👨‍⚖️ Ethical Considerations for Sprint 3

🔐 Data Privacy and Security

Consideration: The data used for Sprint 3 should not violate privacy laws or ethical guidelines 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. 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. We also 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. LLM itself is trained on a large diversity of data, so 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.

🔍 Transparency and Explainability

Consideration: Users should 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 also helps the robot to explain how it arrived at that answer since it can cite the retrieved documents from data sources.

🦽 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. 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.