The Thesis Committee for Yian Wong
certifies that this is the approved version of the following thesis:

Exploring Multiple Perspectives to Mitigate Cognitive Biases through an Integrated Interface to Language Models

SUPERVISING COMMITTEE:

Matt Lease, Supervisor

Jessy Li, Second Reader

Exploring Multiple Perspectives to Mitigate Cognitive Biases through an Integrated Interface to Language Models

by
Yian Wong

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Computer Science

The University of Texas at Austin
May 2024

Abstract

Exploring Multiple Perspectives to Mitigate Cognitive Biases through an Integrated Interface to Language Models

Yian Wong, MSCompSci
The University of Texas at Austin, 2024

SUPERVISOR: Matt Lease

In recent years, large language models (LLMs) have demonstrated remarkable abilities in generating human-like text and supporting decision-making processes. However, their use is often limited by inherent biases and a lack of diversity in presented perspectives. This work introduces a novel system designed to mitigate these issues by leveraging the capabilities of LLMs to simulate a multi-perspective debate format, aimed at providing a balanced view on controversial topics. The proposed system employs a unique integrated interface that facilitates dynamic interactions between multiple AI-generated personas, each representing distinct viewpoints. These personas engage in structured debates, allowing for a comprehensive exploration of a topic that counteracts the cognitive biases typically associated with single-perspective information retrieval systems.

The system incorporates advanced prompt engineering techniques and retrieval-augmented generation to ensure the accuracy and relevance of the information presented. Additionally, the interface is designed with user engagement in mind, featuring interactive elements that allow users to manipulate the debate dynamics and contribute to the discussion. This thesis evaluates the system's effectiveness in enhancing users' understanding of complex issues and its potential in reducing bias in

decision support systems. By simulating diverse viewpoints, the system potentially fosters more critical and informed engagement with topics, thus supporting better decision-making.

Table of Contents

Chapter 1: Introduction

Large Language Models (LLMs), have recently demonstrated remarkable capabilities in automated reasoning and decision support systems. These LLMs have demonstrated state-of-the-art performance in generating human-like text, solving complex problems, and engaging in meaningful dialogue with its users. Despite this prowess, LLMs do have several shortcomings and limitations, which lead to unexpected behavior from gaps in knowledge or difficult tasks. Prompt engineering is the process where inputs (prompts) are carefully crafted to guide the behavior of LLMs. Research in prompt engineering is especially focused on improving performance in automated reasoning tasks, such as mathematical tasks, common-sense reasoning, factuality question answering, or creative writing. For further discussion on related prompt engineering work, see Chapter 2.1. In this work, we develop a prompt engineering method that diverges from the prevelant work on improving automated reasoning performance, and leverage LLMs to develop a decision support system that aims to inform users about controversial topics. To this end, we create a novel debate interface with the following key components:

1. **Multi-agent debate**: Introduced in Du et al. (2024), we create multiple instances of the LLM to debate with each other, by maintaining a chat history for each of the 'agents'. We allow each agent to interact with each other, to contribute and debate within a discussion.

2. **Persona-based contribution**: In the context of multi-agent debate or multi-agent collaboration, rather than each agent statically adopting the same role, we prompt each agent to roleplay as a specific 'persona', conditioning their responses on their background and effectively offering unique perspectives to the debate. (For the purposes of this paper, we will call each instance of the LLM in the debate 'personas'.) Wang et al. (2024) uses persona-based collaboration

in order to effectively solve Trivia Creative Writing tasks, where personas adopt the role of an expert for different aspects of the task.

The potential applications of this system are vast and varied, from navigating public controversies and personal decision-making dilemmas to wrestling with ethical quandaries. This system offers a sophisticated tool that can better simulate multiple perspectives, mirroring a more human-like deliberation process and avoiding confirmation bias. Similar to the approach taken by the Delphi experiment, which gathers and synthesizes expert opinions to make predictions in uncertain scenarios Jiang et al. (2021), our system leverages Large Language Models to present a diverse array of viewpoints. This not only facilitates a deeper understanding of complex issues but also highlights the system's ability to act as a dynamic decision support tool, reflecting the multifaceted nature of human thought and debate.

The thesis is structured as follows. In Chapter 2, we introduce the current state-of-the-art in LLMs, discussing their capabilities, limitations, and the enhancements prompt engineering grants them. In Chapter 3.1, we outline the design goals of our system, and highlight notable challenges when implementing and our solutions. Next, Chapter 4 outlines our system architecture, including the various libraries and technologies used to make this implementation possible. Chapter 5 will outline the system's integration with the user-facing interface, as well as modifications needed to the system to enhance the user experience. Finally, we discuss and reflect on our system, and its implications for future work in the space of prompt engineering and user decision support in Chapter 6. Finally, we conclude in Chapter 6.2.

Chapter 2: Large Language Models: Capabilities and Limitations

LLMs are transformer-based AI models trained on large amounts of text data, and tuned to engage in helpful conversation with users. Their pre-training enables them to grasp a wide range of topics, from mathematical theories to historical literature. The ability to generate coherent and contextually relevant responses has found applications in a vast range of fields including education, content writing, or customer service.

Despite these advanced capabilities, LLMs have inherent limitations. One significant issue is their propensity to inadvertently perpetuate biases that are present in the vast amounts of training data they are fed. This is because these models learn patterns and associations found in the data, including societal stereotypes and prejudices, which can then be reflected in their outputs. Another challenge lies in their struggle with understanding and processing longer context range prompts. LLMs can find it difficult to maintain coherence over extended texts or to accurately remember and apply information from earlier parts of a conversation or document. This limitation often leads to a degradation in the quality and relevance of their responses as the context lengthens. Furthermore, LLMs are known for sometimes substituting gaps in their knowledge with completely nonsensical or false responses, a phenomenon known as 'hallucinations'. This occurs because, despite their sophisticated algorithms, these models do not possess true understanding or reasoning capabilities. Instead, they generate responses based on statistical likelihoods, which can lead to the production of information that appears plausible at first glance, but is ultimately factually incorrect or entirely fabricated. This aspect poses significant challenges for users who rely on LLMs for accurate information, as it can be difficult to distinguish between valid responses and these 'hallucinated' inaccuracies without additional verification.

2.1 Prompt Engineering

Prompt engineering is an important technique and essential framework for interaction with LLMs that enables users to fully or partially circumvent these limitations. Prompt engineering methods identify creative ways for LLMs to elaborate or revisit their thought process, typically achieving improved performance as measured by benchmarks such as mathematical reasoning, creative writing, or factual question-answering. For example, Yao et al. (2024) introduces Tree-of-Thoughts, which prompts the LLM to generate 'thoughts' for tackling the given task, maintaining a tree structure where each node represents a thought, each serving as intermediate steps toward solving a problem. This approach enables the language model to self-evaluate its progress through these thoughts and uses a tree search algorithm to systematically explore different thoughts. Generated Knowledge Prompting (Liu et al. (2021)) prompts the LLM to support its answers in factual question answering tasks by generating relevant facts related to the question before answering. This forces the LLM to leverage its training data to answer its question that is conditioned on its recited fact, rather than just generating an answer that is syntactically and coherently plausible.

Many prompt engineering techniques also involve maintaining multiple instances of the same LLM to interact with each other. By simulating a discussion between multiple instances of the LLM, we achieve a similar effect by emphasizing a greater focus on thought process and elaboration from different perspectives. Multi-agent debate (Du et al. (2024)) uses multiple uniform LLM instances to engage in many rounds of debate to answer math and commonsense reasoning tasks. This encourages LLMs to scrutinize and correct other agent's responses, leading to better performance. Wang et al. (2024) uses an LLM to generate 'personas', or domain experts, who could all collaborate in a task. Rather than having uniform agents, each LLM instead adopts the role of a given persona. This allows each LLM to contribute to the task based on their unique background, leading to a better result in creative writing tasks. Drawing from this, while our goal is not to enable LLMs to achieve better performance in a set of downstream tasks such as commonsense reasoning or creative writing, we use LLMs and their ability to roleplay as a key part of our system. By combining multi-agent debate with persona-based collaboration, we arrive at a system that presents unique viewpoints for a given topic, and how they might interact with each other.

Chapter 3: Design Overview

3.1 Design Objectives

The primary design goal of our system is to create a sophisticated decision support tool that facilitates a nuanced understanding of controversial topics by presenting multiple perspectives in a structured debate format. This approach aims to mitigate the risk of cognitive biases inherent in traditional decision support systems, such as search engines, which often rank search results based on algorithms that may not fully account for the diversity of valid viewpoints on a subject. By simulating a debate between personas, each advocating for a distinct perspective, users are exposed to a rich array of arguments, counterarguments, and the complex interplay of ideas that characterize real-world discussions. This interactive experience is designed to equip users with a comprehensive overview of the main points of each perspective, fostering a deeper, more critical engagement with the topic at hand. To this end, this system's design objectives are as follows:

1. **Accurate Representation**: Ensure that each major viewpoint on a controversial topic is represented by a dedicated persona. This involves a careful and balanced generation of perspectives to avoid unintentional bias in the representation of opinions and facts.

2. **Dynamic Interaction and Real-Time Debate**: Facilitate dynamic interaction among personas, allowing them to engage in real-time debate. This includes the ability for personas to respond to each other's arguments, pose questions, and provide rebuttals, closely mirroring the fluid nature of human dialogue.

3. **Factual Accuracy and Evidence-Based Arguments**: Using integrated technology like Retrieval-Augmented Generation (RAG) ensures that personas draw upon up-to-date and relevant information from credible sources, thereby grounding the debate in verifiable facts.

4. **User Engagement and Interactivity**: Design the system to be highly interactive, enabling users to save notable responses, modify the participants of the debate, and request an additional round of debate. This level of interactivity is key to engaging users actively in the deliberation process, rather than as passive recipients of information.

3.2 Challenges

3.2.1 Context Windows

These design goals present a key technical challenge, specifically concerned with the architecture of LLMs. LLMs are limited by a certain text length before it begins to misremember earlier parts of the input text. This text length is known as an LLMs 'context window'. The context window of an LLM may pose a key problem in our system, especially given that personas need to view the history of the debate in order to make effective responses. To circumvent this, if the history of the debate goes over the LLM's context window, we use an extractive summarizer to compress the debate transcript until it is an appropriate length.

3.2.2 Prompting Specifics

While it may be clear that LLMs are certainly capable of generating personas and acting as each one, it may not be clear how to prompt them to do so. For instance, if we ask an LLM to generate personas for a given topic, it may sometimes produce personas with actual names (such as John Smith), and other times with professional titles (such as Doctor). In another case, when generating a description, the LLM may generate a comprehensive background about the persona, while in other cases it may generate a detailed description about their stance only. In order to enforce consistency between runs, we must add many specifications to the prompt in a suitable manner to get this desired behavior. Luckily, in the two cases mentioned above, it suffices to simply add "generate a title for the persona, not a name", or "only describe their background, rather than their stance" respectively.

Other cases of inconsistency arise in formatting. In general, we want our LLM to generate multiple personas in a manner that is parse-friendly for a program. In our system, we prompt the LLM to output the generated personas as a large JSON text with a specific format. When responding, LLMs can tend to output very long and drawn-out answers, which can sometimes reduce user engagement with the overall topic. To circumvent this, we can also simply add something to the effect of "Generate at most 200 words", as well as to "generate using bullet points (in markdown format)" to make the responses both significantly more readable for the user and parse-friendly for our interface.

Chapter 4: System Architecture

In this chapter, we discuss the architecture of our underlying system.

4.1 System Overview

The process flow of our system is outlined in Figure 4.1.

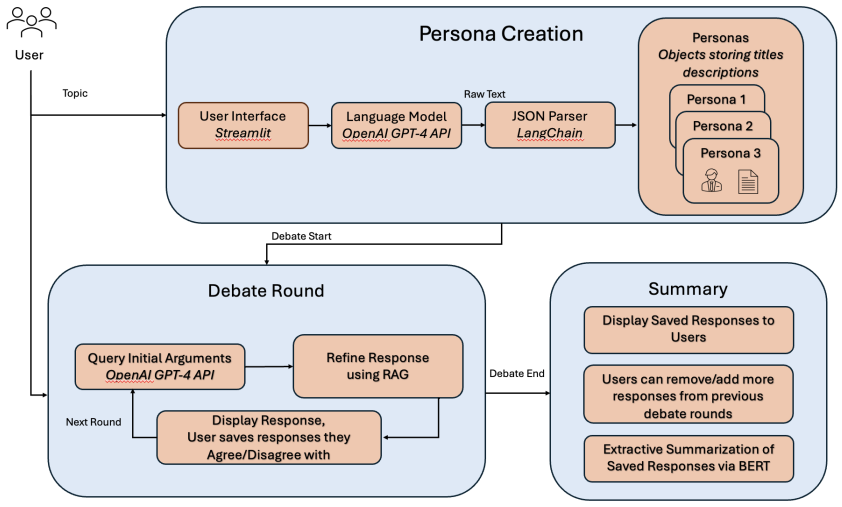

User
Topic
Persona Creation
User Interface
Streamlit
Language Model
OpenAI GPT-4 API
Raw Text
JSON Parser
LangChain
Personas
Objects storing titles descriptions
Persona 1
Persona 2
Persona 3
Debate Start
Debate Round
Query Initial Arguments
OpenAI GPT-4 API
Refine Response using RAG
Next Round
Display Response, User saves responses they Agree/Disagree with
Debate End
Summary
Display Saved Responses to Users
Users can remove/add more responses from previous debate rounds
Extractive Summarization of Saved Responses via BERT

Figure 4.1: Overview of Multi-Persona Debate System

4.1.1 Large Language Model Choice

The key driver behind our architecture is which LLM we use. As of 2024, (Chat)GPT-4 has proven to be one of the most capable LLMs available (OpenAI et al.). Furthermore, GPT-4 now has a 128K token context window, ensuring high performance in longer range tasks such as multi-agent debate. To this end, for our final implementation of this interface, we use GPT-4 as it is a clear choice for the highest performing LLM, not only in downstream tasks, but also coherency. However, we design our system architecture with modularity of the LLM in mind. This means that, as long as an LLM supports a conversational interface between the user and the LLM, our system can easily be modified to support that LLM instead. For instance, during development we also test using Google Gemini Pro and GPT-3.5.

4.1.2 Supporting LLM Outputs and Inputs

As mentioned in Chapter 3.2, our specific inputs to the LLM can greatly influence LLM outputs. In order to be able to parse the LLM persona generation, we use LangChain's JSON parser Langchain (2023) and prompt LLMs to output a JSON using the prompt as follows:

```
Given the topic [TOPIC], create a roundtable debate of different
personas to show key perspectives on the issue.  Output the personas
as a list of JSON objects.  Each JSON object should have the following
structure:
{ "title":  <title of the Persona>,
"description":  <Brief Description of the Persona, only describing
their background (rather than their stance)>,
"emoji":  <Single emoji representation of the perspective the
persona represents>}.
Ensure that the output is formatted as a valid JSON.
Please generate exactly [NUM_PERSONAS] personas.
```

Here, we provide an example JSON structure for the LLM to output the JSON in, the use LangChain's JSON parser to collect the generated content and create an object-oriented strcuture from it.

In order to prompt a persona LLM to debate, we feed it the existing debate history as follows:

```
You are in a continuing roundtable debate on the topic [TOPIC].
You are [NAME], who is [DESC].
Here is the transcript of the debate so far [HISTORY]
Please continue to debate the others, concisely supporting your
```

```
stance on
the topic.  Please limit your response to 200 words, you may use
bullet points and markdown format in your response.
```

As mentioned previously in Chapter 3.2, if the history of the debate exceeds the LLM context window, we use an extractive summarizer to condense the transcript until it fits into our LLM again. In our system, we use BERTExtractiveSummarizer, which provides a good balance between summarization performance and latency Miller (2019). It is also evident from this quote how exactly we prompt the LLM to limit and format its response, which results in consistent and appropriate formatting.

4.1.3 Retrieval Augmented Generation

In the scope of enhancing the factual grounding of persona responses in future iterations of this project, the integration of RAG presents a critical and promising avenue. This approach would involve combining a semantic vectorizer with a search engine like Google Search to dynamically augment response generation with real-time data retrieval. While this functionality was not implemented in the current version of the thesis project and is being developed in a separate work, the impact of RAG integration is clear. RAG can serve as a potential upgrade to enrich the factuality of the system's responses. Such a feature would allow the system to verify and enhance its outputs by referencing up-to-date, relevant information, enhancing both the reliability and the depth of interactions.

Chapter 5: System Interface

We developed the interface of the system using Streamlit Treuille et al. (2019), a popular data science and machine learning web development framework. This framework allows development with an emphasis on interactivity and ease of use. Furthermore, it promotes rapid prototyping, which made it feasible to iterate on the interface design based on user feedback or emerging requirements.

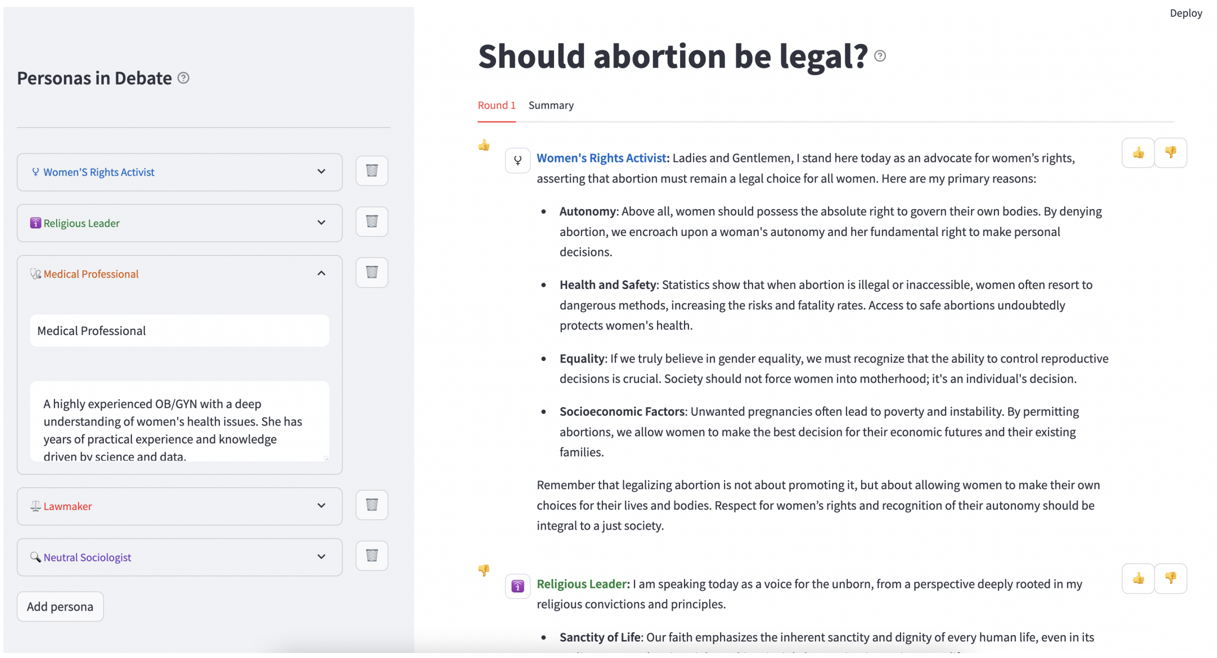
Streamlit provides clear widgets that have natural uses for our system. For example, when inputting a topic for our users, we can use Streamlit's text input widget to pass it to our system. Similarly, when a persona responds in a debate, we can use Streamlit's writing method, which support markdown. Here, we highlight additional features we added in our interface to enhance the system functionality and user experience.

5.0.1 Bookmarks

Our interface is content-rich, and can be overwhelming to users at times. To help users collect their thoughts, we introduce bookmarking in our interface. Users can click whether they agree or disagree with certain responses, and it will be saved in a separate 'Summary' section of the webpage. An example of this is shown in Figures 5.1 and 5.2.

5.0.2 Persona Identification

In addition to colour coding the names of personas to help the user keep track of different personas, we also modify our persona generation prompt to generate and assign an emoji to each persona. This way, it appears next to the name of each persona, ideally engaging the user more and helping them keep better track of the many perspectives in a debate.



Medical Professional

A highly experienced OB/GYN with a deep understanding of women's health issues. She has years of practical experience and knowledge driven by science and data.

Should abortion be legal?

Women's Rights Activist: Ladies and Gentlemen, I stand here today as an advocate for women's rights, asserting that abortion must remain a legal choice for all women. Here are my primary reasons:

- **Autonomy**: Above all, women should possess the absolute right to govern their own bodies. By denying abortion, we encroach upon a woman's autonomy and her fundamental right to make personal decisions.
- **Health and Safety**: Statistics show that when abortion is illegal or inaccessible, women often resort to dangerous methods, increasing the risks and fatality rates. Access to safe abortions undoubtedly protects women's health.
- **Equality**: If we truly believe in gender equality, we must recognize that the ability to control reproductive decisions is crucial. Society should not force women into motherhood; it's an individual's decision.
- **Socioeconomic Factors**: Unwanted pregnancies often lead to poverty and instability. By permitting abortions, we allow women to make the best decision for their economic futures and their existing families.

Remember that legalizing abortion is not about promoting it, but about allowing women to make their own choices for their lives and bodies. Respect for women's rights and recognition of their autonomy should be integral to a just society.

Religious Leader: I am speaking today as a voice for the unborn, from a perspective deeply rooted in my religious convictions and principles.

- **Sanctity of Life**: Our faith emphasizes the inherent sanctity and dignity of every human life, even in its

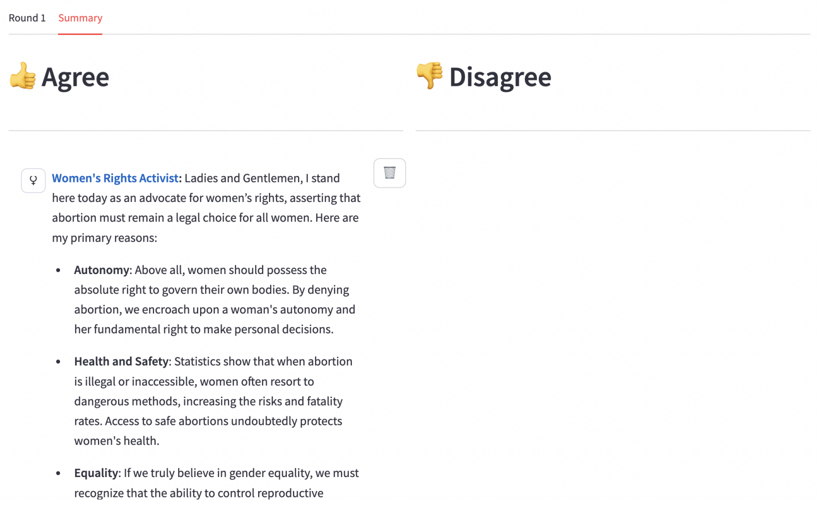
Figure 5.1: Interface for an example topic, "Should abortion be banned?". Personas are generated by our LLM on the left, where users can modify, delete, or add personas. When the debate starts, personas engage in debate, and the users can mark each response with agree or disagree using the thumbs up or thumbs down buttons, which shows up in the Summary tab in Figure 5.2

5.0.3 Persona Generation

The interface is designed to generate personas as soon as a topic is chosen. We set a predefined number of personas at the beginning (5), but users are free to generate as many or as little personas as they want in a debate. Because of this, we create a prompt for persona addition, which gives the LLM existing personas and asks it to output an additional persona in JSON format as before.

5.0.4 Debate Flow and Performance Optimizations

During the development of this interface, we encountered a unique design decision which controls how much each persona interacts with each other within a round. In a natural human debate, responses can be tailored to as much as the most



Women's Rights Activist: Ladies and Gentlemen, I stand here today as an advocate for women's rights, asserting that abortion must remain a legal choice for all women. Here are my primary reasons:

- **Autonomy**: Above all, women should possess the absolute right to govern their own bodies. By denying abortion, we encroach upon a woman's autonomy and her fundamental right to make personal decisions.
- **Health and Safety**: Statistics show that when abortion is illegal or inaccessible, women often resort to dangerous methods, increasing the risks and fatality rates. Access to safe abortions undoubtedly protects women's health.
- **Equality**: If we truly believe in gender equality, we must recognize that the ability to control reproductive

Figure 5.2: Example summary tab with one response saved (agree). Users can save responses for later and view them later on to help support their decision process and synthesize their key takeaways.

recent response. While this is possible in our multi-persona debate, we also recognized that personas could just respond to the responses up until the previous round. The main advantage of this is for latency and performance. From our experiments, when using RAG, responses from a single LLM can take as much as 5 minutes. If each persona responds to the most recent response, then each LLM call must happen in sequence. However, if each persona responds to the same transcript, up until the last round, we have the potential to further optimize our pipeline by concurrently calling the LLM API as well as the RAG pipeline. While not implemented, this should speed-up our interface by many factors.

Chapter 6: Discussion and Conclusion

6.1 Discussion

In this project, we have developed a sophisticated decision support tool that harnesses the capabilities of LLMs to provide multiple perspectives on controversial topics in a structured debate format. Our approach sought to mitigate the risk of cognitive biases that are often inherent in conventional decision support systems by presenting diverse viewpoints through the simulation of a real-time debate among personas.

Our design objectives focused on accurate representation, dynamic interaction, factual accuracy, and user engagement. We believe our system achieves these goals by utilizing the most proficient of LLMs, GPT-4, to ensure a high level of performance and coherence in responses and integrating it with advanced prompt engineering techniques.

One of the key innovations in our system is the use of prompt engineering to guide the LLMs in generating and interacting as distinct personas with balanced and evidence-based arguments. This technique, coupled with the dynamic retrieval of external information, ensures that the debates are not only informative but grounded in factual accuracy.

The interface, developed with Streamlit, enhances user experience through interactivity and ease of navigation, allowing users to engage with the debate actively. Features such as bookmarks and persona identification further aid in making the debates more accessible and understandable to the users, enabling them to synthesize information effectively and form well-rounded opinions.

6.1.1 Limitations and Future Work

- **Context Window Limitations**: As LLMs have a fixed context window, they struggle with long debates or discussions, often losing track of earlier exchanges as the conversation progresses. For transformer-based methods, this can be mitigated with summarization techniques, but these may oversimplify or omit crucial details. However, Munkhdalai et al. (2024) presents novel infinite-context models, effectively circumventing this method entirely while maintaining comparable language modeling performance and capability. Further research and exploration on these models could allow for more meaningful responses in longer context range debates.

- **Unpredictable LLM Behavior**: There's an inherent challenge in prompting LLMs to generate consistent and relevant personas. In our tests, even with our refined prompts, LLMs have the potential to produce personas with real human names (eg, John Smith) compared to titles (eg, Medical Expert). Variations in how personas are prompted can lead to inconsistencies in their background or expertise, which might affect the quality and relevance of the debate outcomes, and potentially introduce unexpected biases in persona generation. Stricter and more refined prompt engineering could ensure more uniform and predictable persona behaviors.

- **Real-Time Latency**: During development of this work, we noticed that on average persona response times vary from 10 to 15 seconds. With 5 or more personas in a debate, this can lead to minutes just for the generation of a single debate round. Furthermore, integration with RAG can increase response times significantly, potentially impacting user engagement and the fluidity of the debate experience. Optimizing these aspects to reduce latency could enhance user experience and system responsiveness.

- **Debate Dynamics and Interaction Quality**: Testers of our interface have noted that the multi-round aspect of the system clearly has diminishing returns.

As the debate round progresses, personas tend to reiterate their points rather than addressing and responding to the points brought up by other personas. Enhancing the model's ability to dynamically adapt and evolve arguments based on the debate's progression could make interactions more realistic and engaging.

6.2 Conclusion

This research presents a novel application of LLMs to facilitate a nuanced understanding of controversial topics through a multi-perspective debate format. By leveraging the latest advancements in AI, we have created a decision support tool that not only addresses the inherent cognitive biases in traditional information retrieval systems but also promotes critical engagement and informed decision-making.

Our work underscores the potential of LLMs in transforming how we interact with and process complex information. The successful integration of technologies like RAG and the innovative use of prompt engineering in our system exemplify the dynamic capabilities of LLMs when appropriately guided.

Reflecting on the research process, it is clear that the development of effective LLM-based applications requires a careful balance between technical precision and user-centric design. Our experience highlights the importance of iterative development and the need for continuous optimization to address performance and user experience challenges.

Looking ahead, there are several exciting directions for future work. The further integration of our system into mainstream platforms like ChatGPT or New Bing could significantly broaden its impact, making sophisticated debate and information synthesis tools more widely accessible. Additionally, enhancing the realism of the debates, possibly through the use of avatars and more interactive formats, could further engage users and enrich the debate experience.

In conclusion, our project represents a meaningful contribution to the ongoing effort to harness the power of AI for social good. By providing a platform for balanced and informed discourse, we hope to foster a more critical and open-minded approach to understanding the complex issues that face our society today.

Appendix A: Source Code

A.1 Code Repository Link

The source code for the project is available on GitHub at the following link: `github.com/1yian/debate-prototype`

A.2 Demo

A preliminary demo of this project is available on: `debate-prototype.streamlit.app`

Appendix B: Prompts

B.1 Prompt Placeholders

Throughout our prompts, we use placeholders, which are automatically replaced by their corresponding value as defined in our initial configuration. For all our current version of the system these values are:

1. [**TOPIC**]: The user-input topic to focus persona generation and debate discussion around.
2. [**NUM_PERSONAS**]: Initial requested number of personas
3. [**CURRENT_PERSONAS**]: Current personas in JSON format
4. [**NAME**]: Name of the persona
5. [**DESC**]: Description of the persona
6. [**HISTORY**]: Transcript of the debate in JSON format
7. [**LIMITER**]: Sentence telling the LLM how long their response should be (ie. Limit your response to 5 sentences).

B.2 Persona Generation

B.2.1 Generating an initial set of personas

```
Given the topic, [TOPIC], create a roundtable debate of different
personas to show key perspectives of it.  Output the personas
as a list of JSON objects.  Each JSON object should have the following
structure:
{ "title":  <title of the Persona>,
"description":  <Brief Description of the Persona, only describing
their background (rather than their stance)>,
"emoji": <Single emoji representation of the perspective the
persona represents>}.
Ensure that the output is formatted as a valid JSON.
Please generate exactly [NUM_PERSONAS] personas.
```

B.2.2 Adding an Additional Persona

```
Given the topic, [TOPIC], we are creating a roundtable debate
of different personas to show key perspectives of it.
Currently, we have personas as follows: [CURRENT_PERSONAS] Please
output exactly one additional persona. Your output should have
the following JSON structure:
{"title": <title of the Persona>,
"description": <Brief Description of the Persona, only describing
their background (rather than their stance)>,
"emoji": <Single emoji representation of the perspective the
persona represents>}.
Ensure that the output is formatted as a valid JSON.
```

B.3 Persona Response

B.3.1 Starting a debate

```
You are in an roundtable debate on the topic [TOPIC].
You are [NAME], who is [DESC].
Please start the debate by concisely presenting your argument
for your stance on the topic. [LIMITER]
```

B.3.2 Continuing the debate

```
You are in a continuing roundtable debate on the topic [TOPIC].
You are [NAME], who is [DESC].
Here is the transcript of the debate so far [HISTORY]
Please continue to debate the others, concisely supporting your
stance on the topic. [LIMITER]
```

Works Cited

Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. Improving factuality and reasoning in language models through multiagent debate, 2024. URL `https://openreview.net/forum?id=QAwaaLJNCk`.

Liwei Jiang, Jena D Hwang, Chandra Bhagavatula, Ronan Le Bras, Jenny Liang, Jesse Dodge, Keisuke Sakaguchi, Maxwell Forbes, Jon Borchardt, Saadia Gabriel, et al. Can machines learn morality? the delphi experiment. *arXiv preprint arXiv:2110.07574*, 2021.

Langchain. Python langchain documentation: Modules - model io - output parsers - types - json, 2023. URL `https://python.langchain.com/docs/modules/model_io/output_parsers/types/json`. Accessed: 2024-04-02.

Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West, Ronan Le Bras, Yejin Choi, and Hannaneh Hajishirzi. Generated knowledge prompting for commonsense reasoning. *arXiv preprint arXiv:2110.08387*, 2021.

Derek Miller. Leveraging BERT for extractive text summarization on lectures. *CoRR*, abs/1906.04165, 2019. URL `http://arxiv.org/abs/1906.04165`.

Tsendsuren Munkhdalai, Manaal Faruqui, and Siddharth Gopal. Leave no context behind: Efficient infinite context transformers with infini-attention. *arXiv preprint arXiv:2404.07143*, 2024.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny

Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, and Ruby Chen. Gpt-4 technical report.

Adrien Treuille et al. Streamlit. `https://streamlit.io`, 2019. Software for machine learning and data science teams to create beautiful data apps in hours, not weeks.

Zhenhailong Wang, Shaoguang Mao, Wenshan Wu, Tao Ge, Furu Wei, and Heng Ji. Unleashing the emergent cognitive synergy in large language models: A task-solving agent through multi-persona self-collaboration, 2024.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, 36, 2024.