21CNTThuc   - Xử lý ngôn ngữ tự nhiên ứng dụng
Group 5   - Teach me
19127620 - Nguyễn Văn Vũ Tường
20127349 - Trần Quốc Thuận

Project ID "H.11" - AI English Tutor

Merits of the topic:
　　　　OpenAI's large language model GPT-3.5 Turbo has been used to make corrections to English sentences, which helps improve the users' English proficiency. The GPT-3.5 Turbo model satisfies the scope of the topic and has a low cost to use.

　　　　ElevenLabs' text-to-speech technology has been used to enhance the learning experience with audio. ElevenLabs have natural-sounding voices.

　　　　The Streamlit library has been used to create beautiful and easy-to-use user interfaces.

Shortcomings of the topic:
　　　　The language model is stateless and has no memory of past interactions. Every interaction is completely independent of previous interactions.
　　　　It is possible to provide the model with context of previous interactions by sending the entire conversation or a summary back to the model. However, this will use more tokens and any references to lost data will cause hallucinations.

　　　　ElevenLabs has a high cost to use and runs out of usage easily.

　　　　The final section contains some additional shortcomings. However, these shortcomings are beyond the scope of the original topic and are therefore not included in this section.

A summary of the topic's methodology:
　　　　The topic uses the "one-shot" prompting method.
　　　　In addition to the instructions and input data, the prompt also contains an example. This example consists of a pair of input data and an answer.
　　　　Based on that example, the model learns and then gives an answer that satisfies the requirements of the topic: making corrections to English sentences, which helps improve the users' English proficiency.

Differences from the original topic:
　　　　The OpenAI API and ElevenLabs API were updated after the initial topic's creation. Our group has updated the source code to run the newest version of these APIs.

　　　　ElevenLabs easily runs out of usage, so Google Text-to-Speech has been added in order to easily demonstrate the audio feature.

　　　　A few features have been added. They are simple and not as powerful as the main feature of the original topic. However, they help users have a richer English learning experience. Those features are:
- Give a piece of advice on learning English.
- Explain a mistake that people often make while using English.
- Explain a word or phrase, then use it in a sentence.
- Convey an idea in a different way.
- Pronounce a sentence.

More shortcomings:

Our group wanted to add some features with the purpose of helping users prepare for the writing section of the IELTS or TOEIC test. However, we did not have enough experience and time to do this (creating suitable prompt structures, testing different prompts, checking the results).

Furthermore, these examinations are important and expensive. Therefore, users should not rely on large language models but consult qualified instructors instead.

The GPT-3.5 Turbo model has certain limitations. It is adept at generating responses based on vast datasets. However, it does not understand text at a human level. It may provide information that sounds reasonable but is inaccurate or fails to grasp complex subject matter.

While assessing sentences and essays answering the writing section of the IELTS or TOEIC test, the language model can't seem to follow the rubrics of the IELTS or TOEIC test. Nor can it accurately assess the content and organization of the essays.

Therefore, such important tasks should be performed by a qualified IELTS or TOEIC instructor. They can provide you with proper feedback and accurate scoring.

Future and potential:

Using large language models to enhance the English learning experience has become a reality.

As of February 2023, Duolingo is the world's most popular language learning app based on monthly downloads, with nearly 13.4 million users downloading it that month. The Duolingo Max service uses the GPT-4 model to interpret answers and practice conversation.

Large language models will continue to improve and become more accessible. In the future, there may be large language models or specialized models capable of assessing the writing section of the IELTS or TOEIC test. To do that, it will require the attention of relevant organizations to create datasets, train and evaluate the models.

On April 23rd, the professor provided some feedback:
- "one-shot" prompting and "few-shot" prompting are different. "one-shot" prompting is correct. See week 5's slide, page 45.
- When you need to run legacy code, you can import specific versions of Python libraries.
- In order to implement the IELTS features, provide about 4 bands and use "zero-shot" prompting. There aren't enough tokens for even "one-shot" prompting.