Daniel Berry
HCC 6600

February 16, 2025

Report 2

February 16, 2025

1. How could algorithmic bias impact the decision-making processes in your project?

   The large-language model that is used can be trained to not use naturalistic Italian that would be seen from a native speaker. I might trust that the speech it is producing is authentic Italian, but since I am not a native speaker, I may not be able to accurately evaluate when the algorithm is incorrect.

2. Which forms of bias (sampling, selection, measurement) are most critical to address in your project and why?

   If the model is trained with text written by non-native speakers, then the produced conversations may be inaccurate. This is an example of sampling bias where a high proportion of non-native speakers could be contributing to the large-language model. Users may get tips or learn poor grammar in the language they are trying to study.

   Gender bias will be another issues to identify as the model may not have been trained to work well in a language like Italian that uses gender as a core construct.

3. How might biases in data collection affect the outcomes of your project's machine learning models?

   The trained Italian chat bot might use incorrect grammar or break cultural norms that do not exist in the United States, but would be out of place in Italy.

4. Identify and explain a societal or systemic bias that could influence your project's algorithm design.

   Interactions between people of different genders or races may receive different interactions from the chat bot. I do not have good insight into how the model was trained and need to evaluate for this possibility.

5. Choose a type of bias discussed this week that poses a significant challenge to your project and describe a strategy to mitigate it.

   Gender bias will be a barrier for this project when it comes to returning chat bot dialogue that is specific to the user. At the start of the website, users will select a gender, and all dialogue from then after will need to be appropriate for that case.

   I will use Word Embedding Association testing to verify that dialogue is accurate before returning it to the user. Then if inaccurate, we will need to ask the LLM to fix it.

6. How will you implement random sampling or reweighting to reduce sampling bias in your project?

   I do not have any control over the training of the inference model that I will be using and cannot implement any changes to the weights or sampling. This is not something I can do for this specific project.

7. Based on the biases covered, which mitigation technique do you plan to apply specifically to your project?

   I will not have any control over the training of the large-language model, but I will have control to limit what responses are given back to the user. I intend to limit topics that trigger certain sentiments. Similarly, testing the word embedding associations will allow the application to

8. Discuss the importance of diversity among your project's technology designers in combating algorithmic bias.

   I will need to get advise from experts in the field of language learning and from native Italian speakers. Without their insight I might let slip issues with the product and how it handles cultural interactions.

9. How will you address representation and data collection biases within the context of your project?

   I am reaching out to a friend who runs a language learning business and online community. Additionally I will lean on previous coworkers who are native Italians that can aid me in testing the prototype. The people I reach out to will be of both American and Italian descent and will be of different genders. Reaching out for input, expertise, and feedback from multiple demographics will help to catch any biases.

10. Outline your plans for integrating fairness-aware machine learning techniques into the next phases of your project.

    The Word Embedding Association Test (Word2Vec) example can help to identify issues of gender bias in the text of the responses from the chat bot.