Balazs Zsigmond Horv Ath, 2023, 11:5
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Duenna AI – An Approach to Use Artificial Intelligence in Language Education

1 Balázs Zsigmond Horváth, 2 Bence Blaske, 3 Botond Csorba
1 Pogány Frigyes Technical School, Budapest, Hungary,
2 SAP Hungary, Budapest, Hungary,
3 ELTE University Department of Arts and Humanities, Budapest, Hungary

Abstract-Nowadays, more and more articles and books are published about the use of artificial intelligence in teaching. With our present research focusing on language learning, we want to examine the possibility of using artificial intelligence to write and act out dialogues for practicing listening comprehension so that classroom tasks can be brought into line with the requirements of later exams in the life of an eager foreign language learner. Only experience over time can be beneficial for accomplishing exams, summative in nature. The dialogues make use of the endless tuning options of the safeguarded guiding prompts used to instruct the AI. We would like to expand on the idea of simply communicating with applications like ChatGPT and provide students with understandable materials that are compatible with any given methodical and ethical standard. Our second aim is to develop a portable version as well for people on the go who want to enhance their vocabulary with the help of giving contexts to expressions. Language styles and moods are also explored. The main question is that can the AI provide good-enough materials for educational use? Teacher control is still a necessity, but the application can help supplement creativity when there is simply no time to just use human imagination during curriculum development. Our research is continuous and concentrates on the development of a combined application that is a simple solution for all users practicing listening comprehension to develop appropriate communication skills and to help overcome anxiety when in any speaking situation.

Keywords- AI-aided language learning, text to speech, language teaching

I. INTRODUCTION

The aim of using AI in dialogue generation was that surely the software will use a much more colourful vocabulary than an ordinary human, as the AI has an unlimited access to different dictionaries and also with every example its capabilities are enhanced. There may be different contexts or cohesions that a teacher may not realise because of constant pressure of preparing every day for his or her classes. Also, we wanted to see how creative or varied the AI can be when composing the dialogues. Skrabut states that there is certainly a huge potential in making AI a part of teaching methods regardless of all the expected benefits and disadvantages. The suggested approach is that it can be used as an aid to build up a curriculum for the language classroom and then suggest tasks to support the advancement of the students. Critical thinking is a must on the human side, but the AI can add ideas that perhaps do not always come into mind of teachers amongstthe flood of other duties. [1] We support his viewpoint but we wanted to be more specific and focus on a specific area which counts the most when students prepare for any examination and that area was comprehension – in this study. mainly listening.

© 2023Bal Azs Zsigmond Horv Ath. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

Background

We tried to provide as much tuning opportunities for the tone of the dialogue, the topic of the whole conversation and also for the gender of the participants. Both same and different sex conversations are applicable to simulate naturalistic speaking or interaction environments. Previously we conducted similar tests, but the main focus of that research was on using TTS (text-to-speech) voices in general to make recording and playing back dialogues easier. [2]When the idea was tried out, the dialogues were all written by language teachers, so the human factor dominated the composition style. Now the situation has changed, as now we were curious to see whether the AI can compose good-enough conversation-pieces for language education when prompted or instructed correctly. The already built-in safeguards provide enough quality assurance and guarantees proper wording and the cleanliness if the exchanges.

TTS voices have evolved tremendously over the last decade, so using them is a valid approach even more so nowadays, as one factor in IT also is using the least amount of any resources available. [3] Thus the software was designed to practice listening comprehension and grammar, but was the best applied as practicing or testing aid for boosting listening skills.

A study by Graham and Macaro explored the effectiveness of dialogues in French as part of a foreign language classroom experience. The findings show that students who engaged with dialogues as listening tasks demonstrated improved comprehension skills, vocabulary acquisition, and overall language proficiency. [4]

Vandergrift investigated the impact of dialogues on second language using listening comprehension tasks. The research concludes that dialogues provided learners with exposure to more authentic language situations, enhancing their ability to understand spoken language and extract meaning from context. It is also noteworthy that the importance of dialogue-based tasks can promote learner engagement and intrinsic motivation (Vandergrift, 2012).

II.METHODS

Altogether 12 school-leaver students were in the control group (1 female and 11 males). The testing group consisted of IT students from the 9th grade who all performed beyond the school's entry-level exam acceptance score/result and had a history of learning English for more than 3 years. The testing group consisted of 15 students, (1 female and 14 males).

All students had access to the internet on their home computers and they all used social media and video sharing platforms. Taking this factor in mind and paying attention to it very seriously, we could observe that there was no significant difference between how well the students performed both in the control or the testing groups. After the initial testing phase was completed, the program was introduced in all groups of which the writer of this article had access to in the last school year. The total number of students who had access to the software was 75 from 5 classes.

Open discussion was held after each of the testing sessions and it was agreed that only the topics of the chosen dialogues made it sometimes harder for the students to understand the spoken dialogues. Articulation and the classroom acoustics were still the same during the traditional listening comprehension tasks and the AI-generated dialogue pieces.At first the dialogues were played three times due to their short nature, but after the first testing sessions once it was agreed that the sound quality was good, one playback/repetition was enough, so the tasks were performed as in real English language exams, especially the Matura, school leaving exam. The hardness of the level of the used vocabulary was one of the key points for the research at hand.

This presents no problem as the curriculum uses listening comprehension for practicing purposes and their texts can contain unfamiliar words. This means that the students always have to listen to the gist of the material first.Listening to different contexts as all the comprehension-related questions in the Hungarian Matura examination are summative in nature.

III. RESULTS

The questions related to the dialogues were written by the researchers after the first dialogues were generated and listened to. Much attention was paid to the quality of the material. The questions were presented to the students on paper in the form of questionnaires. The questions concentrated on different phrases used, the overall turning points of the dialogue narrative, or were yes/no questions. Later students were also asked to compose the topic and other attributes of the dialogues themselves and let's observe together what the AI can come up with.

It was interesting to see that the students could judge whether their generated suggestions are suitable enough to be part of listening comprehension practice runs with the whole group. Two students mentioned that that listening to dialogues this way can help them overcome learner anxiety as they would know what reactions are needed to answer questions correctly. It seems that with the proper prompting there was a 25% chance of repetitions in wording or dialogue outcomes. This is both a curse and a blessing, because repetition may help a newcomer student remember phrases more easily, but the dialogues may become boring for students at more advanced levels. We overcame this issue by regenerating dialogues until they were totally different from previous ones.

The development of the software was done in MVP-style(minimum viable product)and each test was followed by a patching period. Newer tests only took place once new versions were fully completed.

1.Technical Overview

In Duenna AI, the dialogues are generated by an artificial intelligence based on the conditions you specify. Duenna AI uses a variant of the GPT (Generative Pre-trained Transformer) artificial intelligence model, which is a kind of neural network trained to generate natural language text. It uses a transformer architecture and is trained on large text datasets such as books, articles and websites.

During the training, the model learns the patterns and structure of the text in the dataset and then uses this knowledge to generate new text similar to the training data. Given the input data, Duenna's AI model uses this learned knowledge to generate highly lifelike and meaningful conversations in a coherent and context-appropriate way. The model processes the input text token by token and generates the next token based on the probability distribution of all possible tokens, taking into account the input context. However, this type of language model has its limitations, so in some cases the generated text may contain inappropriate or strange details.To make the learning experience even more realistic, Duenna AI makes use of advanced speech synthesis APIs, using which it reads the generated conversation aloud.

2.LLMs and current AI models

The Artificial Intelligence that makes the generated conversations very convincing is a so-called Large Language Model. The current leader of this field of AI is the company called OpenAI.

A large language model is a type of artificial intelligence (AI) model that is trained on vast amounts of textual data and can generate human-like language in response to a given prompt or input. Large language models have significantly improved the capabilities of natural language processing (NLP) and have been applied in various applications such as language translation, sentiment analysis, and chatbots. OpenAI GPT models (Generative Pre-trained Transformer) are a family of large language models developed by OpenAI, a leading AI research laboratory. These models are trained using a technique called unsupervised learning, where the model learns from a vast amount of unstructured text data from the internet. The GPT models use a transformer architecture, which enables them to process and generate text in a context-aware and human-like way. GPT-3, the latest and largest version of the GPT model, has 175 billion parameters, making it one of the most powerful language models in existence. It can generate text that is difficult to distinguish from text written by a human and has the potential to revolutionise the field of NLP.

Large language models can also be used for a wide range of other applications, such as chatbots and virtual agents. By analysing natural language patterns, they can generate responses that are similar to how a human might respond. Of course, like any technology, large language models have their limitations. One of the biggest challenges is ensuring that the content they generate is accurate and reliable. While LLMs can generate content that is similar in style to a particular author or genre, they can also generate content that is inaccurate or misleading. This is particularly true when it comes to generating news articles or other types of content that require a high degree of accuracy.

Nevertheless, due to the inherent stochastic nature of such models, achieving absolute 100% accuracy is presently unattainable. Consequently, it remains imperative to incorporate a human-in-the-loop approach to validate any content generated by large language models before conveying it to end-users. This practice becomes particularly crucial when applying such technology in enterprise settings, where potential liability concerns may arise.

3.Currently available language models

Large language models, transformer-based AI linguistic models have been around for a few years, but only since the release of OpenAIChatGPT in November 2022 have they become widespread. Since then, many big companies (Google, Antropic), as well as smaller startups have been rushing to release more capable and/or cheaper to operate AI models. The biggest vendor in the area is still OpenAI. Their GPT3.5 and the more advanced GPT4.0 LLMs are still the state-of-the-art models for the generic use cases. There are many other AI models that are fine-tuned e.g., for coding or for medical use. In the Duenna AI app, we are leveraging the OpenAI GPT3.5, since it's a relatively quick LLM, with good output quality for the use case we are targeting.

4. Instructing the language model

There's a paradigm shift in the way you can "program" large language models. In the traditional approach, when creating a deterministic computer program, it's sufficient to write instructions expressed in a computer programming language, and one can expect that for the same input, the same output will be returned. Even if the program contains programming errors, the error situations themselves are also deterministic, and can easily be "debugged", the error can be efficiently addressed.

This is in contrast with how one can "instruct" large language models. They can be "programmed" using natural language. While natural language is very expressive, it's also prone to misinterpretation. In programming, misinterpretation by the software compiler can never happen. A piece of programming keyword or expression always bears the same meaning. When working with large language models, this is not the case. Let's explore this a bit further. A particular language model, the ChatGPT 3.5 from OpenAI, was asked the following question, and here we have recorded 3 different replies that were returned by the AI model when asked 3 times.

Prompt example: Is chess a good game? Reply briefly!

Response from AI:

1. Yes, chess is a highly strategic and intellectually stimulating game.
2. Yes, chess is considered a good game because it provides mental stimulation, strategic thinking, and opportunities for skill development.
3. Yes, chess is considered a good game due to its strategic depth, intellectual challenge, and ability to improve cognitive skills.

As can be seen from the above example, that - while the responses are somewhat similar in their core meaning - are quite different even if the same question is asked. This variation of the responses is inherent due to the stochastic nature of the language models. Such AI models have a so-called "temperature" parameter, which can mitigate the randomness of the responses, but even with the temperature parameter turned down to 0, the system still exhibits a certain amount of randomness. Moreover, the temperature parameter is also regarded as the "creativity factor" of these large language models, hence tuning it down can have an adverse effect on the quality of the output, where it's expected that the system responds with a more natural sounding text. It is therefore a trade-off: if we want to have let the system be creative with the responses, we will have to deal with the stochastic nature of the model. This brings us to the main challenge of these large language models, that it is difficult to achieve a fully predictable output. We discuss how this can be mitigated with so-called prompt engineering.

5.Prompt engineering

Prompt engineering refers to formulating effective prompts that can efficiently instruct LLMs to perform desired tasks and a certain type of output. It involves the precise wording, formatting, and structuring of a prompt to optimise the model's output, enhancing relevance and accuracy. The practice of prompt engineering is critical due to the sensitivity of LLMs to input. The slight variation in how a query is presented can lead to significantly different responses, affecting the overall utility of the model. Prompt engineering is not an exact science. It requires many iterations, essentially trial and error, to get the desired output. The high level (iterative) process can be seen in Figure 1.

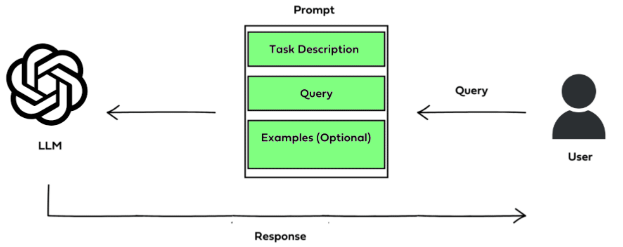


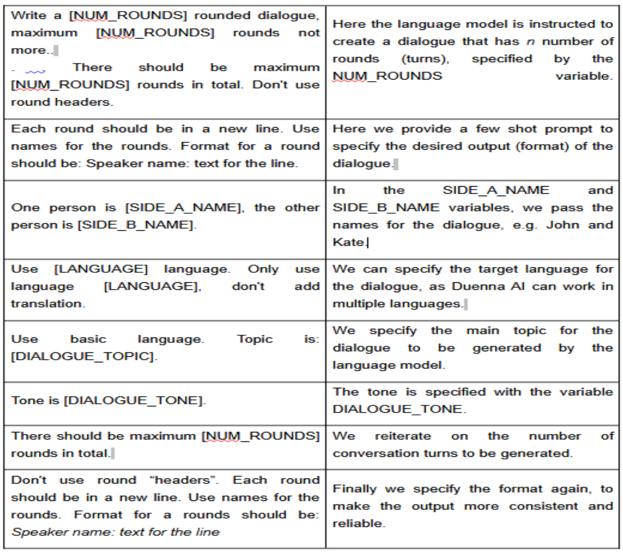
Figure 1: Basic prompt engineering

6.Basic conversation prompt

For the Duenna AI app, the large language model AI is instructed to generate a realistic dialogue between two persons. It is possible to influence the topic of the dialogue, as well as the style (tone). For this, we are prompting the language model to return a dialogue with the specified conditions and in the specified format.

Essentially, (to serve as a simple example), you can prompt the language model like:

"Write a dialogue between 2 persons, where the topic is global warming". This simple sentence will in most cases yield a meaningful output, however we need to be more precise with the instruction than this, in order for the output to be usable in a computer program. The computer program needs to be able to "parse" the output and also the instruction needs to honour the settings from the end user (like topic and style).Therefore, for the app, we have a more elaborate set of prompts that you can see in the next section.



Write a [NUM_ROUNDS] rounded dialogue, maximum [NUM_ROUNDS] rounds not more.. . There should be maximum [NUM_ROUNDS] rounds in total. Don't use round headers.	Here the language model is instructed to create a dialogue that has *n* number of rounds (turns), specified by the NUM_ROUNDS variable.
Each round should be in a new line. Use names for the rounds. Format for a round should be: Speaker name: text for the line.	Here we provide a few shot prompt to specify the desired output (format) of the dialogue.
One person is [SIDE_A_NAME], the other person is [SIDE_B_NAME].	In the SIDE_A_NAME and SIDE_B_NAME variables, we pass the names for the dialogue, e.g. John and Kate.
Use [LANGUAGE] language. Only use language [LANGUAGE], don't add translation.	We can specify the target language for the dialogue, as Duenna AI can work in multiple languages.
Use basic language. Topic is: [DIALOGUE_TOPIC].	We specify the main topic for the dialogue to be generated by the language model.
Tone is [DIALOGUE_TONE].	The tone is specified with the variable DIALOGUE_TONE.
There should be maximum [NUM_ROUNDS] rounds in total.	We reiterate on the number of conversation turns to be generated.
Don't use round "headers". Each round should be in a new line. Use names for the rounds. Format for a rounds should be: *Speaker name: text for the line*	Finally we specify the format again, to make the output more consistent and reliable.

7.Advanced prompting techniques

To realise all the features of the application, the large language model receives a complex prompt-based instruction that contains the user specified configuration for the dialogue. This is just an excerpt that contains the most important parts. To steer the AI model towards a more consistent and reproducible behaviour, some of these instructions are reformulated in a different way that makes the output more reliable.

Table -1: Advanced prompting techniques

9.Prompts for topic

Duenna AI has a variety of built in topics that are selectable as a topic for a dialogue: family, friends, holidays, computer games, risks of AI, shopping, school, jobs, office, food, environment, health, restaurant, sports, money, gifts, weather, films, books, hobbies, music, animals, dreams, love, cooking, computers, museum, travel.With these topics, the generated dialogue will match the specified topic. It is also possible to enter a user specific topic, that will then be used as part of the instruction prompt for the language model.

10.Prompts for tone (style)

The end user is also able to specify the tone / style of the dialogue, which has a significant influence on the generated output. The following styles can be specified: cheerful, happy, gangsta, sad, sarcastic, easygoing, funny, cynical, apathetic, silly, offensive, nostalgic, critical, calm, pirate like

11.Example generated dialogues

Here are 2 prompts and the corresponding generated outputs:

Write a 3 rounded conversation, maximum 3 rounds, not more. One person is Robert, the other person is Emma. Use the English language. Only use the English language, don't add translation. Use basic language. Topic is: family. Tone is happy. There should be a maximum of 3 rounds in total. Don't use round headers. Each round should be in a new line. Use names for the rounds. Format for a round should be: Speaker name: text for the line.

Robert: Hi Emma! How are you today?

Emma: Hi Robert! I'm doing great, thanks for asking. How about you?

Robert: I'm fantastic too! You know, I just had the most wonderful time with my family over the weekend.

Emma: That sounds amazing, Robert! What did you all do together?

Robert: We went on a picnic to the park and played various games. It was so much fun seeing everyone laughing and enjoying themselves. We even had a delicious homemade lunch!

Emma: That sounds like a perfect day with your family, Robert. Spending quality time with loved ones is truly special.

Write a 3 rounded conversation, maximum 3 rounds not more. One person is James, the other person is Avery. Use the English language. Only use the English language, don't add translation. Use basic language. Topic is: food. Tone is offensive. There should be a maximum of 3 rounds in total. Don't use round headers. Each round should be in a new line. Use names for the rounds. Format for a round should be: Speaker name: text for the line.

James: I can't believe you actually like eating that garbage fast food. It's disgusting and unhealthy.

Avery: Well, I enjoy the convenience and taste of fast food. It's my personal preference.

James: Personal preference? More like a lack of taste buds! How can you even stomach that greasy, processed junk?

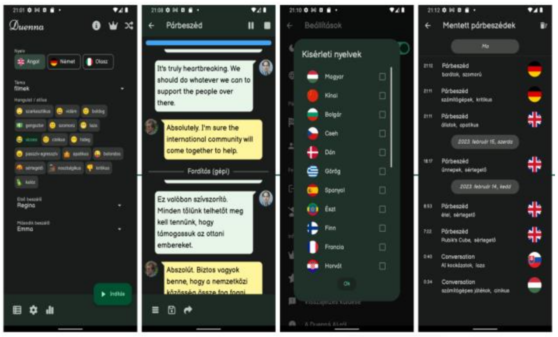
Avery: Look, everyone has different tastes. Just because you don't like it doesn't mean you have to insult my choices. Let's agree to disagree, shall we?

Realistic voice synthesis

Once the dialogue has been generated with the help of the large language model, the text is then acted out by two “voice actors”. These AI based “actors” are using realistic, human-like synthesised voices (text to speech) provided by Microsoft (their Cognitive Services offering). These voices leverage cutting edge, deep learning (neural network) based artificial intelligence models. The voices are available in a large variety of languages.

Development framework

Duenna AI is built using the Flutter UI framework. Flutter is Google's portable UI toolkit for crafting beautiful, natively compiled applications for mobile, web, and desktop from a single codebase. Flutter is free and open source. All the mentioned features are visible in Figure 2.



It's truly heartbreaking. We should do whatever we can to support the people over there.

Absolutely. I'm sure the international community will come together to help.

Ez valóban szívszorító. Minden tőlünk telhetőt meg kell tennünk, hogy támogassuk az ottani embereket.

Abszolút. Biztos vagyok benne, hogy a nemzetközi

Figure 2: Duenna AI features in action and the user interface

IV.CONCLUSION

The initial experience gathered while developing the application is promising. More tests should be carried out in parallel with the observation of new development trends in AI. Not all features are useful for developing dialogues, so the human element can never be ruled out from the process.

However the teachers’ changing role of providing a guidance of what materials could be used for students, may come even more to the front by using AI in educational environments. Student acceptance of Duenna AI as an approach is positive and full with curiosity.The next step is to urge the students to come up with even more creative topics, preferably on their own, and to be familiar with an even greater vocabulary they can use when they speak a foreign language later in their professions and personal lives.

Acknowledgements

We would like to thank all our friends who provided us with profile pictures to be used in the application and for the students who willingly participated in testing out our application.

REFERENCES

[1] Skrabut S. (2023). 80 ways to use ChatGPT in the classroom: Using AI to Enhance Teaching and Learning. Stan Skrabut, ISBN 13: 9798985553765

[2] Horváth B. Zs., Blaske B., Szabó A. (2016). Duenna - An experimental language teaching application. SoftwareX, 2016. 5, 171-177.

[3] Widyana A. (2022). The Application of Text-to-Speech Technology in Language Learning. Proceedings of the Sixth International Conference on Language, Literature, Culture, and Education. ICOLLITE, 2022. 85-92.

[4] Graham S., Macaro E. (2008). Strategy Instruction in Listening for Lower-Intermediate Learners of French. Language Learning, 2008. 58(4), 747-783.

[5] Vandergrift L., Goh C.: Teaching and Learning Second Language Listening: Metacognition in Action. Routledge, 2012.

Author's details

1 Balazs Zsigmond Horváth, Pogány Frigyes Technical School, 11 Thökölyút, 1183 Budapest, Hungary, zsmondi@gmail.com

2 Bence Blaske, SAP Hungary, 7 Záhonyutca, 1031 Budapest, Hungary, bbence84@gmail.com

3 Botond Csorba, ELTE University Department of Arts and Humanities, 4/A Múzeumkörút, 1088 Budapest, Hungary, jasondark5500@gmail.com