Daniel Berry
HCC 6600
March 23, 2025

Report 6

1. How does automatic decision-making improve your project's efficiency?
The project has automated policies in place to analyze user prompts and adjust the response if particular flags are thrown. The grammar check for instance monitors for English prompts and blocks the LLM from continuing on to produce a response.

2. How will you address data biases in your project?
I will have to carefully curate and remove any training data that is aggregated from users.

3. What steps will you take to ensure transparency in your project's decisions?
Anybody can find full access to the source code of the project, tooling and back-end documentation. Any automated decision from the grammar check is displayed to users on the front-end. An explanation of the system is available through the help module.

4. Which probabilistic method will you use for decision-making in your project?
If a prompt is more than 80% English, it will block a response and force the user to work in Italian.

5. How will heuristics simplify decision-making in your project?
When weighting certain interactions to use for training data, I can use different heuristics like app usage and grammar ratings to determine how good the data is to use.

6. How will Bayesian updating enhance your project's model accuracy?
Over time, users will add in more interactions and more data. Over time we can build a more complete model and get a better rating of how a user could possibly be malicious. Over time, a good guess can be made of whether or not a user account is malicious.

7. What role will Markov models play in your project's decision processes?
If users are continually showing different attributes of possible malicious intent, they can be automatically throttled and then banned for a short time from the site. If continued malicious activity continues they can be permanently banned from the site.

8. How will Monte Carlo methods be applied in your project?
I can fuzz and randomly operate the chat bot api to determine how it may react the majority of time and during outliers.

9. How will multiagent systems enhance decision-making in your project?
There are three LLM systems in the project that interact independently and in line with one another. The Rasa system sends it's responses through ChatGPT to verify grammar and check for offensive content. User prompts are similarly graded for grammar through another ChatGPT connection.

10. What actions will you take to prevent over-optimization in your project?
Training will setup the chat-bot to tend towards chit chat and avoid letting the user resolve interactions as quickly as possible. Longer interactions more heavily weighted for later training too. This way the project does not attempt to steer to the quickest possible interaction.