Multi-Agentic LLM Framework to Mitigate Cognitive Biases in Information-Seeking Contexts

Yian Wong[1], Utkarsh Mujumdar[2], Li Shi[2], Matt Lease[2] , Houjiang Liu[2]
[1]Department of Computer Science, [2]School of Information,
University of Texas at Austin

GOOD SYSTEMS
Ethical AI at UT Austin

The University of Texas at Austin
Research Development
Vice President for Research, Scholarship and Creative Endeavors

Introduction

Large Language Models (LLMs) like ChatGPT are revolutionizing the way we seek and receive information. These systems offer significant potential in assisting users with decision-making processes across various domains. However, they tend to amplify cognitive biases such as automation bias and confirmation bias. These biases can skew user perceptions and decisions, especially in complex or controversial topics where a balanced perspective is crucial. This boils down to a core issue of information-seeking behavior with only one perspective. Instead, if multiple perspectives could be presented for a topic or query, it would be more helpful for the users as they'll be able to discover information from a perspective that differs from their own, enabling them to counter their biases.

Research Goal

Our research aims to mitigate cognitive biases in LLMs by introducing an innovative solution that leverages LLMs to argue for multiple perspectives to a complex topic, rather than just presenting a single perspective.

System Design

We use existing prompt engineering methods to offer multiple perspectives on a given topic to the user, in order to present a nuanced aspect of information. Our system creates multiple "personas" for our LLM to adopt, and presents an argument in a debate-style setting for the effective user decision support. We integrate this with Retrieval Augmented Generation (RAG) to allow the LLM to ground their responses in factuality.

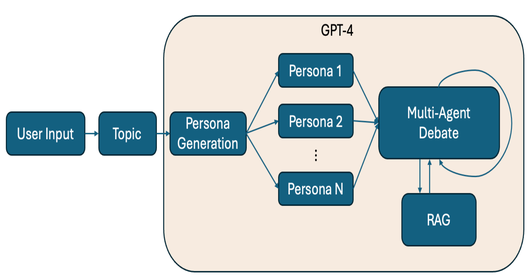
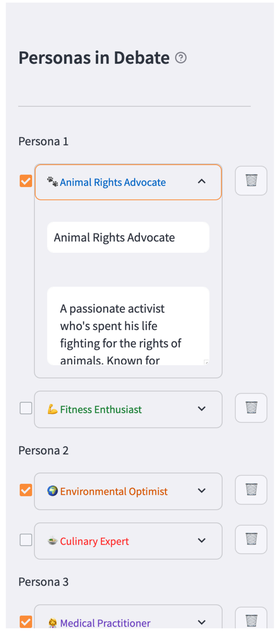


Fig 1: System Design Flow for the Web-App

Interface Design



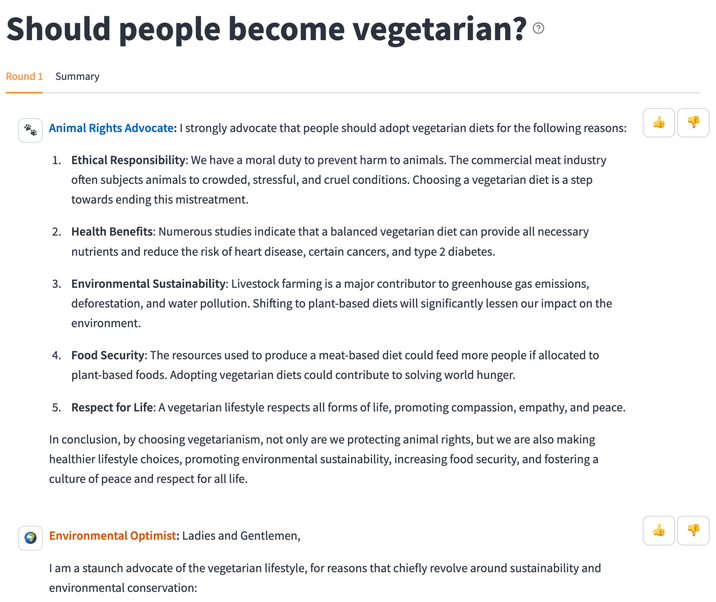


Fig 2: Main User Interface of the Prototype Web App

Our web-based application is engineered to challenge and mitigate cognitive biases in decision-making by engaging users in a simulated multi-perspective debate. The interface is intuitively designed to encourage users to critically evaluate arguments by saving them in 'agree' or 'disagree' categories, fostering a more balanced and informed decision-making process

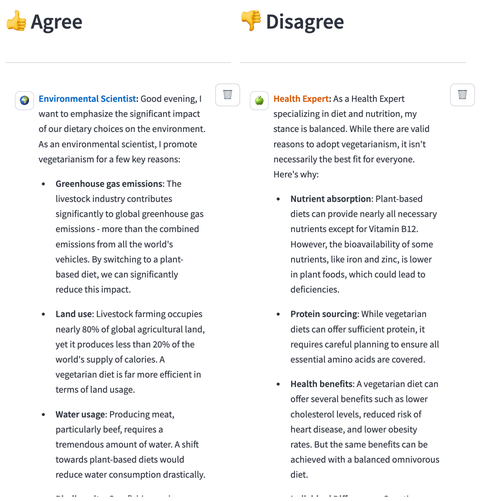


Fig 3: Summary feature showing the arguments the user agreed/disagreed with

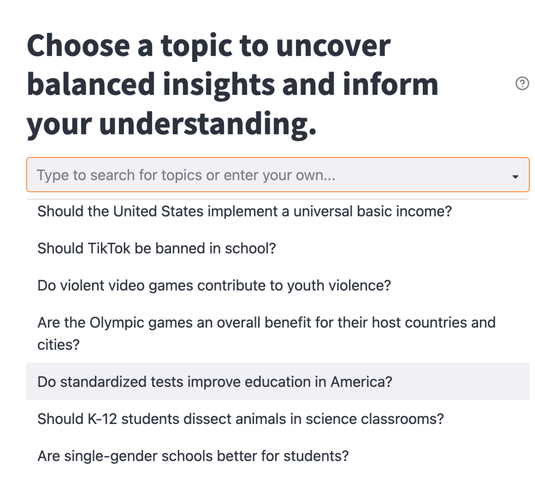


Fig 4: Topics that are available to the users to explore

User Study

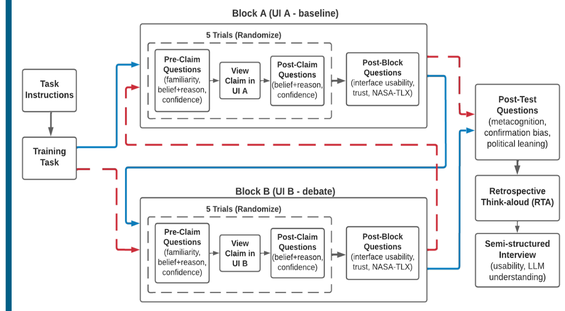


Fig 5: User Study Design

In order to evaluate the effectiveness of our prototype as a tool to mitigate confirmation biases, we are currently performing a user research study that focuses on how users interact with the web app, what features they tend to use most and most importantly – does using the framework help them reconsider their views about the topic at hand. The evaluation of the user experience will be done via a post-test questionnaire and a Retrospective Think-aloud (RTA).

Acknowledgments

This project is being supported by a Microsoft Research Grant in lieu of their partnership with Dr. Matt Lease. Special thanks to Anubrata Das, Gauri Kambhatla and Sooyong Lee from the AI and HCC lab for their valuable contributions and inputs for this project.

References

[1] Bertrand et al. "How Cognitive Biases Affect XAI-Assisted Decision-Making: A Systematic Review." In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, 78–91
[2] Chen et al. "ReConcile: Round-Table Conference Improves Reasoning via Consensus among Diverse LLMs." arXiv, February 21, 2024.
[3] Du et al. "Improving Factuality and Reasoning in Language Models through Multiagent Debate." arXiv, May 23, 2023
[4] Wang et al. "Unleashing the Emergent Cognitive Synergy in Large Language Models: A Task-Solving Agent through Multi-Persona Self-Collaboration." arXiv, January 4, 2024.