
● 正式的 / formal

What do you want your English teacher's tone to be when they are giving you feedback?

THE NEED FOR PERSONALIZATION

反馈的长度应该多长呢？

456 responses

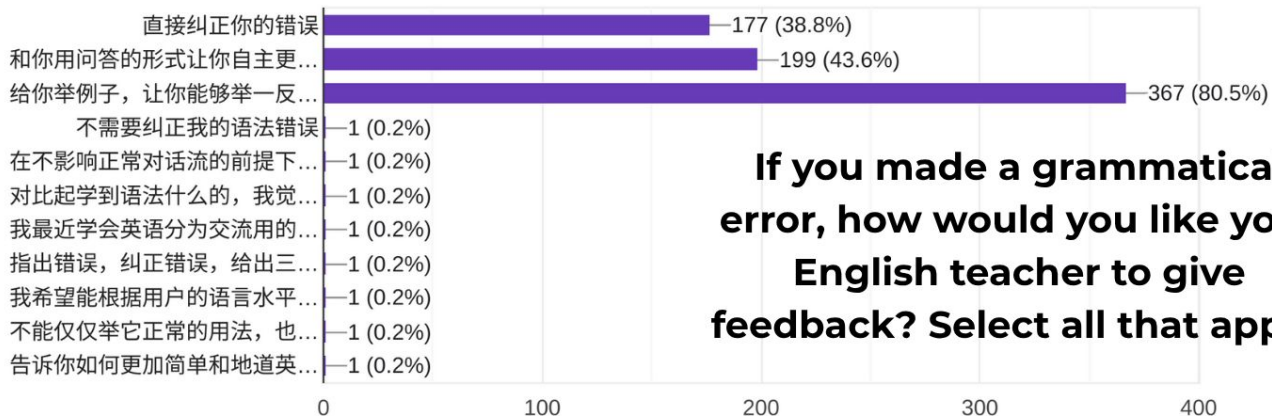


How long should each piece of feedback be?

THE NEED FOR PERSONALIZATION

如果你犯了语法错误，你希望老师怎样给你相对应的反馈？请多选。

456 responses



If you made a grammatical error, how would you like your English teacher to give feedback? Select all that apply.



PERSONALIZED LEARNING

- Has been shown to:
 - Increase student motivation, engagement, and understanding [1]
 - Make learning more effective [2]
 - Maximize learner satisfaction [2]
- However, **difficult to scale to more students**

[1] Falcão, T. P., e Peres, F. M. D. A., de Moraes, D. C. S., & da Silva Oliveira, G. (2018). Participatory methodologies to promote student engagement in the development of educational digital games. *Computers & Education*, 116, 161-175.

[2] Gómez, S., Zervas, P., Sampson, D. G., & Fabregat, R. (2014). Context-aware adaptive and personalized mobile learning delivery supported by UoLmP. *Journal of King Saud University-Computer and Information Sciences*, 26(1), 47-61.



How to scale Personalized Education to EVERY STUDENT to Encourage Engagement & Improve Equity?

PERSONALIZED LEARNING

- Has been shown to:
 - Increase student motivation, engagement, and understanding [1]
 - Make learning more effective [2]
 - Maximize learner satisfaction [2]
- However, **difficult to scale to more students**

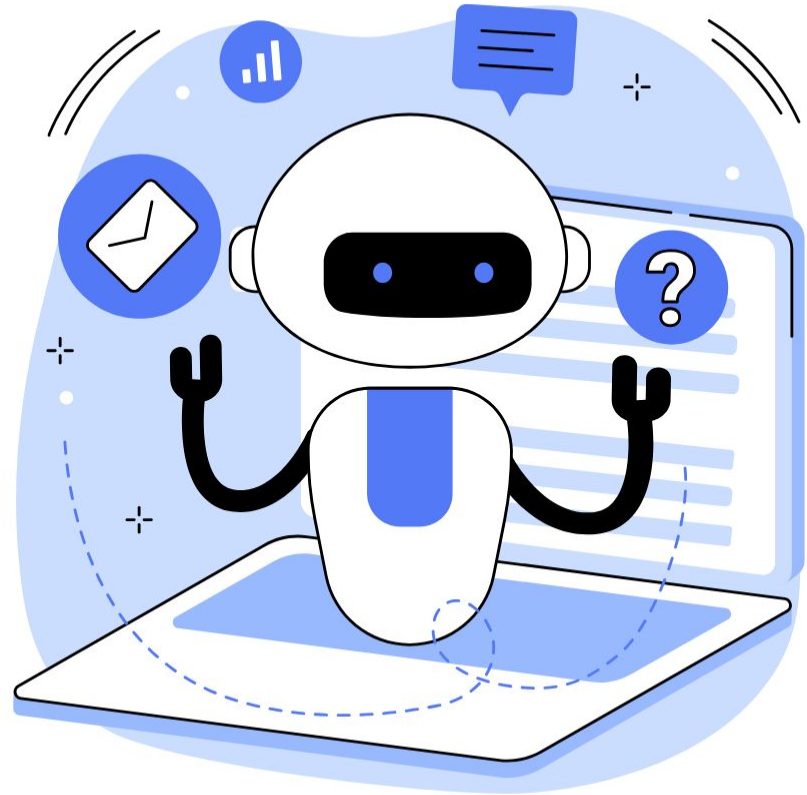
[1] Falcão, T. P., e Peres, F. M. D. A., de Morais, D. C. S., & da Silva Oliveira, G. (2018). Participatory methodologies to promote student engagement in the development of educational digital games. Computers & Education, 116, 161-175.
[2] Gómez, S., Zervas, P., Sampson, D. G., & Fabregat, R. (2014). Context-aware adaptive and personalized mobile learning delivery supported by UoLMP. Journal of King Saud University-Computer and Information Sciences, 26(1), 47-61.

ONE ANSWER: CONVERSATIONAL AI

Conversational AI, often implemented as **Chatbots**, can provide **personalized experiences** by assuming and adapting to different personas [1-2] and following user-specified instructions.

[1] Li, Y., Qu, S., Shen, J., Min, S., & Yu, Z. (2023). Curriculum-driven edubot: A framework for developing language learning chatbots through synthesizing conversational data. arXiv preprint arXiv:2309.16804.

[2] Ait Baha, T., El Hajji, M., Es-Saady, Y., & Fadili, H. (2023). The power of personalization: A systematic review of personality-adaptive chatbots. SN Computer Science, 4(5), 661.



Overview

Two research projects on Chatbots for English Learning:

1. Empathetic Chatbots to Address Learner Frustration [1-2]
2. Proficiency-Adaptive Chatbots to Accommodate Learner English Levels (Ongoing work)

[1] Siyan, L., Shao, T., Yu, Z., & Hirschberg, J. (2024). Using Adaptive Empathetic Responses for Teaching English. BEA Workshop, NAACL 2024.

[2] Siyan, L., Shao, T., Yu, Z., & Hirschberg, J. (2024). EDEN: Empathetic Dialogues for English Learning. Findings of EMNLP 2024.

Empathetic Chatbots

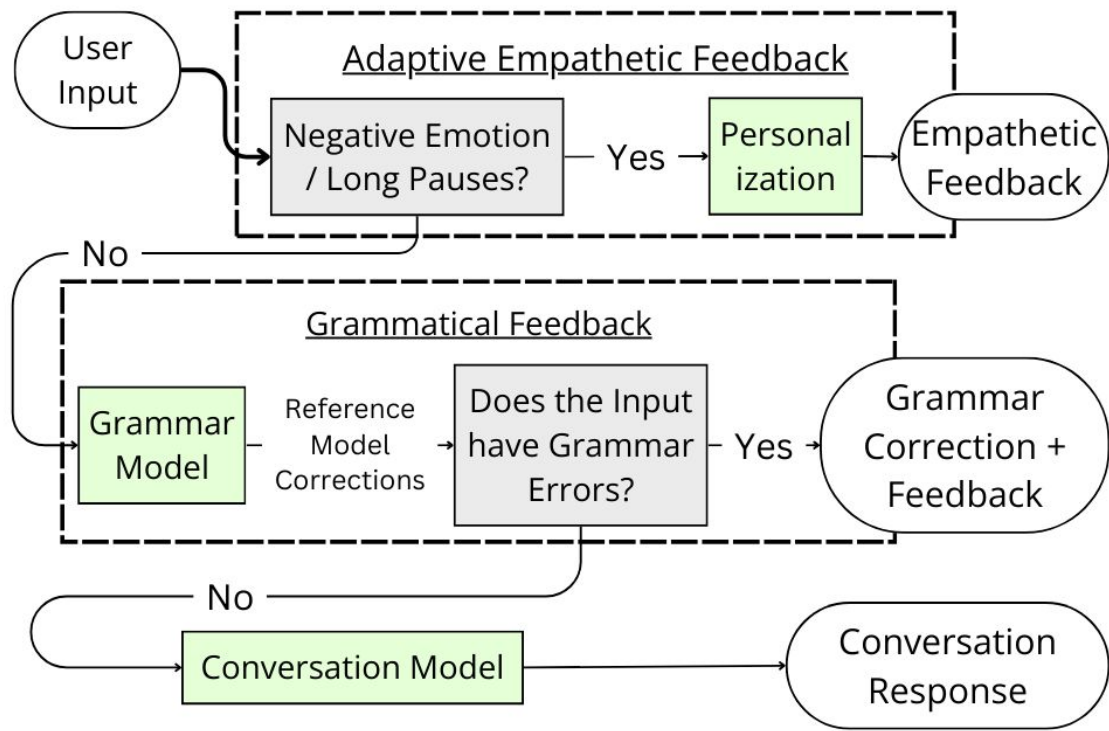
Motivation

- Frustration due to foreign language anxiety is common in English learning (or second-language learning in general)
 - Could make learners give up
- **Wu et al. (2023) : perceived teacher affective support (correlates with empathy)** positively correlates with student **L2 grit** (passion and perseverance in second language learning process) [1]
- This is done with English teachers and Mandarin native speaker learners.

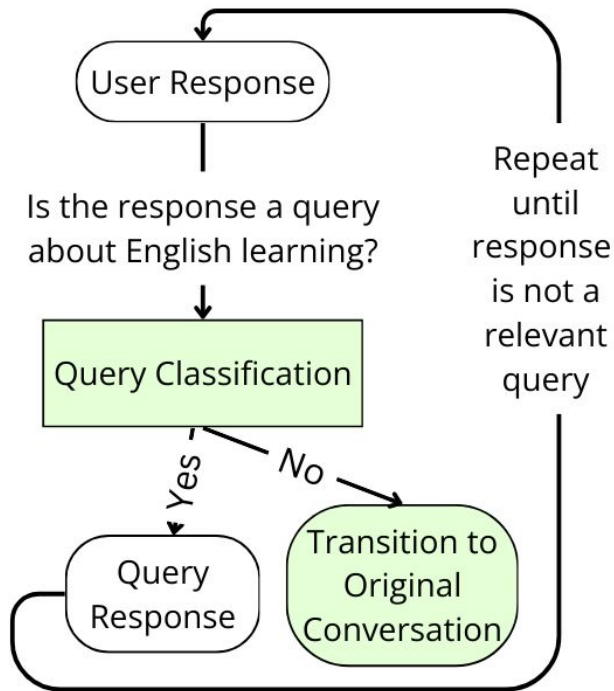
[1] Wu, W., Wang, Y., & Huang, R. (2023). Teachers matter: exploring the impact of perceived teacher affective support and teacher enjoyment on L2 learner grit and burnout. *System*, 117, 103096.

Does a Sufficiently Empathetic English-Teaching Chatbot increase Student L2 Grit?

Chatbot System Overview



After User Responds to Feedback



System Design

A chatbot for users to practice English conversations with, capable of grammar and empathetic feedback. We improve upon Siyan et al. [1]; improvements highlighted in Green.

[1] Siyan, L., Shao, T., Yu, Z., & Hirschberg, J. (2024). Using Adaptive Empathetic Responses for Teaching English. BEA Workshop, NAACL 2024.

Implementation

Empathetic Feedback Mechanisms



I want to eating ice cream.



**No
Emp.**

Maybe you meant "I want to eat ice cream". You have made an error in the verb tense here, it...

**Fixed
Emp.**

I understand that learning English can be a difficult process, but you are doing great! Keep it up!

**Adapt
Emp.**

You are doing well with vocabulary and sentence structures, but try jazzing up your grammar for...

Negative Affect & Pause Detection

- Perceived negative emotions
 - Use existing wav2vec speech emotion recognition model [1]
- Pauses, since pauses are related to foreign language anxiety
 - Use **Silero-VAD** (voice activity detection) for pause duration [2]
 - Use the average pause length as a metric
- If we detect either, then trigger Empathetic Feedback

[1] <https://huggingface.co/ehcalabres/wav2vec2-lq-xlsr-en-speech-emotion-recognition>

[2] <https://github.com/snakers4/silero-vad>

Adaptive Empathetic Feedback

- Based on past user utterances to provide targeted, specific feedback
- Employ a “**criticism sandwich**” structure
- E.g. “You’re doing well with the topic, but work on your grammar to sound smoother. For instance, say "I know what the recipe should taste like" instead of "I know the recipe taste like." Keep practicing to get even better! Does that sound alright to you?”

Grammatical Feedback

- **Synthesized grammar corrections** for real Mandarin native speaker utterances using GPT-4
- Fine-tuned Llama-2-chat-7B to provide grammar corrections
- Use a template-based approach, coupled with POS tagging, to create the grammar feedback

Conversation

- Current dialogue models tend to assume that both speakers in a dialogue are native English speakers
- Used ChatGPT to [synthesize conversations where one party is an ESL speaker](#)
- Fine-tuned **Llama-2-Chat-7B** to conduct conversations with ESL speakers on a variety of topics
 - E.g. TV shows, movies, books, etc
- Spoken chatbot; TTS using **Amazon Polly**, ASR using **Whisper**

User Study

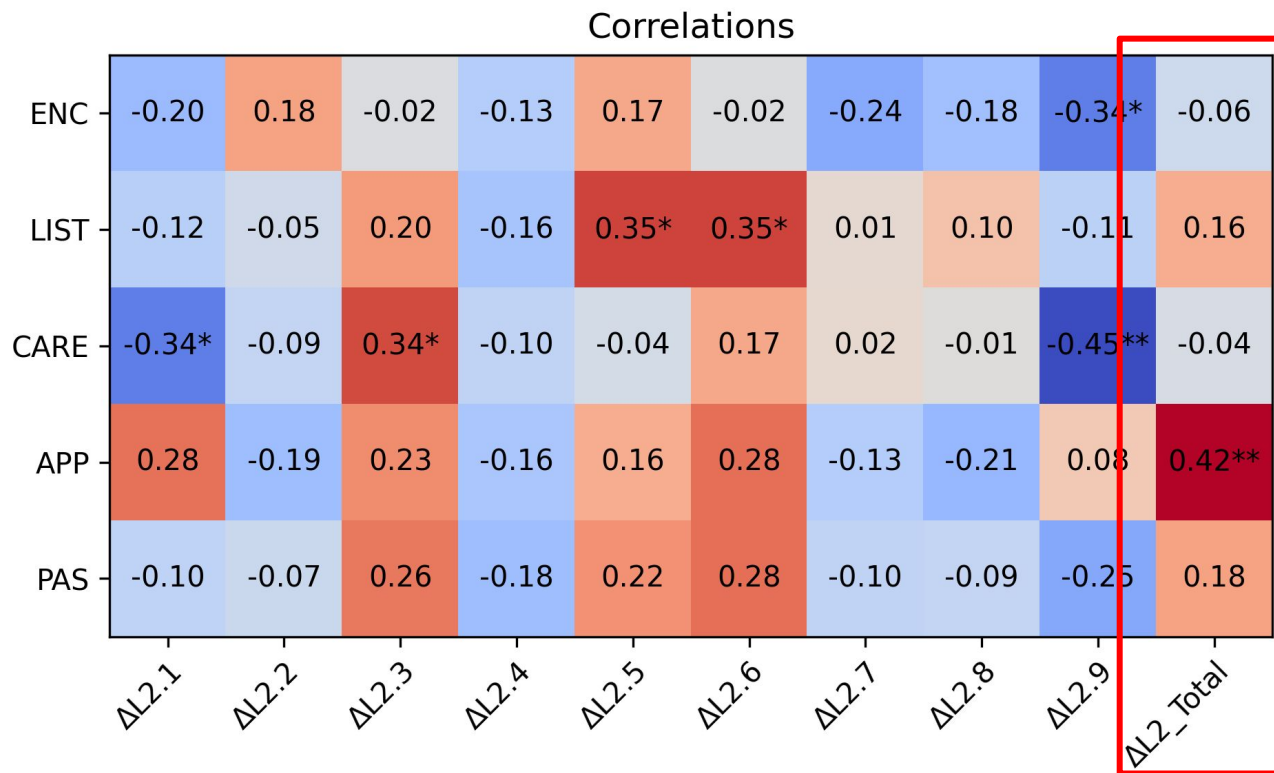
Population

- Recruited through social media and personal connections
- **31 Native Mandarin speakers**
- Self-reported English proficiency at least intermediate
- Filled out questionnaire for perceived affective support from chatbot and L2 grit

PAS and Different Empathetic Feedback Mechanisms

	Encourages me when struggling	Listens to me	Cares about my opinion	Appreciates and praises me when doing well	PAS
None	3.53	4.12	4.00	3.47	3.78
Fixed	3.83	2.83	3.00	3.67	3.33
Adaptive	4.38	4.00	3.88	4.38	4.16

Correlation between PAS and Change in L2 Grit



Takeaways

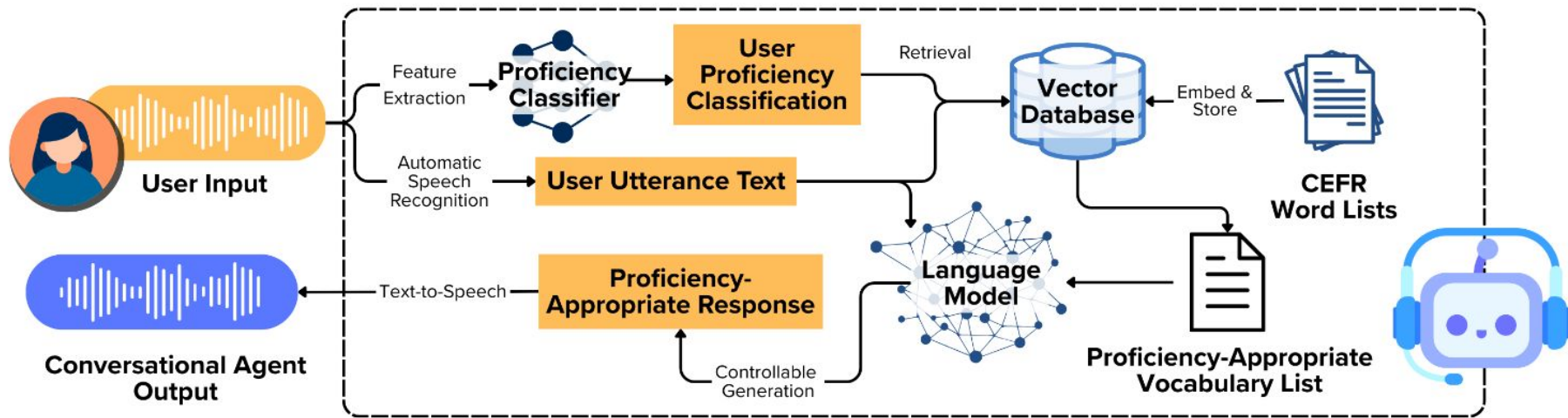
- The **adaptive empathetic feedback** approach is the most effective at increasing perceived affective support.
- Some aspects of *perceived affective support* correlate with *increasing student L2 grit* post-interactions.
- We can attempt to expand sample size, but this can be difficult for many reasons in the United States.
- Users appreciate how supportive the empathetic chatbot is; they are sometimes pleasantly surprised when they receive encouragement.

Proficiency-Adaptive Chatbot

Slides Text Credit: Zack Rackauckas

Motivation

- English learners have different proficiency levels
- Using vocabulary that are significantly out of their proficiency level can be frustrating and discouraging
- LLMs like ChatGPT often assume that the users are native English speakers
 - And therefore produce prolonged responses with complex vocabulary



Chatbot System Structure

Common European Framework of Reference for Languages (CEFR): Measure for English language proficiency. Levels range from A1 through C2.

Proficiency Classifier

- **Data Collection:**
 - We sampled data from COREFL [1], the Qatar Corpus [2], the EDEN ASR Dataset [3], and LibriSpeech [4].
 - Four native English speakers labeled the perceived proficiency levels according to CEFR.
- **LLM-as-a-judge Annotation:**
 - Used gpt-4o-mini, a large language model (LLM) to label proficiency levels using features extracted from audio clips; validated annotation alignment with humans; and labeled the rest of the dataset.
- **Random Forest Classifier** trained on the annotated dataset yields 0.95 Macro F1 score on the test set.

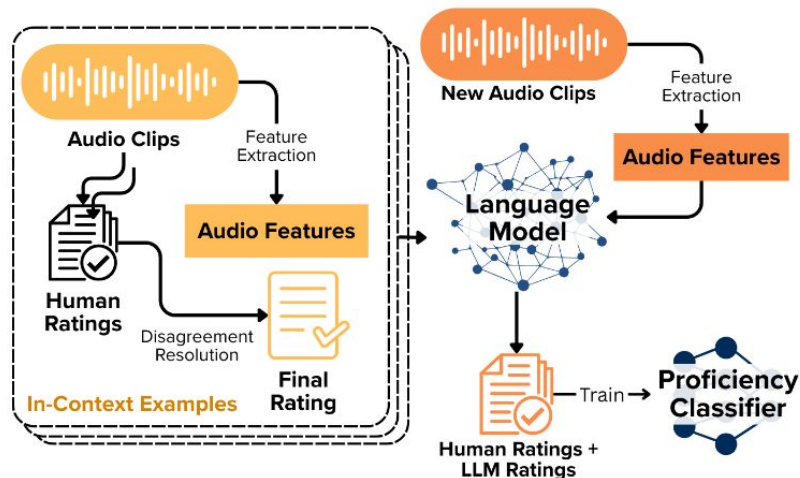


Figure 2. Training the Proficiency Classifier.

Proficiency-Based Output Adaptation

- Upon selecting the retrieved proficiency-appropriate vocabulary that is also relevant to the conversation context:
 - The **logits** corresponding to the vocabulary are **boosted**, so those words have a higher probability of appearing in the assistant's response.
 - The **proficiency-appropriate vocabulary** is also passed into the language model's prompt.
 - The language model generates two sentences: the first **addressing the user's utterance**, and the second asking the user back a **relevant question** to continue the conversation.
- We leverage the Qwen 2.5-14B instruct LLM.

Thank You!

Contact: siyan.li@columbia.edu