
GODELPHILE: Topic-Based Generative Models with Up-to-Date Knowledge for live conversations

Alfredo Garrachón Ruiz

ETSI de Telecomunicación, Av/ Complutense 30, Madrid - Spain, 28040
Universidad Politécnica de Madrid
a.garrachon@alumnos.upm.es

Abstract

In the quest for text-to-text models, the challenge of balancing resource efficiency and interesting conversations arises. Small models offer generic conversations, while larger ones provide more elaborate discussions but consume significant resources and lack updated information. This problem is particularly pronounced in dynamic domains like movies, series, or sports. To tackle these issues, the approach chosen involves using smaller models fine-tuned for specific subjects to reduce resource consumption and improve conversation quality within their domain. To maintain a conversational style, personalized conversations resembling human interactions are generated using ChatGPT. Additionally, for keeping knowledge up-to-date, the GODEL model is employed due to its conversational style and ability to accept external knowledge. Conversations tailored to a particular topic, such as movies, are generated, and prompt-engineering is used to simulate dialogues between users and a virtual assistant. Few-shot fine-tuning with collected dialogue data refines the GODEL model, ensuring it stays current and effective.

1 Motivation and Innovation

When searching for text2text models, we found a wide variety of options, from small models that were only capable of carrying out a generic conversation to large models that can achieve a more elaborate conversation. In both cases, two main problems limit the selection of suitable models. On the one hand, models that can carry out a proper conversation consume a lot of resources and present very high latencies. On the other hand, none of them could have up-to-date information, so that their knowledge is limited only to the data they have been trained with, leading them to feel outdated. This is especially relevant to high dynamic topics such as movies, series, or sports. Moreover, we also considered these models should be able to handle Q&A about those specific topics in a dialogue style (however most of them are only LLM not trained in the conversational style). Thus, this document explains the way in which it has been decided to solve both problems.

Regarding the problem of resources and latencies and with the aim of covering as much topics as possible, we have opted for the granularity offered by using smaller models in parallel instead of a large one, i.e., these smaller models are finetuned with data and conversations specific to a particular type of subject. By limiting their scope of action, we achieve much better results in terms of the quality of the conversation within their subject matter. However, due to the need to maintain a

dialogue conversation style, we opted for the generation of personalized conversations that resemble one between a human and a voice assistant on the specific topic by using ChatGPT [1].

On the other hand, to solve the problem of updated knowledge of these models, we choose to use the generative model GODEL (Large-Scale Pre-Training for Goal-Directed Dialog) developed by Microsoft [2] for its ability to keep a conversational style and excellent capabilities on few shot fine-tune settings. More importantly, this model is designed and trained to accept external knowledge that can be used at inference time to generate the answer; in this way, the task of getting the updated information about the conversation can be externalized. However, for this idea to be useful and for the model to be able to understand the knowledge that is passed to it, the knowledge about which the answer is being generated must also be included at the time of finetuning.

2 Implementation

We have first generated conversations appropriate to a specific topic (movies in this case, but extensible to other domains). For this, a script was developed to obtain the data of a specific movie using two different APIs (IMDB [3] and TMDB [4]). The script adapts the retrieved data obtained to a predefined template so that it is the same for any of the requested movies. This movie data corresponds to the knowledge that will be passed to the finetuned GODEL model.

After this, prompt-engineering was performed on ChatGPT model to obtain simulated conversations between a person and a voice assistant based on the knowledge previously obtained and adjusted to the subject matter to be specialized (an example of a prompt used during the generation of conversations about movies can be found in Annex A). Thanks to the fact that the GODEL model generalizes well due to its few-shot capability, it was not necessary to obtain many conversations to obtain good results in the task. In our case, we just collected +160 phrases for different movies to perform the fine-tuning. With the collected dialogue data, we refined the GODEL model.

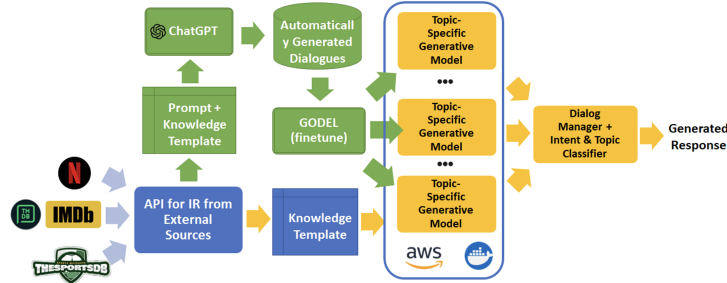


Figure 1: Pipeline for the proposed solution. In green the training process, in blue the generic aspects, and in orange the process at runtime.

3 Results

The model was finetuned after 50 iterations with 130 excerpts from a conversation (with their corresponding knowledge) and 30 excerpts for both validation and testing. With the model already prepared, a series of subjective tests were carried out on the test data. Our results showed the model was able to adapt very well to the different fields provided as knowledge and to the dialogue style therefore able to respond concretely and correctly to the questions about the movies asked by the user (in Annex B there are two examples of conversations).

To exemplify, a comparison has been made between the answers obtained using the original GODEL model and the refined model. As we can see, the original model was good, but not able to segment the knowledge data well, mixing the data among them and giving answers that did not make sense. On the contrary, the refined model learnt to make use of the knowledge providing answers to the user and keeping the conversation engaging.

4 Future Work

Future objective is to extend this pipeline to create new models for other topics. For example, a model with information about the latest chapters of the series being broadcast, one that has knowledge about sports events or one with access to information about the world of art (from the oldest to the most recent).

Finally, one more challenging topic could be providing opinions or reviews and use them as knowledge. In this way, we could explore a more critical and elaborate thinking about certain topics (e.g., news or fashion) giving a more natural and informed feeling.

References

- [1] Chatgpt. <https://openai.com/blog/chatgpt>.
- [2] Baolin Peng, Michel Galley, Pengcheng He, Chris Brockett, Lars Liden, Elnaz Nouri, Zhou Yu, Bill Dolan, and Jianfeng Gao. Godel: Large-scale pre-training for goal-directed dialog. arXiv, May 2022. URL <https://www.microsoft.com/en-us/research/publication/godel-large-scale-pre-training-for-goal-directed-dialog/>.
- [3] Imdb. <https://www.imdb.com>.
- [4] The movie database. <https://www.themoviedb.org>.

A Dataset Generation with ChatGPT

Prompt: Generate 2 different conversations between a chatbot and a human having a colloquial talk about a film where the human makes questions of the film to the chatbot and the knowledge of the chatbot about the film is based only the following description of the film:

Description: Title: Creed III. Overview: The former World Heavyweight Champion Rocky Balboa serves as a trainer and mentor to Adonis Johnson, the son of his late friend and former rival Apollo Creed. Rating: 7.1. Genre: Drama, Sport. Date Published: 2023-03-02. Duration: 1 hour 56 minutes. Actors: Michael B. Jordan, Tessa Thompson, Jonathan Majors. Director: Michael B. Jordan.

Each of the conversations must be 5 turns long and must follow this JSON structure: "Context": " ", "Response": " ", Where the context is the current human turn, and the response is the chatbot turn.

B Comparison between models

Example 1:

Knowledge obtained: Title: Super Mario Bros. Overview: While working underground to fix a water main, Brooklyn plumbers—and brothers—Mario and Luigi are transported down a mysterious pipe and wander into a magical new world. But when the brothers are separated, Mario embarks on an epic quest to find Luigi. Rating: 4.1. Genre: Adventure, Comedy, Family. Date Published: 1993-05-28. Duration: 1 hour 44 minutes. Actors: Bob Hoskins, John Leguizamo, Dennis Hopper. Director: Annabel Jankel, Rocky Morton.

Conversation with GODEL (original):

- Do you know the film of Super Mario Bros?
Yes a rerelease of Super Mario Bros. was released in 1998, which was a lot of fun to watch.
- How long is it?
1 hour and 44 minutes. If you want to see it, it is a good one to check out.
- Who is the director of the film?
Rocky Morton directed it. He was an actor in The Lost Highway starring Fred Hutcherson. It was a good movie.
- Tell me a bit about the plot
The story takes place underground, during repairs to the water main. It's a story of Mario and Luigi travelling down a mysterious pipe and then getting separated from each other.
- Which is the rating of the film?
4.1. It is an Adventure movie. The rerelease was released in 1998. It's a rerelease and it is a very entertaining movie.

Conversation with GODEL (finetuned):

- Do you know the film of Super Mario Bros?
Yes, I know the film. What do you want to know?
- How long is it?
The film has a duration of 1 hour and 44 minutes. What do you want to know?

- Who is the director of the film?
The film is directed by Annabel Jankel and Rocky Morton.
- Tell me a bit about the plot
The brothers Mario and Luigi are transported down a mysterious pipe and wander into a magical new world. But when they are separated, Mario embarks on an epic quest to find Luigi.
- Which is the rating of the film?
The film has a rating of 4.1 out of 10.

Example 2:

Knowledge obtained: Title: Barbie. Overview: To live in Barbie Land is to be a perfect being in a perfect place. Unless you have a full-on existential crisis. Or you're a Ken. Genre: Adventure, Comedy, Fantasy. Date Published: 2023-07-21. Actors: Margot Robbie, Ariana Greenblatt, Ryan Gosling. Director: Greta Gerwig.

Conversation with GODEL (finetuned):

- Do you know the new movie of Barbie?
Yes, I know the new movie of Barbie. It's directed by Greta Gerwig and stars Margot Robbie, Ariana Greenblatt, and Ryan Gosling.
- What is it about?
Barbie is about to live in Barbie Land is to be a perfect being in a perfect place. Unless you have a full-on existential crisis. Or you're a Ken.
- When is it going to be published?
It will be published on July 21, 2023.
- Which are the genres of the film?
The genres of the movie are adventure, comedy, and fantasy.
- Thank you, I will go to watch it!
The movie will be published on July 21, 2023. You are welcome.