👨‍⚖️ Ethical Considerations

This is the whole project ethical considerations, details of ecah sprint can be accessed through sprint2 👨‍⚖️ Ethical Considerations for Sprint 2 and sprint 3 👨‍⚖️ Ethical Considerations for Sprint 3

Introduction

In the development of the Furhat Q&A platform, our team has endeavored to align with the highest ethical standards. This document details the specific ethical considerations addressed throughout the various phases of our project. The focus here is on practical, project-specific issues such as data management, user privacy, bias mitigation, and the implications of AI integration with robotic technology.

🔐 Data Privacy and Security

Consideration: The data used for our project should not violate privacy laws or intellectual property rights of the data sources and effective measures should be made to ensure data security.

Action: Our web scraping method only obtains data from the given URLs. These URLs are all publicly available webpages containing no personal informations that may involve privacy issues. We also avoid any content that requires login credentials or is marked private. In the development process, we have implemented strict access controls to ensure that only authorized personnel have access to the data and the backend of the system. The use of this data is strictly confined to the project's scope—powering the language model for the Furhat robot. For sprint 3, we ensure that no personal information is collected when the user interacts with our QA system.

⚖️ Bias and Fairness

Consideration: Our QA system should not perpetuate or amplify biases present in the retrieved context provided by RAG, avoiding unfair treatment of certain groups.

Action: The bias of the answers generated by LLM is inherited from the retrieved documents of RAG, since RAG will not actively create data, the bias is mainly from the given URL. For sprint 3, considering that LLM itself is trained on a large diversity of data, we rely on its capacity to avoid possible bias as much as possible by modifying the content of the prompt of LLM to adjust the final answer. We can also finetune this process by mannually testing with data from different sources.

🔍 Transparency and Explainability

Consideration: Maintaining transparency throughout the development process is essential to build trust with all stakeholders. Good explainability helps users understand how and why the Furhat robot provides specific responses.

Action: We provide clear documents about the implementation details of our system including data processing (the source from which data is scraped and data processing steps), how RAG retrieves the relevant data, and how LLM use it to generate the final answer. By only using responsible third-party packages, we developed the system with explainable AI principles. The RAG technique we used for our project helps the robot to explain how it arrived at that answer since it can cite the retrieved documents from data sources. In sprint 3, we tested different RAG methods to ensure the reliability of our system.

🦽 Accessibility

Consideration: Ensuring that the system is usable by people with various disabilities, including those with visual, auditory, cognitive, and physical impairments.

Action: With the limited time, our system currently only has features like speech recognition and text-to-speech for visual, auditory impairments, which are provided by the Furhat robot. In sprint 3, we have tested these features to ensure effectiveness.

🟦 Inclusivity

Consideration: The system should be inclusive, catering to users from diverse cultural, linguistic, and socio-economic backgrounds.

Action: The furhat robot provides both virtual and real ways for users to interact. In virtual way, a webpage is provided with simple and intutive design to minimize the cognitive load, making it easy for user to navigate. For real interaction, the furhat robot is served as a guider, providing multiple language support and different communication skills to better help users to understand how to interact with it.

Conclusion

In conclusion, our project's approach to ethical considerations is deeply ingrained in every phase, from data collection to deployment. By maintaining a steadfast focus on privacy, fairness, transparency, accessibility and inclusivity, we strive to ensure that the Furhat robot not only serves its intended purpose but does so in a manner that is respectful and ethical. This commitment to ethical excellence is fundamental to our project's success and is a cornerstone of our ongoing development efforts. Through continuous ethical evaluation and adaptation, we aim to set a standard for responsible AI development in interactive and robotic technologies.