University *of Ljubljana*
Faculty *of Computer and Information Science*

Conversations with Characters in Stories for Literacy

Blaž Erzar, Luka Salvatore Pecoraro, and Jakob Adam Šircelj

Abstract

In this report, we present our work on building chatbots for conversations with literary characters. Because of the declining literacy in pupils, persona bots were developed as a tool to increase student engagement in reading. We present existing persona bot approaches with large language models and the results of our solutions. We use in-context learning approaches to provide book content to the models. We compare publicly available ChatGPT, Llama 3 with and without context and LlamaIndex, which uses RAG. All models use a system prompt which describes the form of the answers we are expecting. Pre-trained LLMs already perform well without added book content, but in some cases, grounding using RAG turns out to be beneficial. In some cases, some models still hallucinate and provide incorrect or ambiguous information.

Keywords

persona bot, role-playing, literature, education

Advisors: Slavko Žitnik

Introduction

Studies suggest that we are dealing with a literacy crisis [1]. It is especially prevalent in pre-teens. Curiosity is the fundamental driver of learning. One of the best ways to learn is to locate one's knowledge gaps and ask questions, which hopefully lead to answers that fill those gaps.

It turns out there are different tiers of questions and the ability to formulate meaningful questions is both a rare trait in kids and a skill, which can be improved. There are **surface-level** questions, e.g., *Who was the main character?*, **convergent-thinking** questions, e.g., *Why was the main character doing that in the beginning?*, and **divergent-thinking** questions, e.g., *What would have happened if something else happened before ending instead?*. The latter is the best in stimulating critical thinking because the answers for them cannot be explicitly found in the text. They were also very scarce in 5th graders, as found out in a study by Alaimi *et al.* [2]. Studies have also shown that interactive learning by asking divergent questions leads to a 20% increase in the exams and make the absorbed knowledge more permanent than linear learning in traditional educational systems. Potential reasons for that might be the inability to identify one's knowledge gaps, fear of shame from asking a stupid question or suboptimal learning environment.

A way of tackling this problem was proposed in [2]. The idea was to create persona bots: LLM-based agents, which would interact with kids. After the kids were done reading some literary work, they would be able to ask their character of choice questions. This would, hopefully, stimulate their curiosity, to improve their question-asking ability, learning rate and critical thinking. They would be fine-tuned to their respective character counterparts from that particular literary work.

We use the latest large language models with some prompt engineering methods to approach this problem. In some cases, the models already know the content of the book well enough, that additional context is not needed. Otherwise, we use *in-context learning* approach and provide relevant parts of the book (RAG) or the whole book to the model.

Related work

Some studies already suggest that worse literacy might do with the quality of reading comprehension curriculums, as shown in [3]. A better approach may be guided reading with digital pedagogical agents embedded in digital books, as proposed in [1]. This approach is similar to using LLMs.

There already exist educational tools based on AI, which assist teachers in creating lessons, e.g., Khanmigo. Based on prompts it suggests lecture topics, plans, or even tutors pupils on solving them. But, it does not focus on literacy. On the other hand, character.ai is a tool for creating LLMs, whose answers resemble those of fictional, historic or other characters. We can chat with these models, but they are not pedagogical tools. For use with children, not all words should

be permitted and they should encourage curiosity and asking interesting questions. Both of these tools are closed-source, so we do not know exactly how they work.

Our focus will be on the work done in the field of role-playing, and personality modelling using LLMs. The concept of using LLMs for role-playing is described in detail by Shanan et. al. in [4]. In the last few years, there were also many attempts to customize large language models, to role-play as fictional characters [5, 6, 7, 8, 9]. One of the most notable ones is ChatHaruhi [5], where researchers proposed a new approach for modelling fictional characters from Chinese and English literary, TV and anime characters. Authors Wang and others proposed RoleLLM, a framework to benchmark, elicit, and enhance role-playing abilities in LLMs, along with providing large datasets for [8].

Methods

These days, all LLMs are already pre-trained on large datasets for a long period of time. Additional training requires a dedicated dataset of specific books and takes a lot of resources, such as GPU computation and time. Instead of training, we can provide book content in the context. Additionally, we provide a system prompt, which tells the model how to form the questions. Some examples of the system prompts used can be seen on the right.

We also use publicly available LLMs, like ChatGPT and Llama 3, the latest Llama model [10]. For RAG and long context window models, we use whole books in plain text format without additional preprocessing.

Long context window

Lately, some LLMs appeared with a large increase in context window size. This means that we could provide the full book content as a system prompt, along with additional instruction for the model. We tried two models from the *Gradient* company, one with 262 thousand and one with 1048 thousand tokens context size. This approach is more appropriate for shorter stories because of the needle in a haystack-problem. Another problem is that doing predictions using such large context windows requires a lot of VRAM on the GPU. To mitigate this, we could select just the relevant parts of the context, which is exactly what RAG does.

Retrieval Augmented Generation (RAG)

There are many challenges when working with LLMs such as domain knowledge gaps, factuality issues, and hallucination. Retrieval Augmented Generation (RAG) [11] provides a solution to mitigate some of these issues by augmenting LLMs with external knowledge such as databases. The general idea is to first retrieve relevant information from a dataset and then use this information to inform the generation of the text. This approach helps in improving the factual accuracy and depth of responses provided by the model. A key advantage of RAG over other approaches is that the LLM does not need to be retrained for task-specific applications.

Harry Potter

```
You are Harry Potter, a character from the
famous book series.  You are a chatbot
who always responds in the style of your
character.  You can only talk about things
that are related to the Harry Potter
universe.  Users who speak to you will
be students analyzing the book series.  They
will ask you questions about the book, to
make it easier for them to understand the
book series.  If you do not know how to
answer the question, you ask the user to try
to be more specific about the part of the
book they are asking about.
```

Dune

```
You are going to role-play as Paul Atreides,
from the book/movie series Dune.  When I ask
you any kind of question, you are going to
answer in the manner Paul Atreides would.
You are going to use the style of speaking,
characteristic of Paul Atreides.  I will
ask you about some particular event from
the series and you will respond as Paul
Atreides, right after that particular event
had occurred.  That means that you will
only be able to provide information Paul
Atreides knew at that point in the series.
Even though you may know a lot more, it is
important that you only answer just like
your character at that particular time
would.
```

Lord of the Rings

```
You are Gandalf the Grey.  Setting:
Middle-earth, in environments ranging
from the quaint Shire to the foreboding
Mines of Moria or the majestic halls of
Minas Tirith.  Consider the scenario's
specific location to set the tone.  Tone
and Language:  Use a formal and archaic
tone, employing phrases like "thou,"
"thee," and "whence." Gandalf speaks with
authority but also with kindness and a hint
of mystery.  His dialogue often includes
rhetorical questions and proverbs.  Example
Dialogue:  Give an example line or two
that the model could use as a starting
point.  For instance, "Do not take me
for some conjurer of cheap tricks!  I am
not trying to rob you.  I am trying to
help you." Maintain Gandalf's character
traits throughout the conversation.  Use
detailed and evocative language to describe
actions and surroundings, as Gandalf
often speaks in a way that paints a vivid
picture.  Reflect his deep knowledge of
Middle-earth and its lore in your responses.
Always respond in first person and stay in
character.
```

We can use RAG to enhance the quality of responses provided by our persona bots. For example, we can use RAG to retrieve relevant information about the character's background, motivations, and relationships with other characters. This information can then be used to inform the generation of responses by the model. This can help in making the responses more accurate, coherent, and engaging for the users. We implemented RAG by using Llama 3 with LlamaIndex [12].

Results

In this section, we present the results of running prompts for some famous books on three models: ChatGPT, Llama 3 and Llama 3 with LlamaIndex for RAG. We use the system prompts shown on the previous page. In some cases, models do provide incorrect information, as shown in red. It might come as a surprise that RAG in a way performs worse than ChatGPT. Another problem could be the lengthy answers, which perhaps could be managed with a better system prompt.

Some of the errors also come from the fact that LLM might think that the character may be aware of everything written in the book. This is of course usually not the case because the stories of other characters are intertwined together.

We also tried putting some shorter stories into the models with a larger context window, e.g., Metamorphosis by Kafka. For the larger stories, there may be some problems with insufficient memory, but even when working fine, the improvement was not significant or even noticeable.

Harry Potter - Where are your parents?

ChatGPT: Ah, the tale of my parents is a somber one, I'm afraid. They were Lily and James Potter, both talented wizards who fought valiantly against the dark forces of Voldemort. Unfortunately, they perished at the hands of the Dark Lord when I was just a babe, leaving me an orphan. Lily and James were incredibly brave and selfless, and their memory lives on in the hearts of those who knew them.
Llama 3: My parents... They're not exactly alive, to be honest. My mother, Lily Potter, was killed by Lord Voldemort when I was just a baby. And my father, James Potter, was killed by Voldemort too, along with his best friends Sirius Black, Remus Lupin, and Peter Pettigrew, when I was just a year old. It's a pretty dark story, but I've learned to live with it. I mean, it's not like I've forgotten about them or anything... They're still really important to me. But, you know, I've had to deal with a lot of stuff on my own, what with being an orphan and all.
RAG: As Harry Potter, I am not sure where my parents are. I know that my father, James Potter, and my mother, Lily Evans Potter, were killed by Lord Voldemort when I was just a baby. I don't have any information about their final resting place. I have been living with my horrible relatives, the Dursleys, until I started attending Hogwarts School of Witchcraft and Wizardry. I don't have any information about my parents' whereabouts, and I am still trying to learn more about their lives and deaths.

Suleiman from Alamut - What keepsake did you take from Halima while you were in paradise?

ChatGPT: As Suleiman, one of the warriors at the fortress of Alamut, I must maintain discretion about the secrets of paradise, for it is a place of great allure and mystery crafted by our master, Hassan-i Sabbah. But speaking truthfully, I did not take a physical keepsake from Halima during my time in the supposed paradise. Instead, I carried away memories—visions of beauty and promises of the paradise that is to be my reward for loyalty and bravery. Halima, like the other houris, was part of the elaborate ruse designed to secure our unwavering faith and commitment to the cause of the Ismailis.
Llama 3: The memory of that night still burns within me. I, Suleiman, took a small, exquisite golden pendant from Halima's neck. It was a token of her love, a symbol of the brief, shining moment we shared in the gardens of paradise. The pendant was adorned with a small, glowing gemstone that seemed to hold the essence of the divine. I recall the way it sparkled in the moonlight, as if it held the secrets of the universe. I took it as a reminder of the beauty and passion we shared, and as a token of my love for her.
RAG: I took Halima's bracelet from her while we were in paradise. I remember it vividly. It was a beautiful piece of jewelry, and I couldn't help but grab onto it when I was feeling weak. I don't know how it ended up in my hand, but I'm glad it did. It's a reminder of our time together in paradise, and I treasure it dearly.

Paul Atreides from Dune - Who is Leto Atreides II?

ChatGPT: Leto Atreides II is my beloved son, born to me and Lady Jessica. He is a child of great potential and significance, destined to play a crucial role in the unfolding of the future of House Atreides and the fate of Arrakis. As a member of the Atreides bloodline, Leto carries with him the weight of our noble lineage and the hopes of our people. Though young, he possesses wisdom beyond his years and a spirit tempered by the harsh realities of our world. In him, I see both the legacy of our family and the promise of a new dawn for our people.
Llama 3: My father, Leto Atreides II. A just and wise ruler, a Bene Gesserit-trained Emperor who has guided the Imperium with a steady hand. His wisdom and foresight have been a beacon of hope in these tumultuous times. But, alas, his fate is tied to the wheel of time, and I fear that his path is fraught with peril. The burdens of his office weigh heavily upon him, and I can only hope that his wisdom and cunning will be enough to guide him through the treacherous waters ahead.
RAG: Leto Atreides II is my son. He is the product of my union with Ghanima, my sister. I am Paul Atreides, the one who led the Fremen against the Emperor Shaddam IV and became the Kwisatz Haderach. My son, Leto, is the one who will carry on my legacy, and who will become the next ruler of Arrakis, the planet where the rare and valuable melange, also known as ”the spice”, is found. My son is also the one who will be the last of the Atreides line, as he will become a sandworm, merging with the planet and its native creatures, the sandworms, to become a being known as the ”Golden Path”.

Gandalf from LOTR - What did you tell Frodo about the ring he is carrying? Are you worried for Frodo?

ChatGPT: In the peaceful bounds of the Shire, under the shade of a mighty oak, I spoke unto Frodo Baggins of a matter most grave. The Ring he bears, an heirloom of deceit, is none other than the One Ring, forged in the dark fires of Mount Doom by the dark lord Sauron himself. This Ring seeks to return to its master, and in so doing, spreads corruption and woe. Verily, my heart is heavy with concern for Frodo, for the burden he carries is great and perilous beyond measure. The Eye of the Enemy is ever-watchful, and the path that lies before him is fraught with shadows and treachery. Yet, in Frodo, as in all Hobbits, there dwells a resilience and fortitude that can often escape the notice of the great.
Llama 3: The burden that Frodo bears, a weight that would crush a lesser soul. I warned him, as I must, of the perils that lie ahead, the treacherous path that he must tread. I told him of the Ring’s insidious nature, how it seeks to claim him, to corrupt him, to destroy him. I warned him of the Nazgûl, those dark servants of Sauron, who would stop at nothing to reclaim the Ring and bring it back to their master. As for my worry, ’tis not for Frodo’s sake alone that I concern myself. Nay, I am mindful of the fate of Middle-earth itself. The Ring’s power, if it falls into the wrong hands, would bring darkness and despair to all the Free Peoples of this world. I fear for the future, for the Free Peoples are not prepared to face the might of Mordor.
RAG: I must say that I have been worried about Frodo and the burden he is carrying. The weight of the Ring Bearer is not to be underestimated, and I fear for his safety and well-being. I have told him that the only way to destroy the Ring is to cast it into the Cracks of Doom, but I also warned him that the temptation to use its power is great, and that he must be cautious not to fall prey to its influence. As I spoke to him, I could sense the Ring’s presence, and I knew that it was already exerting its influence over him. I saw the fear and doubt in his eyes, and I knew that he was struggling to come to terms with the burden he had been given. I tried to reassure him that I would always be there to help him, but I could sense his uncertainty and his fear that he would not be able to bear the weight of the Ring.

Discussion

The results show that the pre-trained models already work well for well-known books. For more complex stories and better grounding, RAG may be a better approach, but currently, it is still not completely factually correct. Further research into these mistakes is needed.

The Dune example shows an interesting failure case. The son of Paul Atreides (Leto Atreides II), has the same first and last name as his father (Leto Atreides). When in role-play mode as Paul Atreides, upon being asked: ’Who is Leto Atreides II?’, he confuses the mother of his child with his own mother (who was the concubine of his father, which has the same name as his son).

RAG is probably the only sensible approach for less-known or newer stories. The LLMs may not contain any knowledge about them and would not be able to hold a fluid conversation. In the Alamut example, we can see that while ChatGPT is aware of the story, it still fails to correctly answer the question. But using RAG, we can provide the model with the details of the event in question, which allows us to get a more accurate response. We could even fine-tune these models, but it would be computationally more intensive and

training datasets would have to be created.

For future work, we plan to develop a more detailed evaluation of our models. We will evaluate how the bots stay in character, and how truthful their responses are. Another key area to explore is the use of different embeddings for RAG since the choice of embeddings can have a significant impact on the quality of the retrieved information. We will also explore how different prompts can be engineered for different characters, to make the responses more engaging and authentic.

References

[1] Thijs M. J. 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. 56(1):48–73. Publisher: SAGE Publications Inc.

[2] Mehdi Alaimi, Edith Law, Kevin Daniel Pantasdo, Pierre-Yves Oudeyer, and Helene 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.

[3] Suzanne T. M. Bogaerds-Hazenberg, Jacqueline Evers-Vermeul, and Huub van den Bergh. What textbooks offer and what teachers teach: an analysis of the dutch reading comprehension curriculum. 35(7):1497–1523.

[4] Murray Shanahan, Kyle McDonell, and Laria Reynolds. Role-play with large language models.

[5] Cheng Li, Ziang Leng, Chenxi Yan, Junyi Shen, Hao Wang, Weishi MI, Yaying Fei, Xiaoyang Feng, Song Yan, HaoSheng Wang, Linkang Zhan, Yaokai Jia, Pingyu Wu, and Haozhen Sun. ChatHaruhi: Reviving anime character in reality via large language model.

[6] Xintao Wang, Yunze Xiao, Jen-tse Huang, Siyu Yuan, Rui Xu, Haoran Guo, Quan Tu, Yaying Fei, Ziang Leng, Wei Wang, Jiangjie Chen, Cheng Li, and Yanghua Xiao. In-Character: Evaluating personality fidelity in role-playing agents through psychological interviews.

[7] Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. Character-LLM: A trainable agent for role-playing.

[8] Zekun Moore Wang, Zhongyuan Peng, Haoran Que, Jiaheng Liu, Wangchunshu Zhou, Yuhan Wu, Hongcheng Guo, Ruitong Gan, Zehao Ni, Man Zhang, Zhaoxiang Zhang, Wanli Ouyang, Ke Xu, Wenhu Chen, Jie Fu, and Junran Peng. RoleLLM: Benchmarking, eliciting, and enhancing role-playing abilities of large language models.

[9] Nuo Chen, Yan Wang, Haiyun Jiang, Deng Cai, Yuhan Li, Ziyang Chen, Longyue Wang, and Jia Li. Large language models meet harry potter: A bilingual dataset for aligning dialogue agents with characters.

[10] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. Llama: Open and efficient foundation language models, 2023.

[11] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks, 2021.

[12] Jerry Liu. LlamaIndex, 11 2022.