AdaptLingo: *An Adaptive Speech-to-Speech Conversational Agent Tailored to Users' English Proficiency Levels*

Columbia University in the City of New York

Zackary Rackauckas, Riya Raj, Siyan Li, Julia Hirschberg

Columbia Engineering

Language students have **varying levels of proficiency**.

How can an English conversation partner agent **adapt to different student proficiency levels**?

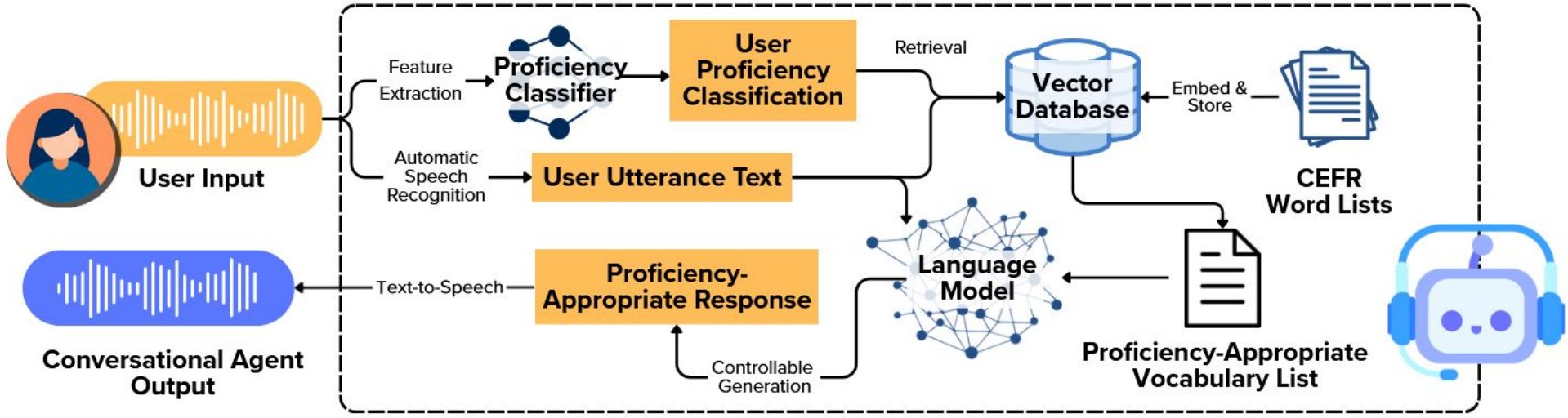

Figure 1. Overall Conversational Agent Workflow.

Step 1: Scan QR Code

Step 2: Click “Visit Site”

Step 3: Scroll down.

Step 4: Use the chatbot!

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 the Common European Framework of Reference for Languages (CEFR) for 301 audio clips.
- **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.

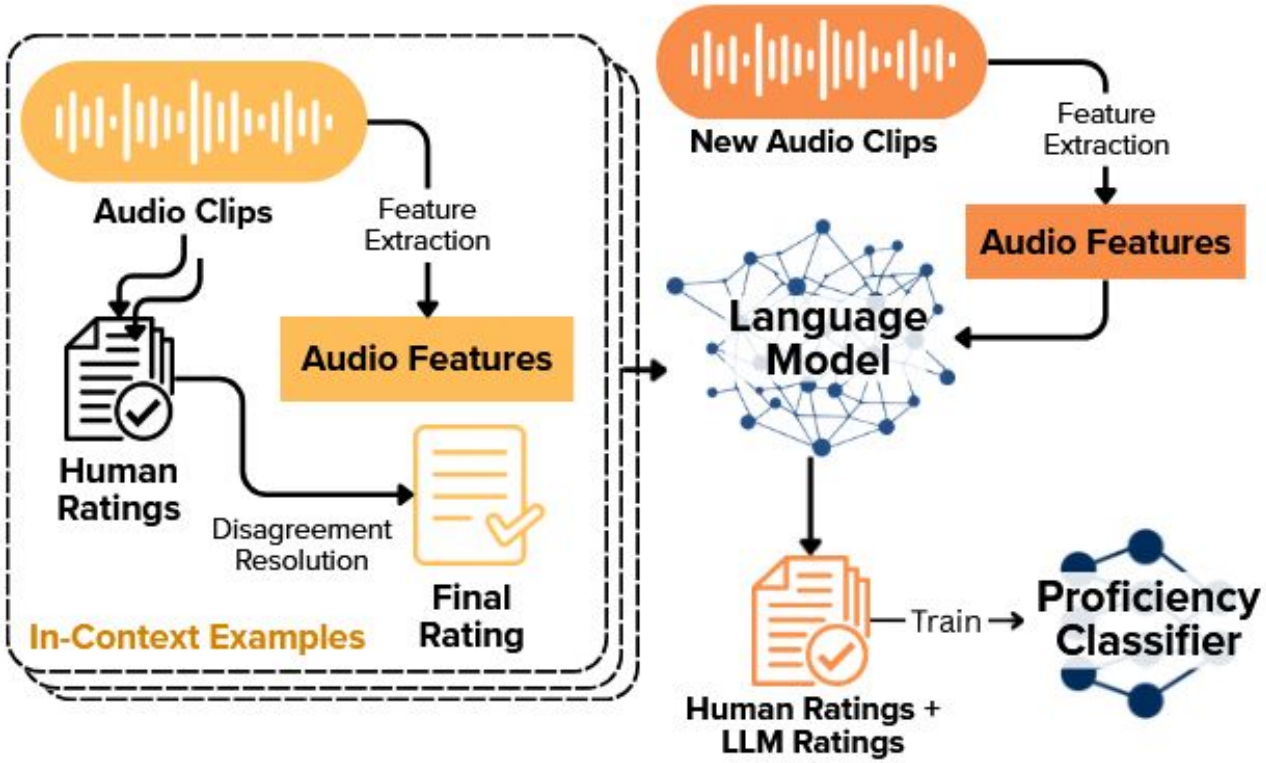

Figure 2. Training the Proficiency Classifier.

Classifier	Test Accuracy	Test Macro F1	Cross-Validation Accuracy Mean	Cross-Validation Macro F1 Mean
SVM	0.93	0.92	0.90	0.89
XGBoost	**0.95**	0.94	**0.93**	**0.92**
Random Forest	**0.95**	**0.95**	0.92	**0.92**

Table 1. The performance of different Proficiency Classifiers on our Test set.

Retrieval-augmented Generation (RAG)

- We use the *all-mpnet-base-v2* Sentence Transformers model to **embed individual words** in CEFR vocabulary lists.
- We use Chroma for vector storage and *k*-nearest-neighbors (KNN) search for **retrieval**.
- KNN search finds the top *k* words that have the most similar meaning to the **full user's utterance**.
- 40 words are retrieved to construct a **proficiency-appropriate vocabulary** for generation.

Language Model 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 **Qwen2.5-14B instruct LLM.**

I want to practice English conversation.

We can chat anytime. What topics interest you?

I'm interested in cooking.

Cooking sounds fun. Do you like baking too?

Yes, I like baking too!

Great, baking is enjoyable. What kind of cakes do you bake?

I bake chocolate lava cakes.

Get in Touch

- zcr2105@columbia.edu
- https://2950aafc6f4e.ngrok-free.app//