



Conversations with Characters in Stories for Literacy - Quick, Customized Persona Bots from novels

David Aristovnik, Tara Rolih, Ana Strmčnik

Abstract

This project addresses the declining literacy rates among young readers by leveraging character conversations from novels and stories to create engaging educational tools. Through corpus analysis, we aim to understand the dynamics of textual conversations between characters through their attributes and develop personaBots that embody the specific characters. These personaBots are designed to interact with users in a meaningful way, improving their literacy skills and making reading more appealing. Our approach includes reviewing related works on data augmentation, dialogue agents, and pedagogical agents with personality theory, as well as comparing existing services for creating persona bots. We utilize datasets such as the BookCorpus and The Gutenberg dataset for pretraining and fine-tuning our models on specific characters. The project involves preprocessing steps focused on steps like dialogue extraction and named entity recognition. We intend to use models from the HuggingFace library, augmented with retrieval methods to enhance the bot's knowledge and human interaction capabilities. Our evaluation strategy includes comparing our model's performance with other models and assessing its understanding of character-specific information and ability to simulate human-like conversations.

Keywords

Literacy Enhancement, Conversational Agents, PersonaBots, Data Augmentation, Retrieval-Augmented Generation

Advisor: Slavko Žitnik

Introduction

Because of a growing concern over declining literacy rates among young readers, innovative solutions are required to engage and improve their reading habits. This project introduces a corpus analysis of character conversations within suggested stories and novels, aiming to understand how these interactions can be leveraged to create engaging educational tools. By analyzing textual conversations between characters, we seek to develop a framework for creating personaBots, conversational agents that embody characters from novels. These personaBots are intended to motivate young readers by allowing them to interact with the characters in a meaningful way, helping to make reading a more appealing and engaging activity. Through this analysis, we aim to research the effort of enhancing literacy skills, making reading a more enjoyable and frequent part of young people's lives.

Related work

We reviewed the provided literature [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] and looked at additional literature to get a better grasp of what is currently being done in the field and what we can do to improve it.

The [6] paper presents PEDANT, a novel approach for data augmentation in modeling human personality, leveraging generative pre-trained models and domain expertise to generate high-quality synthetic data for rare personality types, exemplified through the "Dexter Machine".

The paper [11] presents an approach towards constructing human-like dialogue agents by introducing Human Level Attributes (HLAs) based on tropes. These tropes, derived from recurring character traits in fiction recognized by viewers, facilitate the development of dialogue agents that mimic the personalities of fictional characters. The authors propose a dataset combining detailed HLA data with character-specific dialogue, enabling agents to learn language styles from specific characters through HLAs. They introduce ALOHA, a

three-component system consisting of character space mapping, character community detection, and language style recovery, which collectively build a character-specific language model.

We came across an intriguing article [12], which integrates pedagogical agents with personality theory to incorporate emotion and emotional states. They also emphasised that believability is crucial in developing persona-like pedagogical agents, as learners expect human-like behavior such as emotional engagement and empathy, and failure to meet these expectations diminishes trust and ruins the interaction. We will try to make our persona bots follow the same principles.

We also came across an article [13] that provides valuable insights into enhancing language models with external knowledge, specifically through a Retrieval-Augmented Generation (RAG) method. This approach integrates a pre-trained language model with a vast external knowledge base from Wikipedia, significantly enriching the model's outputs in knowledge-intensive tasks. Incorporating this method into our project could substantially elevate the effectiveness of our persona bots. By leveraging this technology, the bots can offer more accurate, contextually relevant, and engaging interactions, thereby fostering a deeper literary engagement and addressing the ongoing literacy crisis.

We've considered integrating the insights from an article [14] that introduces the Forward-Looking Active Retrieval augmented generation (FLARE) method. This innovative approach, which enhances language models by actively retrieving relevant information during the text generation process, could significantly enrich our persona bots. By implementing FLARE, we aim to empower the bots with the ability to provide more contextually rich, accurate, and engaging interactions.

We also compared two of the most commonly used services capable of producing quick, customized persona bots from novels: ChatGPT [15] and Character AI [16]. Character AI offers the option to create a character from your description, a feature also available with ChatGPT. Both services provide answers of similar quality using this approach. Subsequently, we tested both services using a well-established fictional character from a novel, arguably the most famous one, Harry Potter. We obtained the most popular Harry Potter character from Character AI [17] and requested ChatGPT to act as Harry Potter when responding. We posed the following questions to both:

- What was the most horrible experience you ever had?
- When were you born?
- Who are your best friends?

Character AI's response to the first question did not include actual data (his ex-girlfriend was cheating on him), while ChatGPT provided a much more believable answer (his friend Cedric Diggory dying in front of him). Additionally, we found the use of parentheticals in Character AI's responses

to be unnecessary. ChatGPT tends to provide longer answers, typically packed with more information about characters compared to Character AI.

The responses provided by both services were satisfactory in terms of accuracy and demonstrated a level of human-like understanding.

Based on our findings, we concluded that we would prefer our persona bots to resemble ChatGPT rather than Character AI.

Data

We have discovered the BookCorpus, a publicly available dataset compiled by Soshikobayashi [18], containing a collection of novels authored by unpublished writers. This dataset consists 11,038 books, totaling approximately 74 million sentences and 1 billion words, spanning across 16 different sub-genres like Romance, Historical, and Adventure.

We also found an alternative, The Gutenberg dataset [19] that comprises over 15,000 book texts, along with their respective authors and titles.

Our plan is to initially pretrain our model using one of the datasets and fine-tune it on specific novels and characters of interest. By adopting this approach, we aim to assess whether this sequential training methodology yields improved performance. As we rely on novel and story recommendations from our advisor, our strategy may evolve based on the provided data.

Preprocessing

Our preprocessing does not have to be focused on correcting the text, as our data will come from books that have already been professionally proofread.

- **Text extraction** - exclude irrelevant data like page numbers, headers, footers
- **Noise removal** - remove irrelevant data like special characters, white spaces and consistent case (lowercase usually)
- **Tokenization** - split text into sentences and words
- **Part-of-speech (POS) tagging** - assign parts of speech to each word (e.g. noun, verb, adjective)
- **Lemmatization** - convert words to their base or root form
- **Named entity recognition (NER)** - identify and tag named entities (characters, places, etc.)
- **Sentiment analysis** - determine sentiment expressed in sentences or paragraphs.

The most important part would be **dialogue extraction**. We want to try and extract dialogue from the given text and assign it to the character that said it, following a similar principle that was used in this article [11].

Project idea

At present, our understanding of advanced language processing machine learning models is limited. Therefore, we have opted to explore the best options for preprocessing, feature extraction, modeling, and evaluation through experimental testing.

Initially, our objective is to construct a basic model to familiarize ourselves with the dataset and algorithms used. We will start with preprocessed text and then proceed with a pretrained model from the HuggingFace library [20]. Our plan involves constructing a model for each novel or character, based on our requirements, and subsequently fine-tuning that model using retrieval augmented generation (RAG) [13, 14] or by implementing more efficient vector searching or similarity computation approaches.

For evaluation, we will try comparing our model's output with that of other similarly functioning models. We will also assess if our model possesses basic knowledge about characters, such as their age, appearance, etc. Furthermore, we will test its capability to simulate human conversations effectively.

We have identified the following key aspects for our persona bot's functionality:

- **Human-like Interaction** - Our bot's responses should closely resemble those of humans. To achieve this, we intend to select a pretrained model with such capabilities.
- **Self-Knowledge** - Our bot should possess information about itself or specific characters, including their manner of speech and relevant details. We aim to accomplish this through methods like RAG.

As we progress, our aim is to enhance our results through iterative development, conducting experiments, and gaining deeper insights into our project understanding.

References

- [1] Mehdi Alaimi, Edith Law, Kevin Daniel Pantasdo, Pierre-Yves Oudeyer, and Hélène Sauzeon. Pedagogical agents for fostering question-asking skills in children. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.
- [2] Suzanne TM Bogaerds-Hazenbergh, Jacqueline Evers-Vermeul, and Huub van den Bergh. What textbooks offer and what teachers teach: an analysis of the dutch reading comprehension curriculum. *Reading and writing*, 35(7):1497–1523, 2022.
- [3] Laura C Engel and Matthew O Frizzell. Competitive comparison and pisa bragging rights: sub-national uses of the oecd's pisa in canada and the usa. *Discourse: Studies in the Cultural Politics of Education*, 36(5):665–682, 2015.
- [4] Lewis R Goldberg. An alternative “description of personality”: The big-five factor structure. In *Personality and Personality Disorders*, pages 34–47. Routledge, 2013.
- [5] Jane Murray. Literacy is inadequate: young children need literacies, 2021.
- [6] Yair Neuman, Vladyslav Kozhukhov, and Dan Vilenchik. Data augmentation for modeling human personality: The dexter machine. *arXiv preprint arXiv:2301.08606*, 2023.
- [7] Thijs MJ Nielen, Glenn G Smith, Maria T Sikkema-de Jong, Jack Drobisz, Bill van Horne, and Adriana G Bus. Digital guidance for susceptible readers: Effects on fifth graders' reading motivation and incidental vocabulary learning. *Journal of Educational Computing Research*, 56(1):48–73, 2018.
- [8] Ioannis Papaioannou et al. *Designing coherent and engaging open-domain conversational AI systems*. PhD thesis, Heriot-Watt University, 2022.
- [9] Zhenyu Wu, YaoXiang Wang, Jiacheng Ye, Jiangtao Feng, Jingjing Xu, Yu Qiao, and Zhiyong Wu. Openicl: An open-source framework for in-context learning. *arXiv preprint arXiv:2303.02913*, 2023.
- [10] Rolf A Zwaan, Gabriel A Radvansky, Amy E Hilliard, and Jacqueline M Curiel. Constructing multidimensional situation models during reading. *Scientific studies of reading*, 2(3):199–220, 1998.
- [11] Aaron W Li, Veronica Jiang, Steven Y Feng, Julia Sprague, Wei Zhou, and Jesse Hoey. Follow alice into the rabbit hole: Giving dialogue agents understanding of human level attributes. *arXiv preprint arXiv:1910.08293*, 2019.
- [12] Ana Paiva, Isabel Machado, and Carlos Martinho. Enriching pedagogical agents with emotional behaviour—the case of vincent. In *AIED*, volume 99, pages 47–55. Cite-seer, 1999.
- [13] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.
- [14] Zhengbao Jiang, Frank F Xu, Luyu Gao, Zhiqing Sun, Qian Liu, Jane Dwivedi-Yu, Yiming Yang, Jamie Callan, and Graham Neubig. Active retrieval augmented generation. *arXiv preprint arXiv:2305.06983*, 2023.
- [15] OpenAI. ChatGPT: OpenAI's GPT-based chatbot. Available online at <https://openai.com/chatgpt>, 2022. Accessed on March 21, 2024.
- [16] Character AI. Character AI. <https://character.ai/>, 2022. Accessed on March 21, 2024.
- [17] Character AI. Character AI Chat. https://character.ai/chat/suAUJzAPwFm-rDAAzKByHqAN64dYBg_.lC_83ClfBzg, 2024. Accessed on March 21, 2024.

- [18] Sosuke Kobayashi. Homemade bookcorpus. <https://github.com/soskek/bookcorpus>, 2018.
- [19] Matej Bejan. 15000 gutenbergs books. <https://www.kaggle.com/datasets/mateibejan/15000-gutenberg-books>, 2021.
- [20] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Huggingface’s transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*, 2019.