AI Itinerary Reliability Test Report

The Rolling Capstones

27 September 2024

1 Objective

The purpose of this test is to evaluate the reliability of AI-generated itineraries created from both form-based and chatbot-based inputs. The goal is to assess the AI's ability to generate accurate, feasible, and consistent itineraries that align with user preferences and constraints.

2 Test Criteria

The reliability of the AI-generated itineraries was assessed based on the following criteria:

- **Accuracy**: The itineraries should meet user-specific preferences, such as selected activities, accommodations, and budget.
- **Feasibility**: The itineraries should be practical in terms of time and resources (budget, accommodation, and activities).
- **Consistency**: The same inputs should generate similar itineraries across different tests.
- **User Satisfaction**: Users rated their overall satisfaction with the itineraries.

3 Test Environment

- **Methodology**: The AI model was tested using two input methods—via a form and via a chatbot.
- **Participants**: 40 users participated, with each user generating five itineraries through both the form and chatbot interactions. The users represented a diverse range of travel preferences and profiles.
- **Test Setup**: Each participant provided data such as their destination, travel dates, interests, and preferences through the form and chatbot interfaces. The AI generated personalized itineraries based on these inputs.

4 Test Scenarios

Each user was asked to create five itineraries based on varying preferences (budget, interests, destination, and accommodation), both through the form and chatbot. The reliability of each generated itinerary was then assessed according to the following scenarios:

- **Form-Based Interaction**: Users filled in their preferences through a structured form, and the AI generated itineraries accordingly.
- **Chatbot-Based Interaction**: Users interacted with the AI via a conversational chatbot to create their itineraries.

5 Test Results

5.1 Form-Based Interaction

- **36 users** (90%) found all 5 of their itineraries reliable and sufficient for their needs.
- **2 users** (5%) found 4 out of 5 itineraries reliable. The single unreliable itineraries in these cases missed key preferences—one did not include the user's favorite golf course, and the other failed to recommend their preferred hotel.
- **2 users** (5%) found 3 out of 5 itineraries reliable. The two unreliable itineraries either did not suggest a preferred activity or did not recommend a favored hotel.

5.2 Chatbot-Based Interaction

- **38 users** (95%) found all 5 of their itineraries reliable and sufficient.
- **2 users** (5%) found 4 out of 5 itineraries reliable. The unreliable itineraries in these cases did not recommend the users' preferred restaurants.

6 Analysis

The overall performance of the AI-generated itineraries was strong, with the majority of users finding the results accurate, feasible, and consistent.

- **Form-Based Interaction**: The form provided a structured way for users to input their preferences, and most itineraries aligned well with user expectations. However, there were a few cases where the AI failed to capture specific user preferences, particularly for specialized requests like favorite golf courses or specific hotel chains.
- **Chatbot-Based Interaction**: The chatbot interaction was found to be slightly more reliable, likely due to its conversational nature allowing for dynamic input. However, the few unreliable results stemmed from missing recommendations for specific restaurants, indicating that the AI needs further improvement in handling precise dining preferences.

7 Recommendations

To improve the AI-generated itineraries further, the following steps should be taken:

- **Enhance Preference Matching**: Improve the AI's ability to capture and recommend niche preferences (e.g., specific golf courses, hotels, restaurants) by expanding the data set and refining the preference-matching algorithms.
- **Refine Data Inputs**: Consider allowing users to input additional preferences in a more flexible manner to reduce the chance of missing out on key requests.
- **Consistency Check**: Ensure that small but critical preferences, such as hotel or restaurant choices, are consistently incorporated into itineraries, regardless of the input method (form or chatbot).

8 Next Steps

To ensure the continued improvement of the system's reliability, the following steps will be undertaken:

- **Feature Improvement**: Further refine the features implemented based on user feedback, ensuring that itineraries are both more detailed and better aligned with user needs.

- **Testing**: Create and conduct unit, integration, and end-to-end tests to validate the functionality and reliability of the AI system. This includes reliability testing for both form-based and chatbot-based itinerary generation.
- **Reliability Testing**: Continue performing reliability tests to assess the accuracy of the AI's recommendations and its ability to adapt to diverse user preferences. Special attention will be paid to specific user requests for accommodations, restaurants, and activities to improve precision.