Daniel Berry
HCC 6600
March 2, 2025

Report 4

1. How does XAI improve trust and transparency in your project?
   Users are able to determine which aspects of the project came from the AI back-end. They should be able to understand the logic behind the conversations that will allow them to build up faith in the effectiveness of the software.

2. Which XAI technique (visual, textual, example-based) is most applicable to your project and why?
   The open source RASA chat-bot has tooling that allows for easily follow-able stories and conversation threads. These can be explained textually and through visual decision tree diagrams.

3. How does the need for effective human oversight influence the XAI methods you will use in your project?
   The human oversight aspect of the project is the availability to report incorrect or harmful suggestions from the chat-bot to a real person. This is an easy to access feature that users will know about from the beginning.

4. What are the key biases and limitations in your project that XAI could help identify?
   Differences in how people might react to different situations in another culture and language. The conversation flow that is mapped out in the RASA dialogue trees helps to explain a natural flow of conversation and how things are expected to go.

5. Choose an XAI method that would significantly enhance the interpretability of your project's AI system. Why this one?
   The DIME method would be able to analyze the Natural Language Understanding data and give insight into how different words will affect the flow of conversation. I chose this tool because it was the only thing I could find that would work with the Rasa project.

6. How does the challenge of balancing interpretability and model complexity manifest in your project?
   It shouldn't have a detrimental effect towards the complexity of the model. The conversations will be modeled to be simple to progress through and understand.

7. In what ways will XAI facilitate better human-AI collaboration in your project?
   Users will have a better understanding of the role of the chat-bot in the interaction.

8. What role does XAI play in understanding and mitigating biases within your project's AI system?
   Users may understand why a certain response was given and understand the importance of providing feedback or corrections to a chat-bot response.

9. How will you implement model-agnostic XAI methods like LIME or SHAP in your project?

   There is an application called Dime-XAI that is setup to work with classifiers for Rasa NLU: https://github.com/DIME-XAI/dime-xai It will require some further research to get working and determine if its a viable tool for the project.

10. What steps will you take to ensure the explanations provided by your project's AI system are robust and reliable?
    During feedback interviews we can include questions that ask for users to give their best understanding of the project before and after reading through any explanatory text.