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

# Draft: Literary Conversational Agents

Žiga Trček, Matej Urbas, and Jan Vasiljević

**Abstract**

The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here. The abstract goes here.

**Keywords**
Keyword1, Keyword2, Keyword3 ...

*Advisors: Slavko Žitnik*

## Introduction

Literacy among young people is declining, as highlighted in [1]. Many young people have a disinterest in reading and seldom read for enjoyment. A potential strategy to encourage reading is to involve them in conversational interactions with digital pedagogical agents that imitate well-known literary figures. Although numerous studies, such as those [2, 3], discuss the advantages of pedagogical agents, detailed technical implementation aspects are often overlooked. Our work aims to explore various methods for creating pedagogical agents and to provide a comprehensive technical implementation for the method we choose.

### Related work

In [4] they built a social bot, called Alana, which is able to engage in an open-domain conversation with their users over various popular topics. The key requirements for such a bot are to maintain the context, provide coherent responses, and be engaging and knowledgeable. The final bot is an ensemble of many different bots, each of which has a different purpose. A ranker is used to determine the best bot response for the given user input. For context maintaining, a state object with information from previous conversations is stored and accessible to every bot in the ensemble.

Retrieval augmented generation (RAG) is often used to correct factually inaccurate, outdated or halucinated Large Language Model (LLM) outputs. A survey on different RAG methods is conducted in [5]. RAGs can be seprated into three categories: pre-training, fine-tunning and inference. Now-days inference RAGs are mostly used. [6] proposed FLARE, Forward-Looking Active REtrieval Augmented Generation, which re-prompts the language model with extra retrieved data about the subject when some tokens in LLM's output have a low probability.

A straightforward way to create agents is by training or fine-tuning LLM. In [7] the authors developed conversational agents that resemble historical figures like Beethoven, Cleopatra, and Caesar, with personalized profiles, experiences, and emotional states. They introduced three new methods for training specialized agents. Experience Reconstruction extracts scenes in the style of memory flashes, such as profiles, scenes, or interactions. Protective Experience aims at teaching the model to forget or ignore information not relevant to the character to prevent knowledge hallucinations. Experience Upload uses the previous two techniques to fine-tune an existing LLM. They fine-tuned a LLaMa 7B model [8] on a dataset of 750 000 words per character using eight A100 80GB GPUs for 1 hour per character. To assess the models, they used an interview process.

To avoid training a LLM, [9] suggests using prompt engineering, specifically Chain-of-Thought (COT), on an existing model to incorporate more contextual information. They recommend employing Information-Rich Prompts (IRP) that include the emotional state, the character's relationship with the interlocutor, and the character's memories. Memories are categorized into short-term, which are a limited number of the most recent conversations with the interlocutor, and long-term, which are recursively summarized memories of longer

conversations from the character's perspective. Although not explicitly stated, implementing the Big Five personality model [10] could further refine the character's responses. This model would detail the character's Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

Previous methods that do not involve fine-tuning could be enhanced by using the OpenICL framework [11]. In-Context Learning (ICL) is an approach used with LLMs where the model learns a specific task without the need to update its weights. Instead, the model is shown examples of how the task should be performed. OpenICL offers the tools needed to construct ICL tasks, including key components like retrieval strategies and inference methods. For retrieval, it incorporates heuristic-based methods (such as BM25 and Top-K), random sampling, and model-based retrieval (using embeddings, RAG, Minimum Description Length (MDL), and entropy-based selection). For inference, OpenICL facilitates the integration of COT and other methods along with a prompt template.

> **Method proposals**
> We have a few ideas on how to tackle the problem of creating an Agent. These include:
>
> 1. Using ICL with OpenICL and RAG
> 2. Training our own LLM - can be quite expensive
> 3. Using another LLM to correctly format text - e.g., using Mistral [12] on HPC to give us a dataset.
> 4. Preprompting an LLM and giving it a context, such as character memories
>
> *This box here is only temporary and is meant only for the first phase. It will be removed in the next revision.*

## Methods

After the 1st phase

Use the Methods section to describe what you did an how you did it – in what way did you prepare the data, what algorithms did you use, how did you test various solutions ... Provide all the required details for a reproduction of your work.

Below are LaTeX examples of some common elements that you will probably need when writing your report (e.g. figures, equations, lists, code examples ...).

### Equations

You can write equations inline, e.g. $\cos \pi = -1$, $E = m \cdot c^2$ and $\alpha$, or you can include them as separate objects. The Bayes's rule is stated mathematically as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{1}$$

where $A$ and $B$ are some events. You can also reference it – the equation 1 describes the Bayes's rule.

### Lists
We can insert numbered and bullet lists:

1. First item in the list.
2. Second item in the list.
3. Third item in the list.

- First item in the list.
- Second item in the list.
- Third item in the list.

We can use the description environment to define or describe key terms and phrases.

**Word** What is a word?.

**Concept** What is a concept?
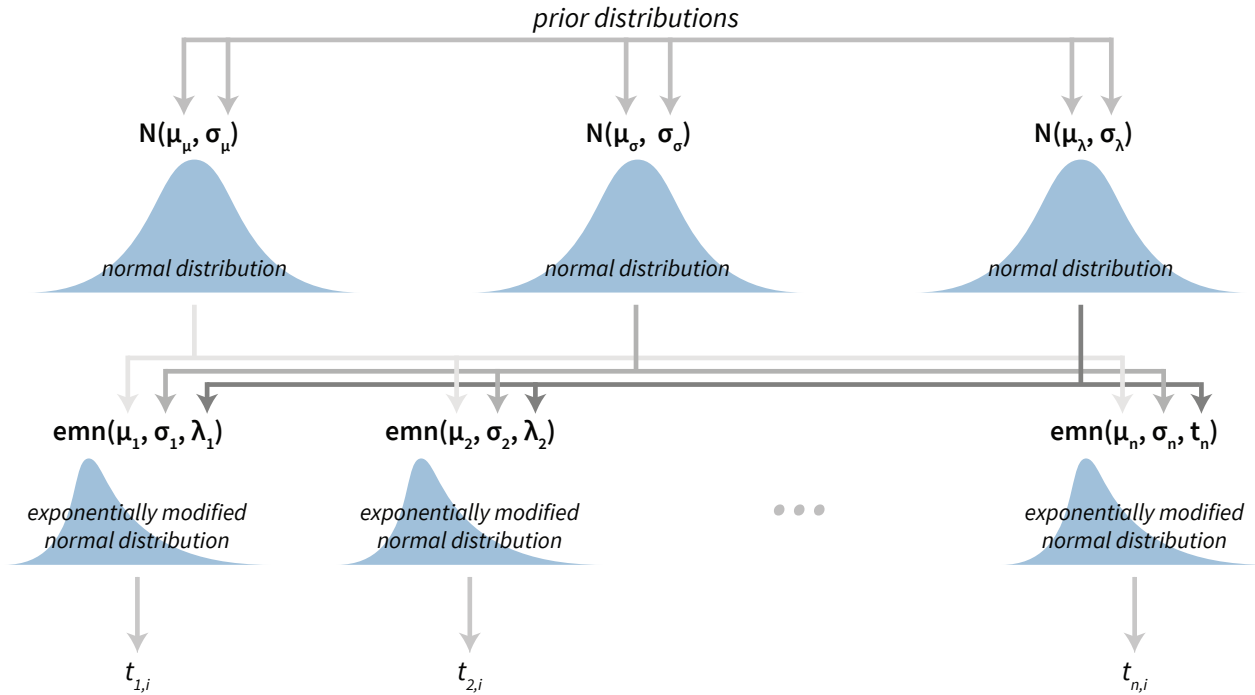
**Idea** What is an idea?

### Random text
This text is inserted only to make this template look more like a proper report. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam blandit dictum facilisis. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Interdum et malesuada fames ac ante ipsum primis in faucibus. Etiam convallis tellus velit, quis ornare ipsum aliquam id. Maecenas tempus mauris sit amet libero elementum eleifend. Nulla nunc orci, consectetur non consequat ac, consequat non nisl. Aenean vitae dui nec ex fringilla malesuada. Proin elit libero, faucibus eget neque quis, condimentum laoreet urna. Etiam at nunc quis felis pulvinar dignissim. Phasellus turpis turpis, vestibulum eget imperdiet in, molestie eget neque. Curabitur quis ante sed nunc varius dictum non quis nisl. Donec nec lobortis velit. Ut cursus, libero efficitur dictum imperdiet, odio mi fermentum dui, id vulputate metus velit sit amet risus. Nulla vel volutpat elit. Mauris ex erat, pulvinar ac accumsan sit amet, ultrices sit amet turpis.

Phasellus in ligula nunc. Vivamus sem lorem, malesuada sed pretium quis, varius convallis lectus. Quisque in risus nec lectus lobortis gravida non a sem. Quisque et vestibulum sem, vel mollis dolor. Nullam ante ex, scelerisque ac efficitur vel, rhoncus quis lectus. Pellentesque scelerisque efficitur purus in faucibus. Maecenas vestibulum vulputate nisl sed vestibulum. Nullam varius turpis in hendrerit posuere.

### Figures
You can insert figures that span over the whole page, or over just a single column. The first one, Figure 1, is an example of a figure that spans only across one of the two columns in the report.

On the other hand, Figure 2 is an example of a figure that spans across the whole page (across both columns) of the report.
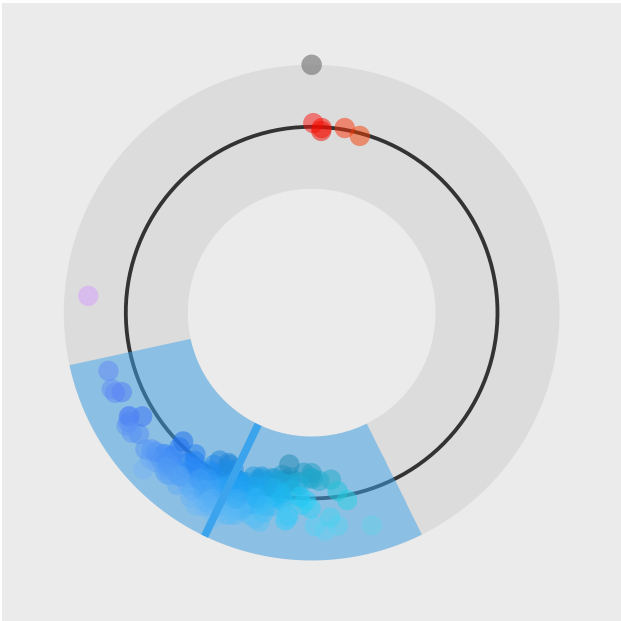
**Figure 1. A random visualization.** This is an example of a figure that spans only across one of the two columns.

### Tables

Use the table environment to insert tables.

**Table 1.** Table of grades.

| Name | | |
|---|---|---|
| First name | Last Name | Grade |
| John | Doe | 7.5 |
| Jane | Doe | 10 |
| Mike | Smith | 8 |

### Code examples

You can also insert short code examples. You can specify them manually, or insert a whole file with code. Please avoid inserting long code snippets, advisors will have access to your repositories and can take a look at your code there. If necessary, you can use this technique to insert code (or pseudo code) of short algorithms that are crucial for the understanding of the manuscript.

**Listing 1.** Insert code directly from a file.

```python
import os
import time
import random

fruits = ["apple", "banana", "cherry"]
for x in fruits:
  print(x)
```

**Listing 2.** Write the code you want to insert.

```r
import(dplyr)
```

```r
import(ggplot)

ggplot(diamonds,
        aes(x=carat, y=price, color=cut)) +
  geom_point() +
  geom_smooth()
```

## Results

> After the 1st phase

Use the results section to present the final results of your work. Present the results in a objective and scientific fashion. Use visualisations to convey your results in a clear and efficient manner. When comparing results between various techniques use appropriate statistical methodology.

**More random text**

This text is inserted only to make this template look more like a proper report. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam blandit dictum facilisis. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Interdum et malesuada fames ac ante ipsum primis in faucibus. Etiam convallis tellus velit, quis ornare ipsum aliquam id. Maecenas tempus mauris sit amet libero elementum eleifend. Nulla nunc orci, consectetur non consequat ac, consequat non nisl. Aenean vitae dui nec ex fringilla malesuada. Proin elit libero, faucibus eget neque quis, condimentum laoreet urna. Etiam at nunc quis felis pulvinar dignissim. Phasellus turpis turpis, vestibulum eget imperdiet in, molestie eget neque. Curabitur quis ante sed nunc varius dictum non quis nisl. Donec nec lobortis velit. Ut cursus, libero efficitur dictum imperdiet, odio mi fermentum dui, id vulputate metus velit sit amet risus. Nulla vel volutpat elit. Mauris ex erat, pulvinar ac accumsan sit amet, ultrices sit amet turpis.

Phasellus in ligula nunc. Vivamus sem lorem, malesuada sed pretium quis, varius convallis lectus. Quisque in risus nec lectus lobortis gravida non a sem. Quisque et vestibulum sem, vel mollis dolor. Nullam ante ex, scelerisque ac efficitur vel, rhoncus quis lectus. Pellentesque scelerisque efficitur purus in faucibus. Maecenas vestibulum vulputate nisl sed vestibulum. Nullam varius turpis in hendrerit posuere.

Nulla rhoncus tortor eget ipsum commodo lacinia sit amet eu urna. Cras maximus leo mauris, ac congue eros sollicitudin ac. Integer vel erat varius, scelerisque orci eu, tristique purus. Proin id leo quis ante pharetra suscipit et non magna. Morbi in volutpat erat. Vivamus sit amet libero eu lacus pulvinar pharetra sed at felis. Vivamus non nibh a orci viverra rhoncus sit amet ullamcorper sem. Ut nec tempor dui. Aliquam convallis vitae nisi ac volutpat. Nam accumsan, erat eget faucibus commodo, ligula dui cursus nisi, at laoreet odio augue id eros. Curabitur quis tellus eget nunc ornare auctor.

## Discussion

> After the 1st phase

*prior distributions*

$N(\mu_\mu, \sigma_\mu)$  $N(\mu_\sigma, \sigma_\sigma)$  $N(\mu_\lambda, \sigma_\lambda)$

*normal distribution*  *normal distribution*  *normal distribution*

$emn(\mu_1, \sigma_1, \lambda_1)$  $emn(\mu_2, \sigma_2, \lambda_2)$  $emn(\mu_n, \sigma_n, t_n)$

*exponentially modified normal distribution*  *exponentially modified normal distribution*  ...  *exponentially modified normal distribution*

$t_{1,i}$  $t_{2,i}$  $t_{n,i}$

**Figure 2. Visualization of a Bayesian hierarchical model.** This is an example of a figure that spans the whole width of the report.

Use the Discussion section to objectively evaluate your work, do not just put praise on everything you did, be critical and exposes flaws and weaknesses of your solution. You can also explain what you would do differently if you would be able to start again and what upgrades could be done on the project in the future.

## Acknowledgments

## References

[1] Jane Murray. Literacy is inadequate: young children need literacies, 2021.

[2] 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.

[3] 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.

[4] Ioannis Papaioannou et al. *Designing coherent and engaging open-domain conversational AI systems*. PhD thesis, Heriot-Watt University, 2022.

[5] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2023.

[6] 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.

[7] Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. Character-llm: A trainable agent for role-playing, 2023.

[8] 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.

[9] Seokhoon Jeong and Assentay Makhmud. Chatbot is not all you need: Information-rich prompting for more realistic responses, 2023.

[10] L R Goldberg. An alternative "description of personality": the big-five factor structure. *J Pers Soc Psychol*, 59(6):1216–1229, December 1990.

[11] Zhenyu Wu, YaoXiang Wang, Jiacheng Ye, Jiangtao Feng, Jingjing Xu, Yu Qiao, and Zhiyong Wu. Openicl: An open-source framework for in-context learning, 2023.

[12] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. Mistral 7b, 2023.