SEVEN
Anna, Aundre, Kai, Kaitlyn, Ye Wei

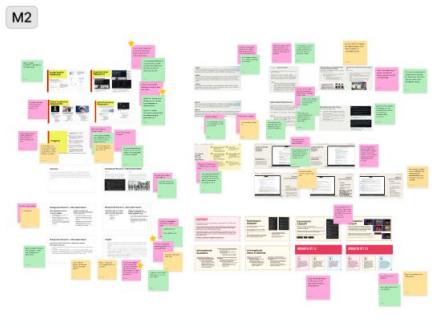
UCRE M7 Deliverable C: Speed Dating

Walk the Wall

Figjam Link

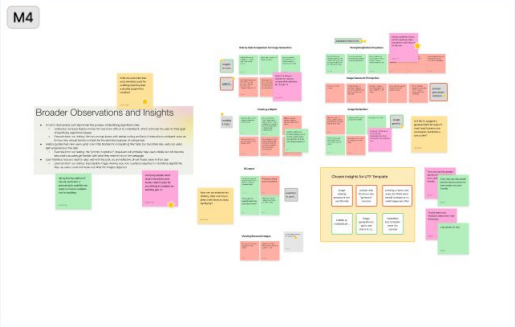
The Wall

M2



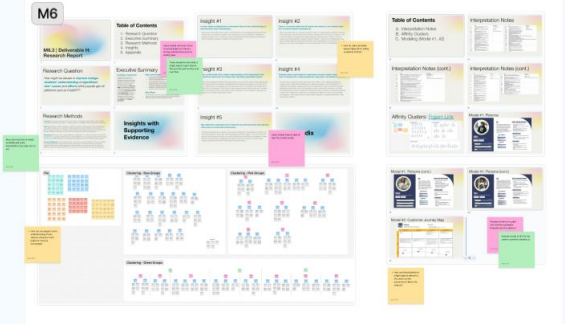
M3

M4



M5

M6



List of Needs, Design Ideas, Questions

Needs (Pink Post its)

- Users need to know what AI bias is in order to identify it
- Solution needs to be quick and intuitive to properly integrate into the platform
- Users should have an idea of how the system works
- Users should not have to face technical jargon or require a strong understanding of AI to identify bias
- Users need a constant reminder that bias exists in order to put in the work to identify it

Design Ideas (Green Post its)

- Solution needs to fit into the current user flow naturally to
- Platform should show examples to demonstrate what bias looks like
- Platform should provide multiple outputs to allow users to easily compare and recognize bias
- Bias reporting feature needs to visible and easily accessible to any user, new or old
- There should be more than a single way to report bias as fits with the user intuition and user flow

Questions (Yellow Post its)

- How can we support users' understanding of bias without using too much jargon or industry knowledge?
- What kinds of visual or interactive prompts can help casual users evaluate output for bias?
- When is bias considered harmful enough to highlight for users?
- How do users' personal biases impact their ability to identify AI bias?
- How can GenAI platforms adapt outputs tailored to the user's current awareness of bias in its outputs?

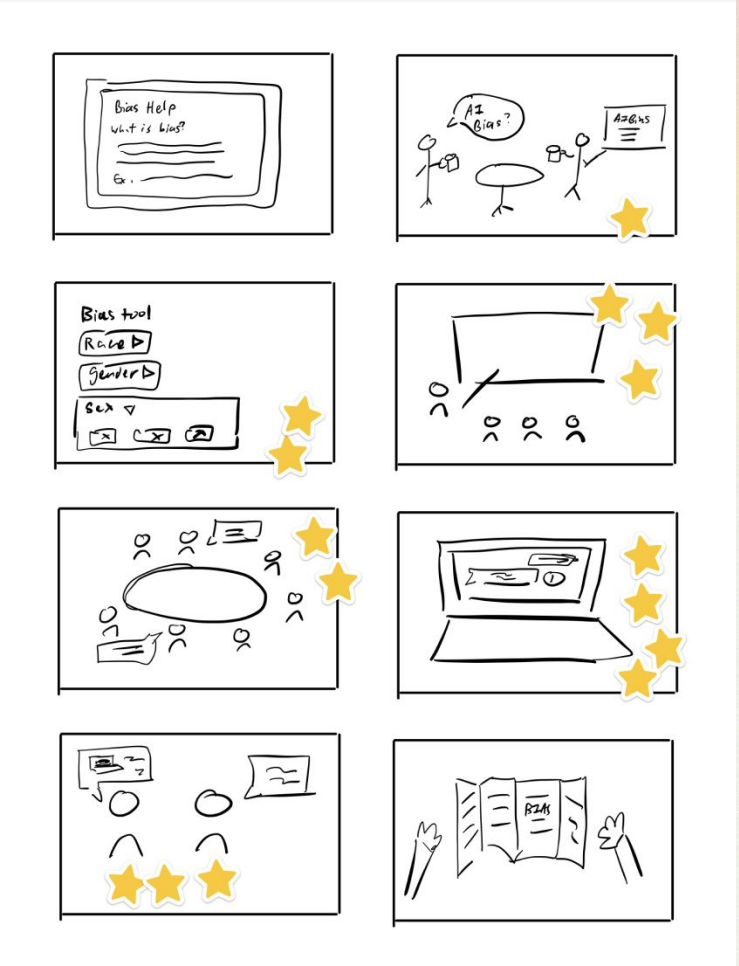
Crazy 8s

User Needs

1. Desire to learn more about AI bias
2. Desire to learn inner workings of AI systems (including source of results)
3. Multiple outputs in gen-AI results
4. Examples to demonstrate AI bias
5. Quick, intuitive, and engaging delivery method

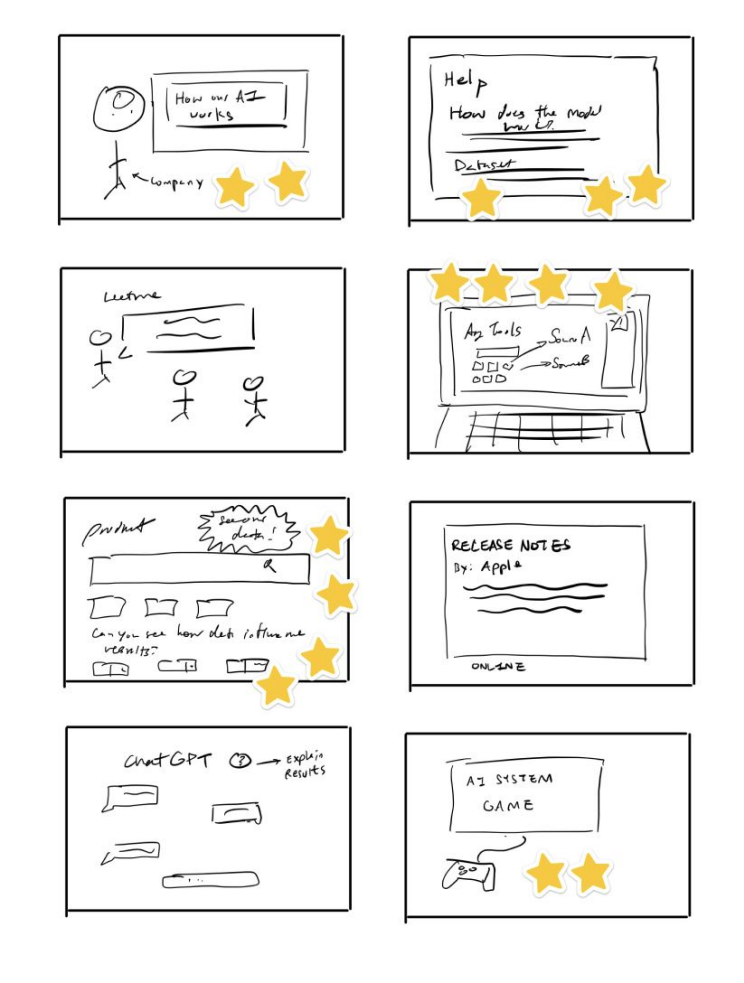
Need #1

Desire to learn more about AI bias



Need #2

Desire to learn inner workings of AI systems

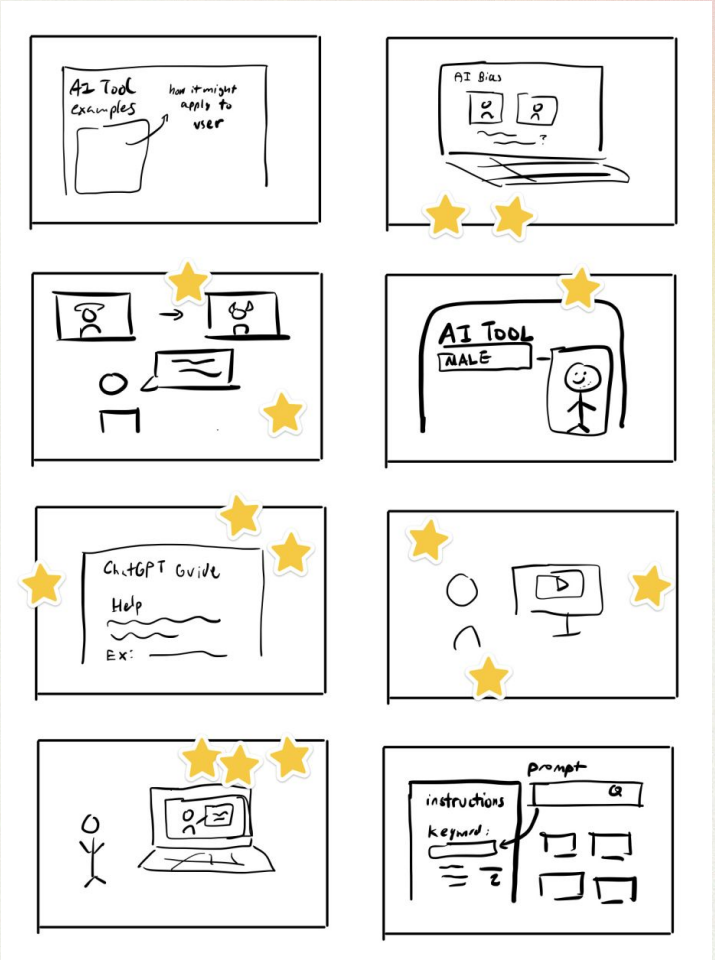


Need #3

Multiple outputs in gen-AI results

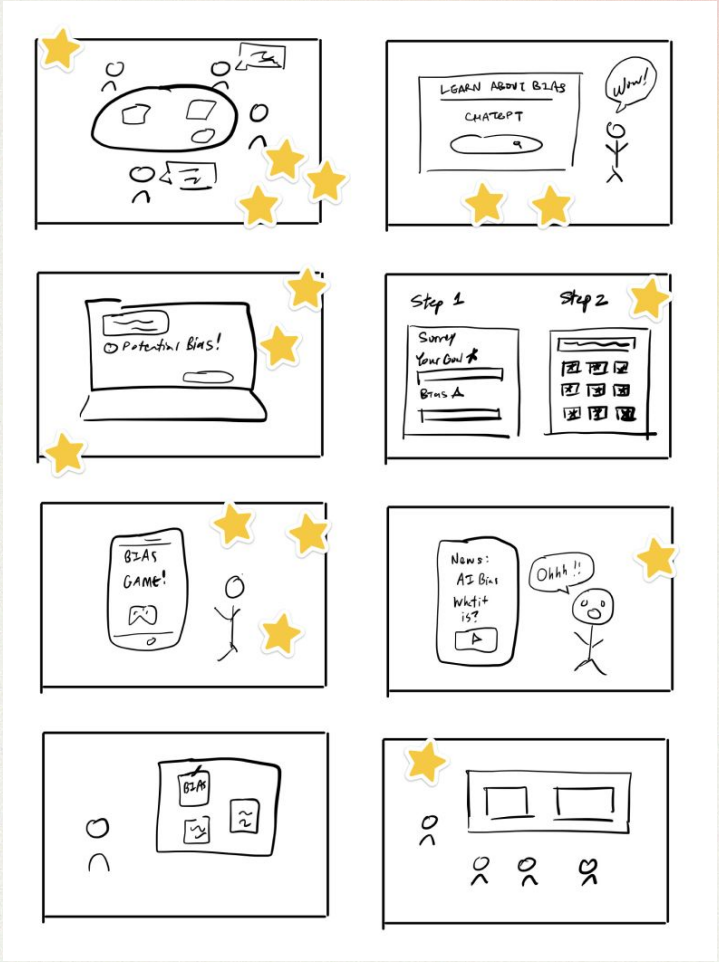
Need #4

Examples to demonstrate AI bias



Need #5

Quick, intuitive, and engaging delivery method



Storyboards

Need #1

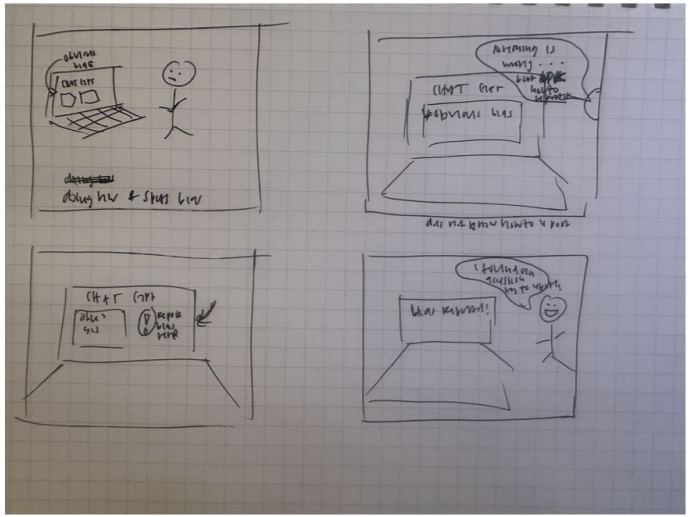
USER NEED: Users should be able to report bias more easily, as well as spot it.

LEAD QUESTION: Have you ever been utilizing an LLM and noticed something off with the answer? Would you and how would you go about it?

DISCUSSION POINTS:

- Feasibility of reporting
- How is the bias detected
- WIll users genuinely engage with reporting?

1st Scenario



Need #1

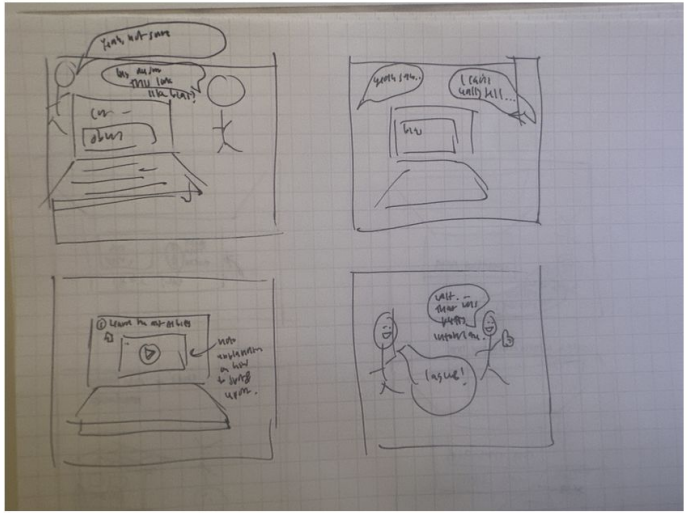
USER NEED: Users should be able to report bias more easily, as well as spot it.

LEAD QUESTION: Have you ever been doing work with a friend and notice that an answer from ChatGPT seems biased? How would you go about reporting bias or seeking knowledge about it?

DISCUSSION POINTS:

- Feasibility of clicking on the video
- Feasibility of accessing video
- WIll users genuinely engage with reporting?

2nd Scenario



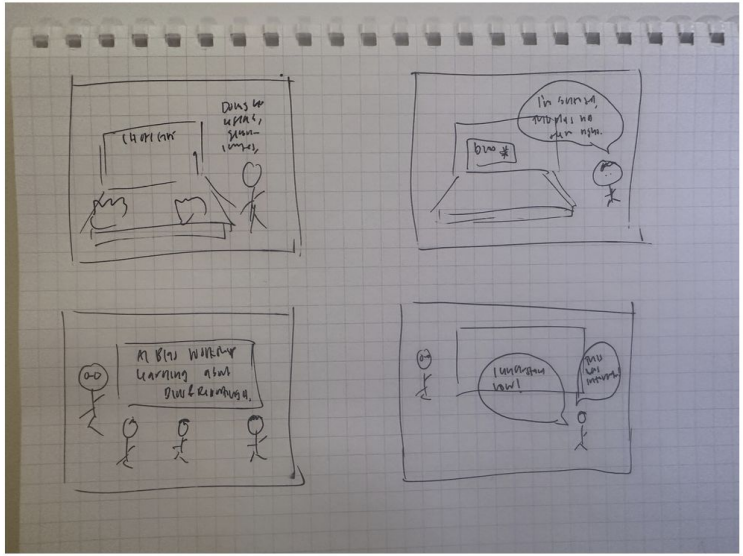
Need #1

USER NEED: Users should be able to report bias more easily, as well as spot it.
LEAD QUESTION: Have you ever encountered a form of bias, but did not know exactly how and what it was? Would you attend an AI bias workshop?

DISCUSSION POINTS:

- Overall motivation to attend
- Accessibility of these types of events
- Will it be internalized?

3rd Scenario



Need #2

Needs

Users want to use AI to generate diverse results, and they need help to find tiny bias problems in the result.

Leading question

Can we just use a suggestion to encourage the user to rethink or regenerate the result?

Discussion Points

- Will this suggestion be effective?
- Can users notice bias problems after seeing the suggestions?
- Will this method interfere user's motivation?

Need #2

Needs

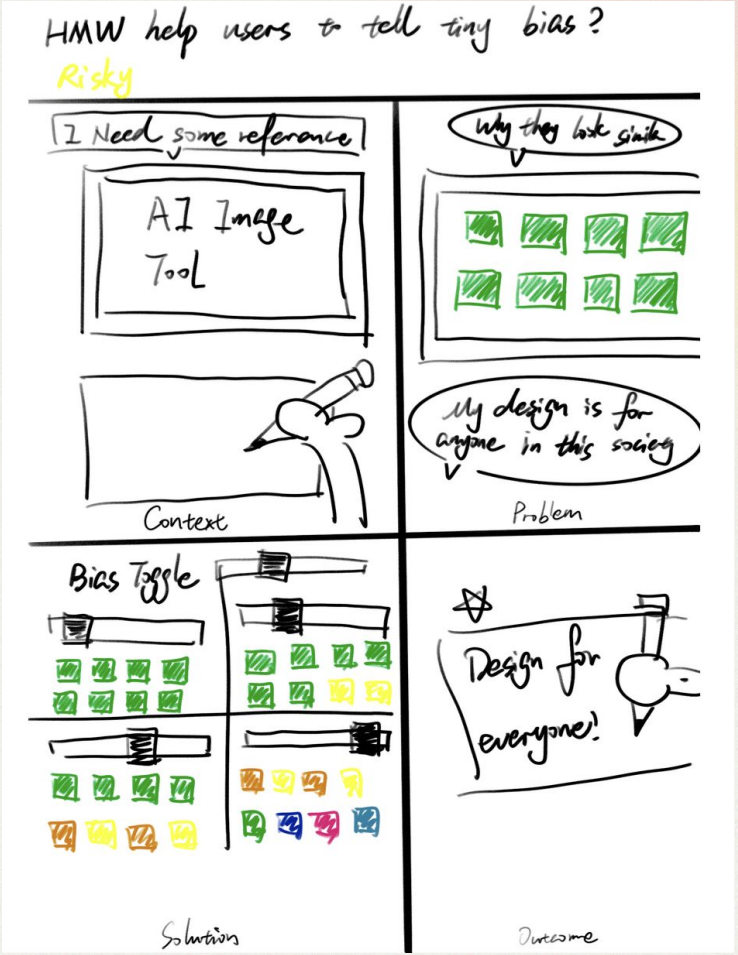
Users want to use AI to generate diverse results, and they want to compare the result to see the bias problems in the result.

Leading question

Can we show diverse result to help user find bias problems?

Discussion Points

- Will users find bias problems when they see different results?
- Will user feel lost then they see so many results?
- Will users find the problems and improve the prompts?



Need #2

Needs

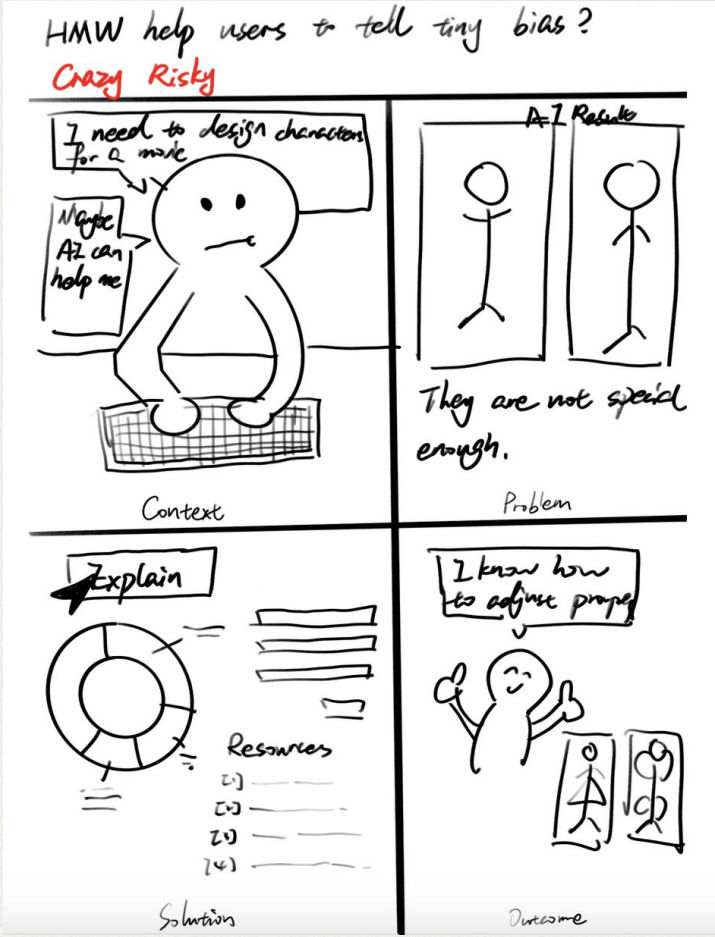
Users want to use AI to generate diverse results, and they want to know how AI generate the result and find the bias problems.

Leading question

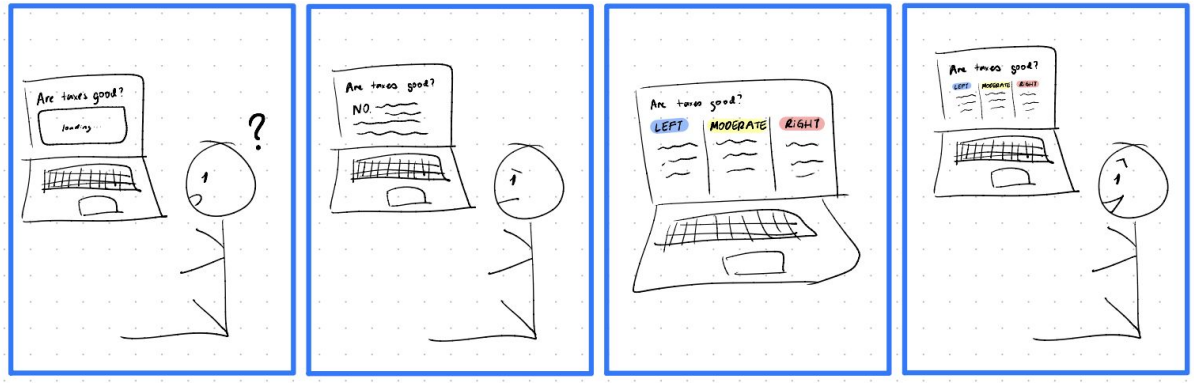
Can we show how AI thinks to help user find bias problems?

Discussion Points

- Will users understand so many information?
- Will they feel overwhelmed?
- What will user learn from these information?



Need #3



context

User is using Gen AI to search up political topics for personal education

problem

User isn't sure if the content generated is biased or not

solution

Platform sees three answers, clearly labelled by the perspective

resolution

User sees three different perspectives, understanding the biases that exist in each

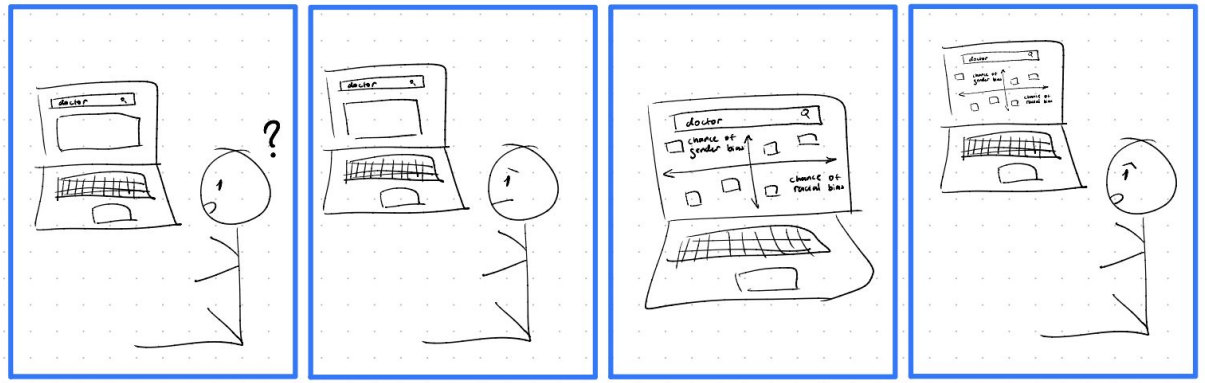
NEED: multiple outputs in Gen AI results

LEADING QUESTION: HMW provide multiple outputs to help users better understand and identify bias?

DISCUSSION POINTS:
- how easily can these perspectives be identified + compared?
- does this increase your trust in gen AI output?

SAFE

Need #3



context

User is using Gen AI to generate images of a doctor for a website

problem

User isn't sure if the content generated is biased or not

solution

Platform shows multiple images of a doctor and suggest how they might rank on certain spectrums of bias

resolution

User compares different results, chooses whichever one best suits theirs needs, and better understands what bias looks like

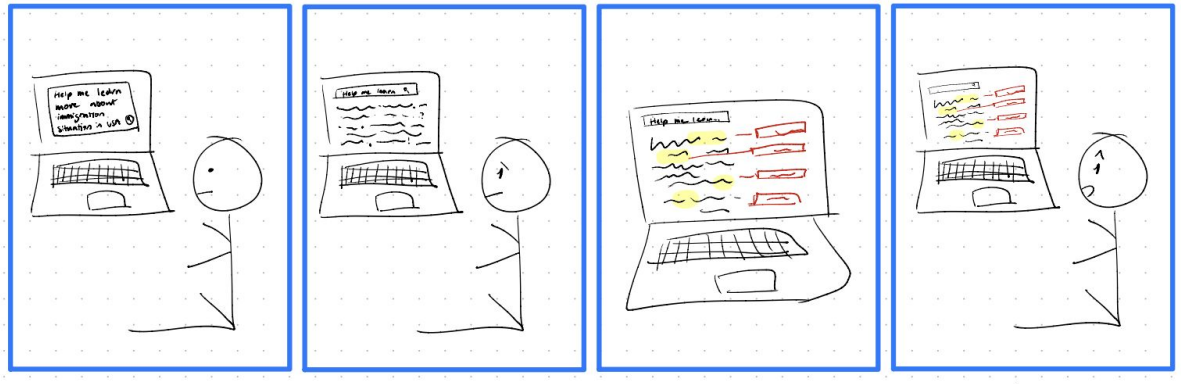
NEED: multiple outputs in gen AI results

LEADING QUESTION: HMW provide multiple outputs to help users better understand and identify bias?

DISCUSSION POINTS:
- would you be comfortable seeing bias talked about so openly?
- would you correct any incorrect bias labels?
- do you think AI can accurately evaluate itself for bias?

RISKY

Need #3



context	problem	solution	resolution
User is using Gen AI to understand what immigration policy is in the USA for personal learning	User isn't sure if the content generated is biased or not	Platform generates its initial result, then highlights specific parts of its generated text and provides an alternate perspective, identifying what kind of bias it might be correcting itself for	User sees what the most expected/ predictable answer might be, but also sees several modified versions of it adjusting for various biases in perspective

NEED: multiple outputs in gen AI results

LEADING QUESTION: HMW provide multiple outputs to help users better understand and identify bias?

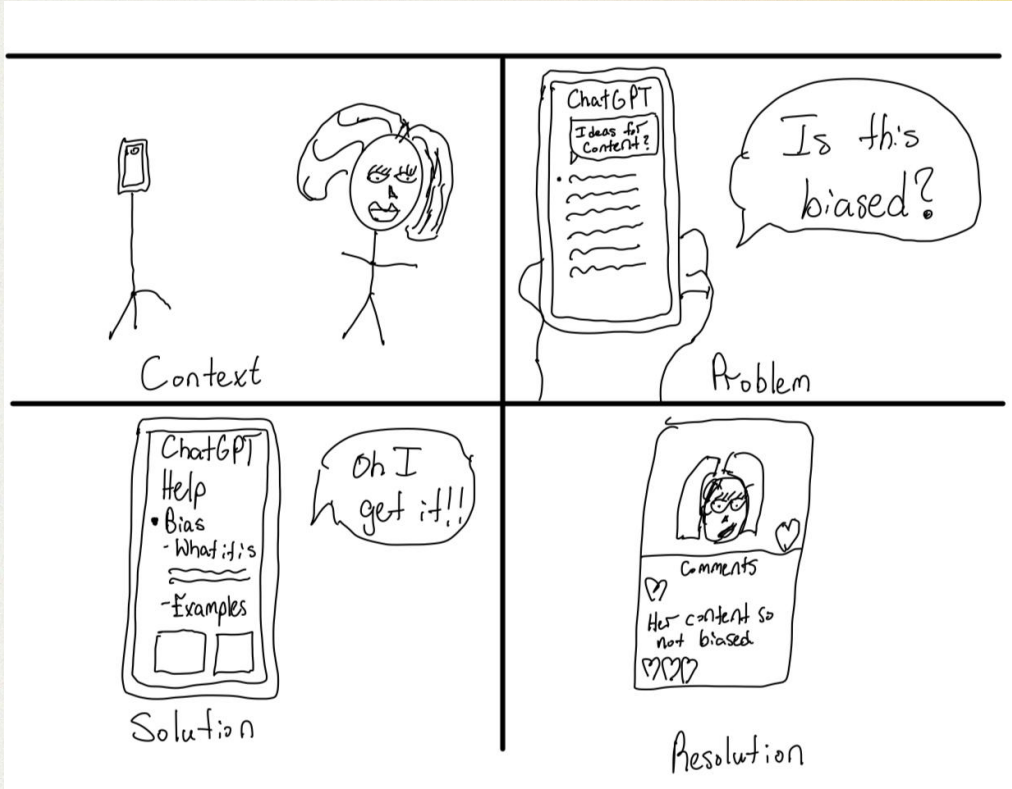
DISCUSSION POINTS:
- is it helpful to see what is considered the "default" answer?
- how would you feel if the output was what you expected, but the modified versions were identifying it for bias?

OUT THERE

Need #4

Safe Storyboard

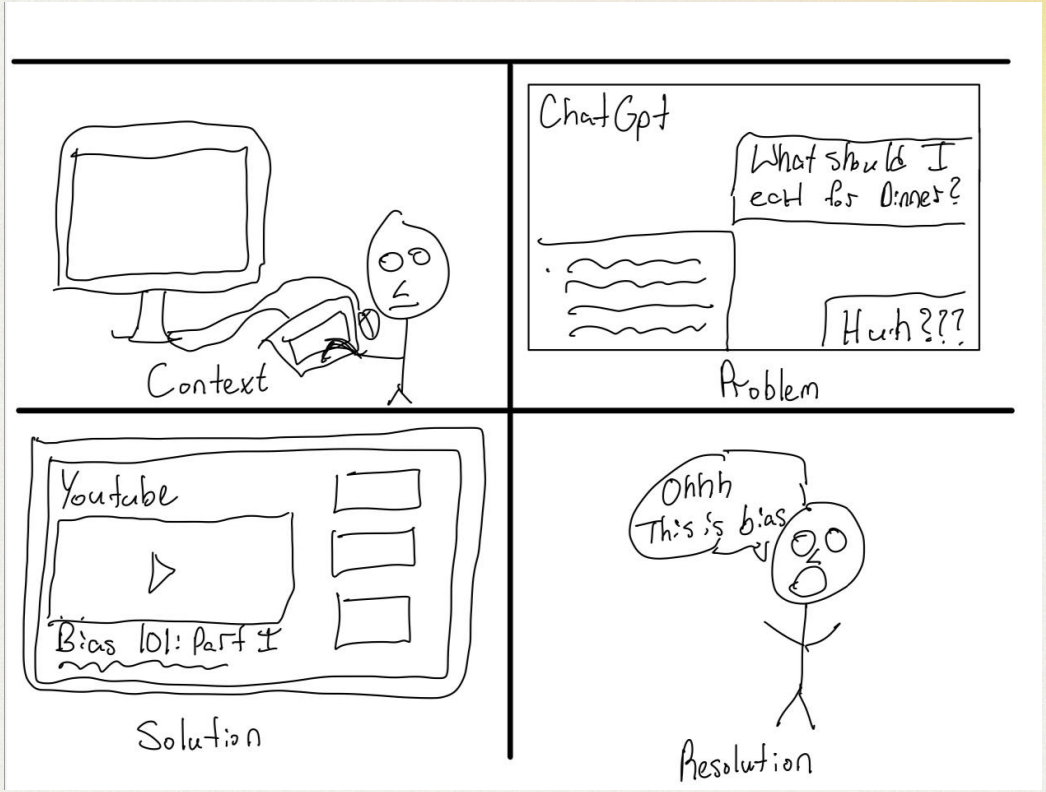
- **User Need:** Examples to demonstrate AI bias to increase familiarity with auditing tasks
- **Lead Question:** What would you do if you were uncertain of possible bias in generated content?
- **Discussion Points:**
  - What do you take away from the drawings?
  - After being told the story, can you envision yourself in a similar position?
  - Would a help page on the applications be useful?



Need #4

Risky Storyboard

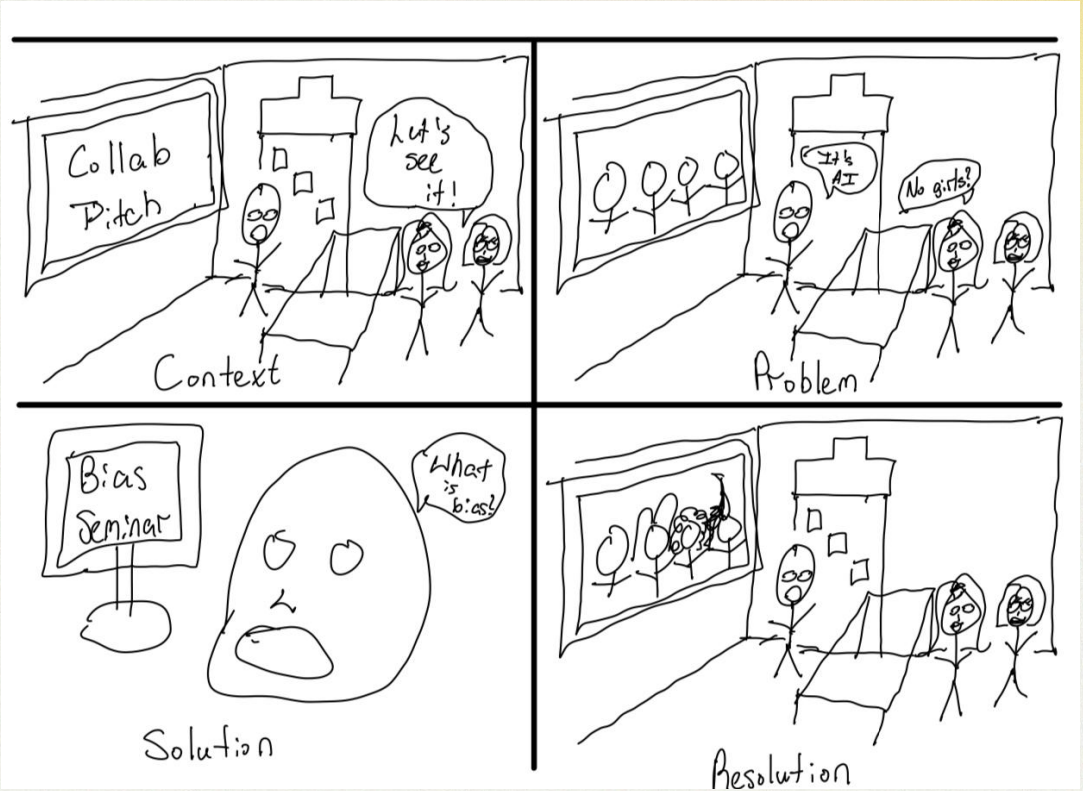
- **User Need:** Examples to demonstrate AI bias to increase familiarity with auditing tasks
- **Lead Question:** Would people spend time to learn more about possible biases via a Youtube series?
- **Discussion Points:**
  - Takeaways from the drawings?
  - Could such a series benefit a large number of users?
  - Is there another format people would gravitate more to for the purposes discussed thus far?



Need #4

Out There Storyboard

- **User Need:** Examples to demonstrate AI bias to increase familiarity with auditing tasks
- **Lead Question:** Do you think attending a virtual workshop on bias would help people better identify bias?
- **Discussion Points:**
  - What do you takeaway from the drawings?
  - What would incentivize you to participate in such a meeting?
  - Do you see such meetings as necessary to address uncertainty with bias?



Need #5

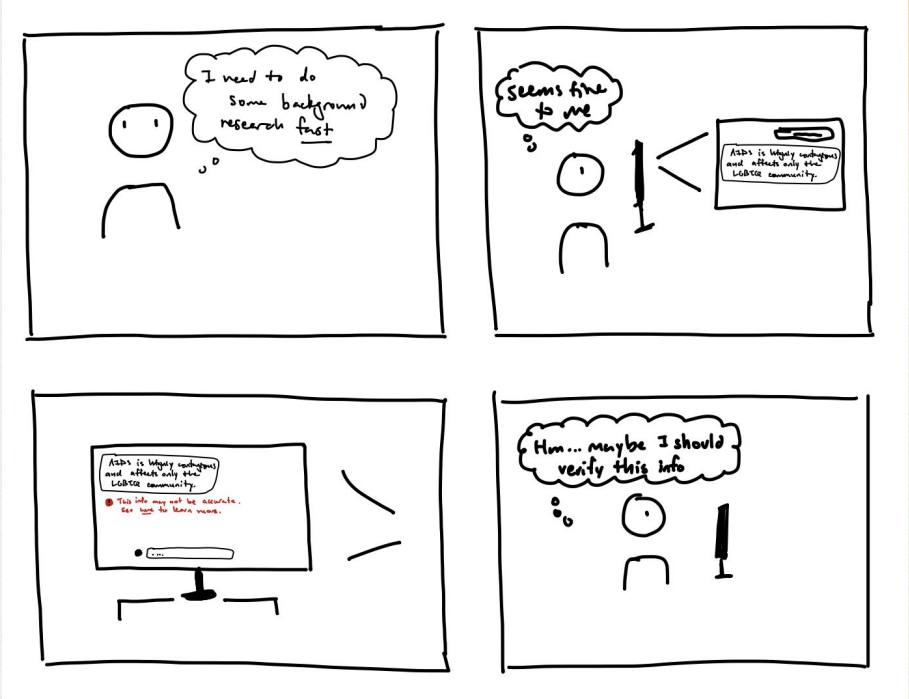
Storyboard #1

Need:
Inform users about bias quickly

Leading Question:
HMW incorporate gen-AI bias information within gen-AI platforms quickly and intuitively?

Discussion Points:

- How likely are users to “learn more” about bias?
- Is it feasible from a developer standpoint to distinguish between biased and unbiased results?
- How might you balance between a “clean” interface vs providing as much information/clarify about bias as possible?



Need #5

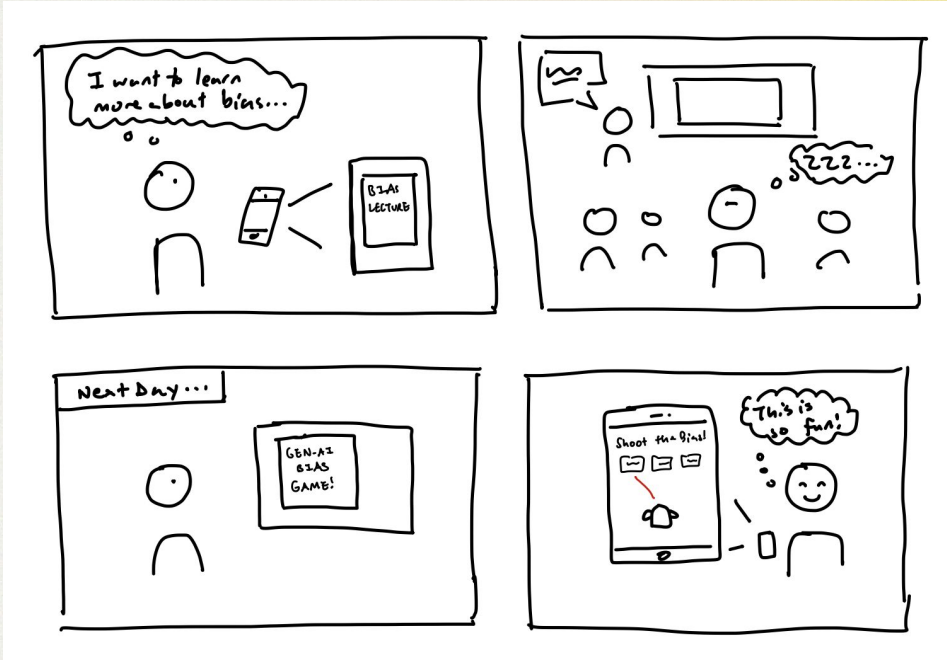
Storyboard #2

Need:
Catch users' attention/engage them regarding gen-AI bias information

Leading Question:
HMW develop an easy-to-use yet educational digital game that teaches users about algorithmic gen-AI bias?

Discussion Points:

- What are some engaging games you've played?
- Have games helped you learn a concept/have games taught you a skill?
- What are some important elements about gen-AI bias to include in an educational game?



Need #5

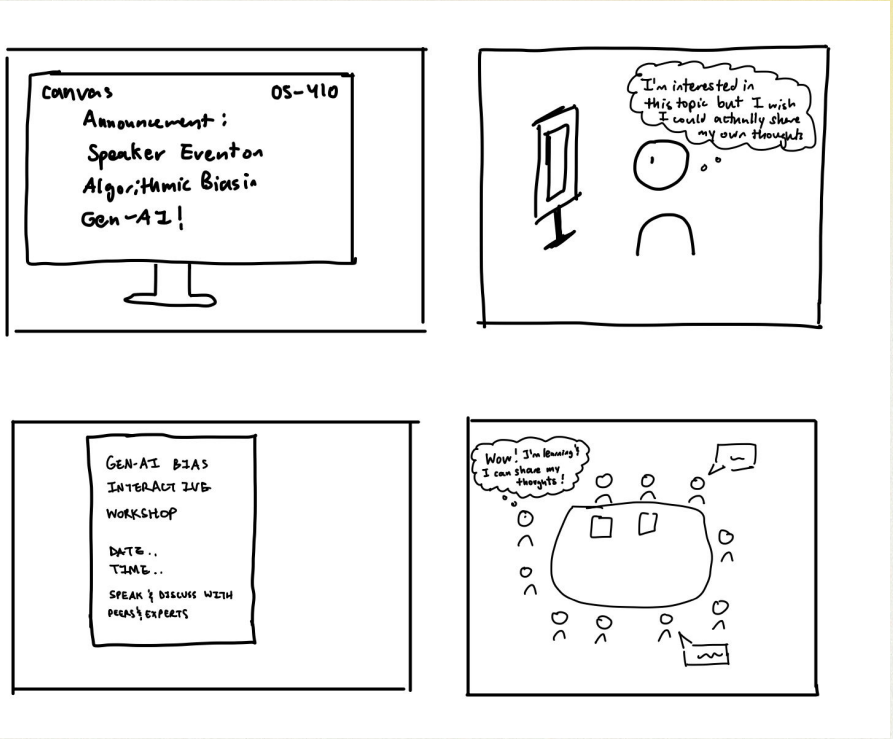
Storyboard #3

Need:
Actively (rather than passively) teach users about algorithmic bias

Leading Question:
HMW design interactive workshops where users come away from them with a new, deeper understanding of algorithmic gen-AI bias?

Discussion Points:

- What would incentivize players to step out of their comfort zone and share their own thoughts on gen-AI bias?
- Would players be willing to take the time to participate in these workshops?
- What are some types of interactive activities that could spark greater conversation about gen-AI bias?



Summary Report

Overall Findings & Insights

Many of our users noted that the most effective solutions addressed user needs without creating much additional work for users, which is typically the main barrier to users reporting bias. Such solutions either added to existing user flows (e.g. adding bias markers to current output in Need #3, Scenario #3) or incentivized users to new user flows (e.g. gamifying bias education in Need #5, Scenario #2).

Meanwhile, most of our users had concerns about live events (e.g. Need #2, Scenario #3) as being particularly risky due to the need for additional time and resources from both the user and the host. They suggested that such solutions would demand a much higher level of commitment to begin with and would likely not be realistic for most users who are using Gen-AI for quick responses in their personal education or project work. Users also had concerns that bias was being defined too rigidly and wanted an acknowledgement of the nuances in bias in Gen-AI output.

Overall, it seems that users tended to gravitate towards solutions that would integrate seamlessly into existing Gen-AI platforms and allow users to identify and understand the nuances of biases without creating an objective definition of bias.

Common Misunderstandings

- **The extent to which users need more bias education**
  - While some users thought that the educational resources were very helpful, others thought that workshops and videos would be too much commitment for a casual user. We need to ensure our solution caters to different user needs as the amount of education is different for everyone.
- **Distinction between incorrect answers and biased answers**
  - Some of the solutions seemed to conflate demonstrating bias with showing correct results. This is something we might want to address in our language to clearly communicate what aspect of the Gen-AI output we want to address (specifically what dimension of bias).
- **Bias reported by the user vs. the platform**
  - Sometimes our solution had the user reporting bias on the platform results, and sometimes it was the user evaluating the platform's ability to report bias. This was a bit confusing to some participants as they tried to evaluate the effectiveness of each. This depends on the solution that we design, but overall we know that recognition usually trumps recall, so our solution may lean towards the latter.

Validated Needs/Design Ideas

- **Diverse presentation formats of AI results can enhance users' understanding without adding significant cognitive load.** By using multiple perspectives and different axes of comparison, users can more quickly identify and comprehend the biases present in the results.
- **Facilitating users in efficiently identifying and editing biased elements within content not only enhances their awareness of bias issues, but also promotes user engagement.**
- **Experts' insights, encouragement from friends, and engaging forms of interaction can motivate more people to actively participate in workshops about AI bias.** This approach directly addresses users' need to understand AI bias and serves as the most straightforward, fast, and effective solution.
- **With its quick and simple mechanics and content rooted in everyday experiences**, the game effectively encourages users to engage in active learning and deepen their understanding of AI bias.
- **Users seek to detect bias issues more efficiently**, as a faster experience preserves their intended use of AI while maintaining productivity.

New Design Opportunities

- **More diverse views for displaying AI results**
  - After generating results with AI, users can choose how the results are displayed. For example, textual content can be shown from different perspectives, while visual content can be organized along horizontal and vertical axes to help users easily distinguish between different outputs.
  - In this process, the system not only presents the generated results, but also enables users to more quickly locate their own results and identify potential bias issues through comparison.
- **Provide users with more options to modify the details of AI-generated results.**
  - After users generate results, provide them with the opportunity to modify the details with AI. For textual information, users can choose the parts they are not satisfied with and let AI make changes. AI will also highlight any potential bias issues in the results. This not only improves users' efficiency but also helps them learn about bias issues during their interaction with AI.

New Design Opportunities

- **A simple and quick mini-game about AI bias.**
  - This game is designed to allow users to play during fragmented time. The content of the game focuses on AI bias, presented through relatable examples from everyday life to encourage users to reflect. The game also includes features for interacting with friends, along with explanatory videos and information about AI bias-related activities.