

# ECE1724 Project – Gollum Chatbot

Sebastian Tampu  
*Department of Electrical & Computer Engineering*  
*University of Toronto*  
Toronto, Canada  
seb.tampu@mail.utoronto.ca

**Abstract—** This paper presents the development and challenges of creating a chatbot that mimics Gollum, a character from J.R.R. Tolkien's "The Lord of the Rings." Using the DialoGPT model, we aimed to capture the unique language style and personality of Gollum in a conversational AI. The project involved curating a dataset from the character's dialogues, fine-tuning the model for character-specific responses, and addressing technical challenges such as data preprocessing and model training on a CPU setup due to hardware limitations. Despite facing issues with input padding affecting response accuracy, the chatbot demonstrates the feasibility of creating character-driven AI interactions. This work contributes to the field of artificial intelligence and natural language processing by exploring the nuances of character-based dialogue generation, offering insights for future research in creating more lifelike and engaging AI personalities.

## I. INTRODUCTION

This project explores the development of a chatbot modeled after Gollum, a character from J.R.R. Tolkien's "The Lord of the Rings." The choice of Gollum was motivated by his unique speech pattern and complex personality, making him an ideal subject for testing the capabilities of AI in character simulation. Unlike standard conversational models, Gollum's distinctive dialogue offers a clear measure of the chatbot's performance. The goal was not just to mimic Gollum's way of speaking, but to capture aspects of his psyche, despite acknowledging the challenges in achieving a comprehensive portrayal.

The project's broader aim ties into the fascination with using AI for creating digital personas that could interact with users as if they were real or fictional characters. This ambition extends beyond textual interactions to potential voice and even visual communications, bringing closer the ideas of conversing with characters from literature or personal memories. Through Gollum, this project delves into the technical and creative aspects of building AI-driven personalities, laying the groundwork for future advancements.

## II. RELATED WORK

In exploring the fusion of artificial intelligence with the nuanced art of language, this project draws inspiration and foundational knowledge from a broad spectrum of previous work in chatbot technology and natural language processing. The advancement of chatbots, from simple scripted interactions to those capable of mimicking complex human behaviors and speech, showcases the dynamic evolution of this field. Central to our inquiry is how these digital entities can not only

comprehend but also replicate the distinct speech patterns and emotional depth of a character as intricate as Gollum from J.R.R. Tolkien's universe.

Among the pivotal studies that underpin this exploration, the work of Vinyals and Le on neural conversational models offers a cornerstone for understanding the mechanics behind engaging and coherent dialogue generation in AI systems [1]. Their model paves the way for creating more sophisticated conversational agents capable of sustaining meaningful interactions. Furthermore, the development of XiaoIce, as discussed by Zhou, Gao, Li, and Shum, represents a leap towards imbuing chatbots with a semblance of emotional intelligence and personality, aspects crucial for replicating a character with as rich an emotional tapestry as Gollum [2]. These studies not only highlight the remarkable strides made in making AI more relatable and human-like but also underscore the potential gaps in creating a chatbot that can accurately reflect the unique linguistic styles and psychological complexities of specific literary figures.

This project, therefore, situates itself at the intersection of advanced NLP techniques and the creative endeavor of character simulation. It aims to address the challenges observed in existing works, such as the replication of distinctive speech patterns and the incorporation of emotional depth in responses, by leveraging and expanding upon these foundational advancements. Through this, the project aspires to contribute to the broader discourse on the capabilities and future directions of AI in understanding and reproducing the essence of complex characters like Gollum, further bridging the gap between technology and the humanities.

## III. METHODOLOGY

### A. Data Collection

The data collection process changed three times throughout the course of the project. The initial process was to take data directly from the "Lord of the Rings" books. As is imaginable, this took a lot of time, especially since I was also trying to capture the details of context, actions, and emotions. Even without considering how long the collection would take, it would have been much more difficult to work with or build models that could make use of all the additional context. The first change was to move to a dataset that consisted of character scripts from the movies [3]. This dataset however was not well cleaned and only had about 100 lines to use for Gollum. As such, the final decision was to merge the available lines with lines from a mock conversation made with Gollum made by ChatGPT-4. This

allowed for a dataset that had enough data for a working chatbot, as well as knowledge of more topics than what is present in the movies.

### B. Model Selection

The core of the project is built on leveraging an NLP model, specifically, DialoGPT, a variant of the GPT-2 model tailored for generating human-like conversational responses. The decision to use DialoGPT over other models, including the more advanced GPT-3, was driven by several factors. Firstly, DialoGPT's architecture is designed for dialogue, making it inherently suitable for a chatbot application. Its training on diverse conversational datasets enables a broad understanding of conversational context and nuances, crucial for replicating Gollum's distinctive dialogue style. Additionally, while GPT-3 offers more power, its accessibility and computational demands make DialoGPT a more practical choice for this project, balancing performance with feasibility.

### C. Training Process

The training of the Gollum chatbot involved several key steps, from data preprocessing to fine-tuning the model:

- The first step was data preprocessing. To tailor the DialoGPT model to mimic Gollum's speech, a curated dataset comprising lines spoken by Gollum in "The Lord of the Rings" series, along with generated lines, was compiled. This dataset underwent preprocessing to ensure compatibility with the model's input requirements. Among these steps was the fetching and storing of context data, or past lines that led up to the responses, providing more depth to the data. Other preprocessing steps included tokenization using the model's tokenizer, which breaks down sentences into sub-word units that the model can understand, and formatting the data into conversational pairs to simulate the back-and-forth nature of dialogue.
- Another important part is environment setup and adjusting model parameters. The model was fine-tuned on a modest computational setup, not utilizing a GPU for training or testing. Given the relatively focused scope of the dataset, the fine-tuning process required careful adjustment of parameters to avoid overfitting. Key parameters included the learning rate, set to  $5e-5$  based on empirical evidence of its effectiveness in similar tasks, and the number of training epochs, limited to 5 to maintain a balance between learning the dataset nuances and preserving the model's general conversational abilities.
- Lastly, the fine-tuning process was highly important. Utilizing the Hugging Face Transformers library, the DialoGPT model was fine-tuned on our Gollum-specific dataset. The training process involved iteratively passing the dataset through the model, adjusting the model's internal parameters to minimize the difference between the generated responses and the expected dialogue patterns of Gollum.

### D. Challenges

The challenges involved in data collection were mentioned earlier and so will not need to be mentioned here. The other significant challenge, which is still not solved, deals with padding. The model expects left-padding for any text inputs. Originally, the assumption was that the issue was found in collation functions inside training and evaluation functions. However, the model had no issue with the overall training and learning. As such, focus was shifted to user input when interacting with the chatbot. Several attempts were made to fix the issue, such as explicitly setting the padding style when initializing the tokenizer. Despite this, no solution was found. This issue seems to effect the ability of the model to fully understand the input and thus to return a proper response.

## IV. IMPLEMENTATION

### A. System Architecture

The chatbot's system architecture is designed to encapsulate the complexity of Gollum's character while ensuring efficient and responsive interactions. The system is structured around four main components: the Input Processor, the Dialogue Manager, the Model Core (DialoGPT), and the Output Processor. This architecture facilitates a seamless flow of information through the system, from receiving user input to delivering Gollum's responses.

The Input Processor is the system's entry point. It handles raw user inputs, performing necessary preprocessing to transform the inputs into a format suitable for the model. Preprocessing steps include tokenization, where user inputs are broken down into tokens (sub-word units recognized by the model), and encoding, which converts tokens into numerical IDs that represent each token in the model's vocabulary. This component ensures that even complex or out-of-the-ordinary inputs are coherently understood by the system.

Central to the system, the Dialogue Manager orchestrates the conversation flow and maintains context. It stores the history of the conversation to provide context for Gollum's responses, enabling the model to generate replies that are not only relevant to the immediate input but also coherent with the overall conversation. The Dialogue Manager decides when to invoke the Model Core for generating responses and when to terminate the conversation, based on cues like user commands or conversation length.

At the heart of the architecture is the DialoGPT model, fine-tuned to embody Gollum's unique speech patterns and personality traits. The Model Core generates responses based on the processed input and conversation context provided by the Dialogue Manager. Utilizing a Transformer-based architecture, it predicts the next token in the sequence, iteratively generating Gollum's response one token at a time until a coherent reply is formed or the maximum token limit is reached. The model's parameters and architecture are optimized for generating conversational text, enabling it to produce responses that are characteristically Gollum-like in syntax, semantics, and style.

The final component, the Output Processor, takes the generated response from the Model Core and prepares it for presentation to the user. This includes decoding the token IDs

back into text, cleaning up the response to remove any non-characteristic elements (such as unwanted tokens or syntax errors), and formatting the text to enhance readability. The Output Processor ensures that the final response not only sounds like Gollum but is also presented in a user-friendly manner.

### B. Algorithm Details

The core of the Gollum chatbot's functionality is driven by sophisticated algorithms and techniques designed for natural language understanding and generation. These components are integral to transforming user inputs into character-consistent outputs, effectively mimicking Gollum's unique speech patterns and conversational style. Below, are the specifics of the algorithms employed in the system.

The first component is Natural Language Understanding (NLU). At the inception of each interaction, understanding the user's input is crucial. This is managed through a combination of tokenization and encoding processes facilitated by the DialoGPT's tokenizer. The tokenization process breaks down the raw user input into manageable pieces called tokens. These tokens can be words, parts of words, or even punctuation, depending on the tokenizer's granularity. The Gollum chatbot, employs a sub-word tokenization scheme that balances the vocabulary size with the ability to handle rare or novel words, characteristic of Gollum's language. Following tokenization, the tokens are converted into numerical IDs. This encoding process maps each token to a unique identifier understood by the model, facilitating the subsequent processing steps. The DialoGPT model utilizes a pre-defined vocabulary for this purpose, ensuring consistency in token representation.

The next component is language generation. In the language generation phase, the Transformer-based architecture of DialoGPT employs a decoder mechanism to craft responses, utilizing self-attention mechanisms for contextual understanding and coherence in mirroring Gollum's distinctive speech. Initially, it follows a greedy decoding strategy to choose the most probable next token, but to truly capture the unpredictability inherent in Gollum's language, the process is enriched with stochastic elements like top-k sampling and temperature-controlled sampling. These introduce a controlled randomness into the responses by limiting choices to the top most likely next tokens and adjusting the probability distribution for next token selection, respectively. This blend of strategies ensures the chatbot's responses not only remain true to Gollum's erratic nature but also maintain a coherent narrative flow, adding layers of realism and unpredictability to the interaction.

Lastly, there is optimization and fine-tuning. To adapt DialoGPT to embody Gollum's unique mannerisms and dialogue style, we directed the model's learning with lines drawn from "The Lord of the Rings." This fine-tuning phase is essential for encouraging the model to prefer Gollum's peculiar expressions over generic responses, embedding the character's distinctive nuances more profoundly. Central to this phase is the minimization of a loss function that quantifies the differences between the model's predicted sequences and the genuine dialogues of Gollum. By employing backpropagation and gradient descent techniques, we meticulously adjust the model's parameters, ensuring that the generated language not only

resembles Gollum's speech in accuracy but also aligns with his unique character traits.

### C. Technical Specifications

The components used had a role in determining which model to use, as mentioned earlier, as well as dataset size to an extent. The major components are listed below:

- *CPU*: AMD Ryzen™ 7 5700X
- *GPU*: AMD Radeon™ RX 6750 XT
- 32GB RAM

Given the setup, using CUDA was not an option. As a result, the project was limited to CPU utilization and slower training.

## V. RESULTS AND DISCUSSION

Quantifying the results of a chatbot or speech generation does not seem like good practice by itself. There are many nuances to consider in evaluating the output that numbers might not do justice to. As such, I will be combining performance values with personal impressions.

The effectiveness of training and evaluation was mainly measured through perplexity scores. Several scores are shown in Table 1 below.

TABLE I. PERPLEXITY SCORES FOR VARIOUS HYPERPARAMETERS

Hyperparameters			Perplexity
<i>Context lines</i>	<i>Epochs</i>	<i>Batch Size</i>	
3	3	4	11.02
7	3	4	9.23
7	5	4	6.83
7	5	2	6.12
7	5	1	5.86

The more context lines and epochs were increased, and the more batch size decreased, the better the model learned as seen through the decreasing perplexity scores. This is expected behavior, and so a larger dataset would have helped improve the training even more, despite also meaning a longer training time given the CPU setup. As I mentioned, while the numbers can give a sense of improvement, the text should also be seen. Below is a brief conversation with the Gollum chatbot:

- *User*: Hello Gollum!
- *Gollum*: And my precious, I've been searching for you, What has life come to? Your precious. Now, go back to your cave. It is your destiny. Go in it. You will see. Go back.
- *User*: What's your favourite meal?
- *Gollum*: 's a nice fish, yes. But in the end, it doesn't even matter. We just wanted to be friends, not pets, precious.
- *User*: Do you like the sun?

- *Gollum*: It burns us, yes, but the shadows make us happy, yes'

It is clear that the style of speech of the chatbot is Gollum's, including his usage of words like "yes" and "precious", and referring to himself as "us." The context in the replies is at a good starting point, but can clearly improve as seen through the first response. Sometimes the bot tends to go on unnecessary long tangents that make apparent the need for further improvement.

## VI. CONCLUSION

This project set out to create a chatbot that emulates Gollum from J.R.R. Tolkien's "The Lord of the Rings." By leveraging the DialoGPT model, the project focused on generating dialogue that captures Gollum's unique speech patterns and erratic behavior. Throughout the development, challenges, such as data collection and model fine-tuning, were faced, but they were addressed through a meticulous approach to dataset preparation and model parameter adjustment.

The significance of the work, at least to me, lies in its demonstration of AI's potential to replicate complex character speech in interactive applications. By fine-tuning a model specifically around Gollum's dialogue, it was shown that it is possible to narrow the focus of broad AI systems to reflect specific character traits, contributing to the fields of AI and NLP by providing insights into character-driven dialogue generation.

However, the project also encountered limitations, primarily the issues around input padding that affected the model's ability to understand user inputs effectively. Despite this, the project underscores the importance of ongoing experimentation and refinement in AI dialogue systems.

In summary, while this project has advanced my understanding of AI-driven character simulation, it also

highlights the necessity for continuous exploration in the field. The balance between model complexity and practical implementation presents an ongoing challenge that future research must address. My journey into the realm of AI and NLP, embodied by Gollum's chatbot, serves as a foundation for further exploration and innovation in creating digitally interactive personas from literary and fictional worlds.

## VII. FUTURE WORK

In the future, the chatbot will be refined to achieve a more nuanced understanding and reproduction of Gollum's dialogue, aiming for more accurate and context-aware responses. Key areas of focus include improving data collection to expand the conversational scope and depth, optimizing the model to better handle the character's unique speech patterns, and exploring more advanced NLP techniques for enhanced understanding of complex dialogues. Additionally, there is an aim to address the left-padding issue to improve the model's performance and explore the integration of voice and possibly visual elements for a more immersive interaction with Gollum. Finally, expanding this project to include other characters from literature or historical figures could further demonstrate the potential of AI in bringing fictional and historical personalities to life, opening new avenues for education, entertainment, and personal engagement.

## REFERENCES

- [1] O. Vinyals and Q. V. Le, "A Neural Conversational Model," arXiv preprint arXiv:1506.05869, 2015.
- [2] L. Zhou, J. Gao, D. Li, and H.-Y. Shum, "The Design and Implementation of Xiaolce, an Empathetic Social Chatbot," *Computational Linguistics*, vol. 46, no. 1, pp. 53-93, 2020.
- [3] Kaggle, "Lord of the Rings Data," Kaggle, accessed March 18, 2024, <https://www.kaggle.com/datasets/paultimothymooney/lord-of-the-rings-data>.