SEVEN
Anna, Aundre, Kai, Kaitlyn, Ye Wei

UCRE M8 Deliverable D: Assumption Artifact Test Results

Summary Report

Executive Summary

We wanted to test within the context of where users would frequently utilize gen-AI, so we can accurately gauge how participants would act in real life. For our prototype of our potential solution, we created a small mini-game to complete for first-time users of ChatGPT to complete before they can access the AI tool, meant to teach about AI bias through a series of different scenarios with discrimination and bias. We tested with undergraduate students with a variety of background in AI bias, utilizing a think-aloud protocol, with the feature prototyped on Figma. After the think-aloud protocol, a google form was used to collect quantitative and qualitative data on how much they enjoyed the mini game, along with whether or not their understanding of generative AI increased overall.

Based on our user testing, we found that users were confused on the context of the game in the beginning , and lack of information displayed when they answered that the scenario was biased (only showed when answered incorrectly). The game also seemed a bit easy for users and lacked nuance and complexity that AI bias might really contain.

Moving forward, we will be implementing changes, such as changing information displayed when users select “yes” on the scenarios to still provide context for why it biased (regardless if they answered correctly or not), providing context on the situation of the minigame and a brief overview of what the game entails, and adding more nuanced and complex scenarios that are a bit less obvious as the scenarios go on. We gathered important insights on how we can improve the solution overall, and iterate on it to create a better experience for learning about AI bias. We will improve and change aspects of our prototype, through the visual design along with the flow and information given in the feature.

Test & Prototype

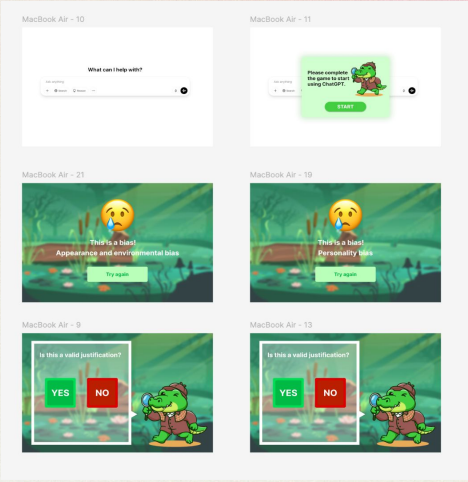
Who: Students @ CMU who've been exposed to gen-AI

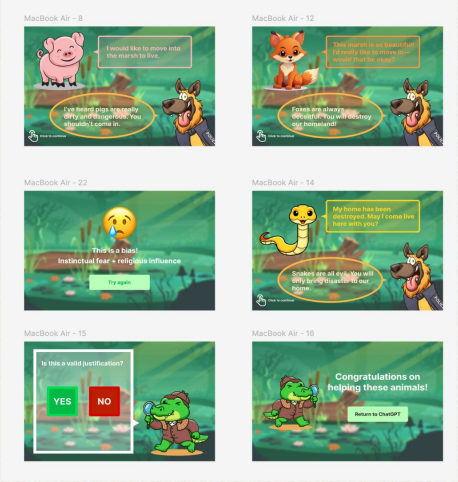
What: Whether participants are engaged and better understand gen-AI bias by interacting with our prototype

Where: Spaces where participants frequently use gen-AI (ie. desk at home, libraries, study spots on campus)

Why: We want to test within the context of where users would frequently utilize gen-AI, so we can accurately gauge how participants would act in real life. Conducting a think-aloud will also allow us to get a sense of what participants will actually do rather than what they say they will do.

How: Think-aloud where participants interact/explore our prototype and commentate their thoughts, feelings, questions. Then we'll conduct a post-session survey/interview.





Successes and Failures

Honest Signal:

a. ***Success*** - Users feel comfortable playing the mini game and are able to learn about AI bias through the experience.
   i. They can understand the bias issue represented by the mini game.
   ii. They can play the mini game without any learning curve.
   iii. Users do not feel overwhelmed by the number of levels in the mini game.
   iv. Users find the game format interesting.

b. ***Failure*** - Users find the mini game boring and are unable to learn anything about AI bias.
   i. Users cannot understand the message the game is trying to convey.
   ii. Users do not want to play the game and feel it is difficult to get started.
   iii. Users feel that playing the mini game is a burden.

Findings & Risks:

- Users find this format very interesting and are able to learn from it.
- More guidance is needed to help users better understand and become immersed in the game.
- The game's difficulty should be increased, and the content should focus more on subtle areas.

Proposed Changes

Based on the testing results (see Appendix Slide 11):

- **Provide context**: Providing a brief overview of the situation of the minigame and what the game entails at the beginning will help users better understand how to connect what they learn from the minigame to evaluating the results of the generated output.
- **Provide explanations**: Regardless if users answered correctly or not, we can provide more information explaining why a scenario is biased and how it can be identified.
- **More nuance**: Potentially more nuanced and complex scenarios that are a bit less obvious as the scenarios go on will be more realistic to the actual output that users may encounter and help them examine subtle bias through a more critical lens.

Next Steps

Risk: We find that many users who are unaware of how AI models generate the responses to user prompts are less confident about identifying bias in said responses. We need to test the relationship between understanding of AI models to confidence in identifying bias to see if a users understanding of the product affects their ability to audit bias within it.

Dimensions: A possible prototype could be a quiz-game visually structured as AI trivia with questions related to the workings of generative AI, and other questions with examples of possible biased outputs from general prompts. For non-technical questions, users would be asked if they detect bias from the response, and how confident they are in their belief. The interaction would be a timed multiple choice quiz to get users authentic reactions to the questions similar to how they would interact with gen-AI. The technical questions would range from basic to advanced to gauge different levels of understandings from users.

Metrics: To determine success we would collect data to rate a user's understanding of how AI models work, and compare it with how confident they rate themself in identifying bias in possible outputs from user prompts

Appendix

Google Drive Links

M8: Deliverable B Assumption Artifact Testing Plan

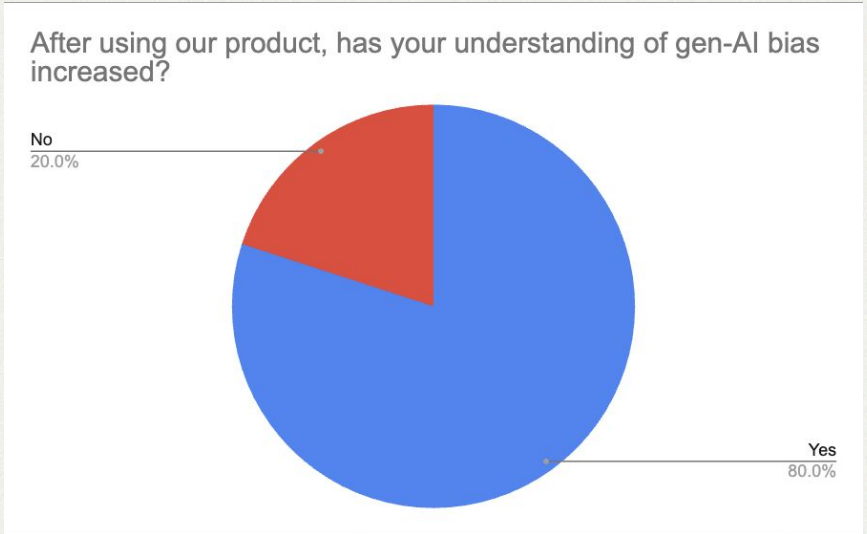
M8: Deliverable C Assumption Artifact [Team]

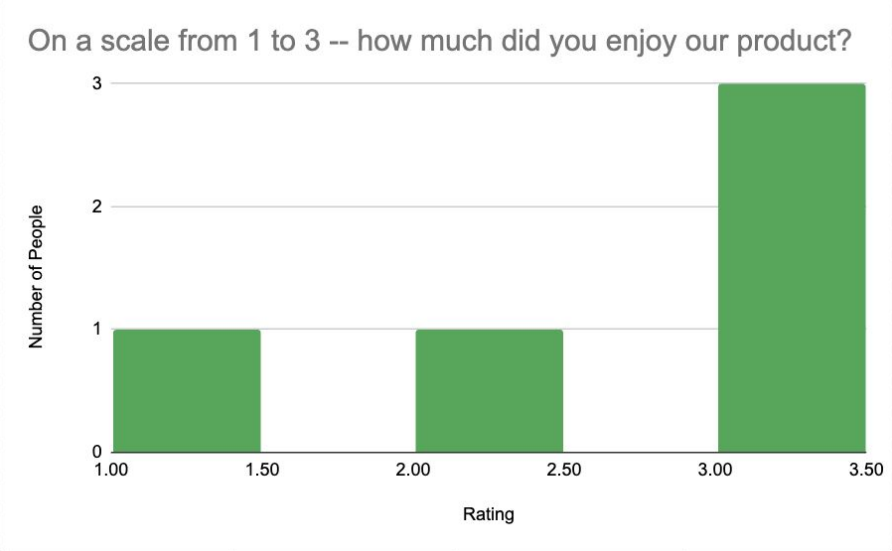
M8: Deliverable D User Testing Consent Forms

M8: Deliverable D User Testing Notes

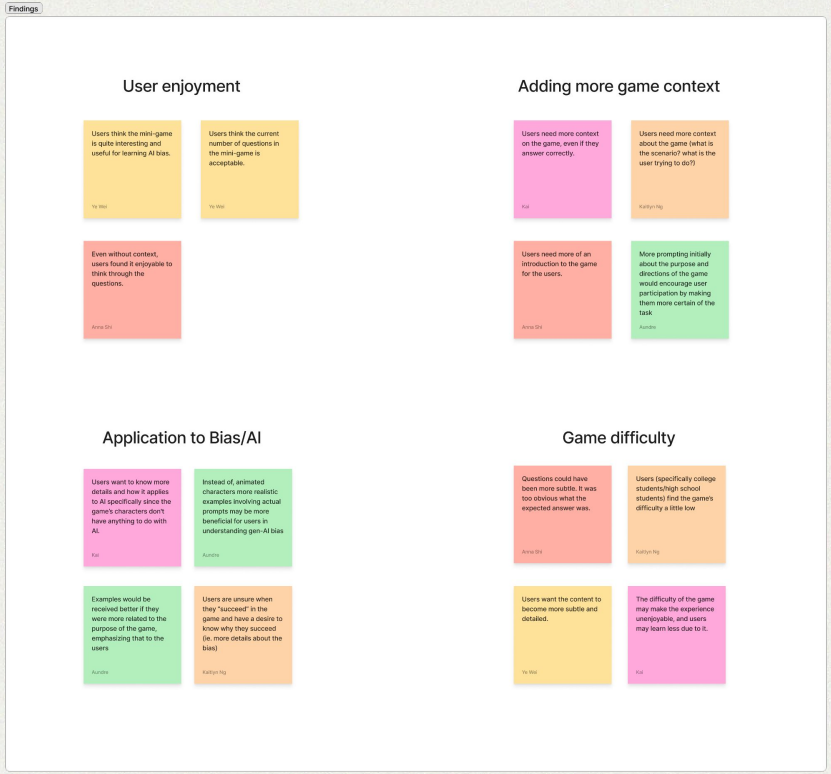
M8: Deliverable D User Testing Post-Session Survey

Quantitative Data Charts





Synthesizing Data/Findings



[Figjam Link]