

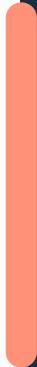
Module 7 | Deliverable C:

Speed Dating [Team]



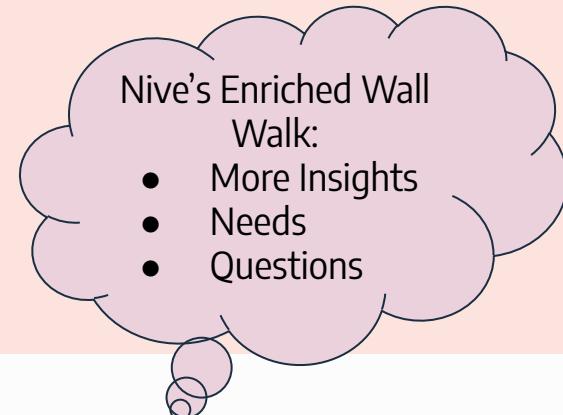
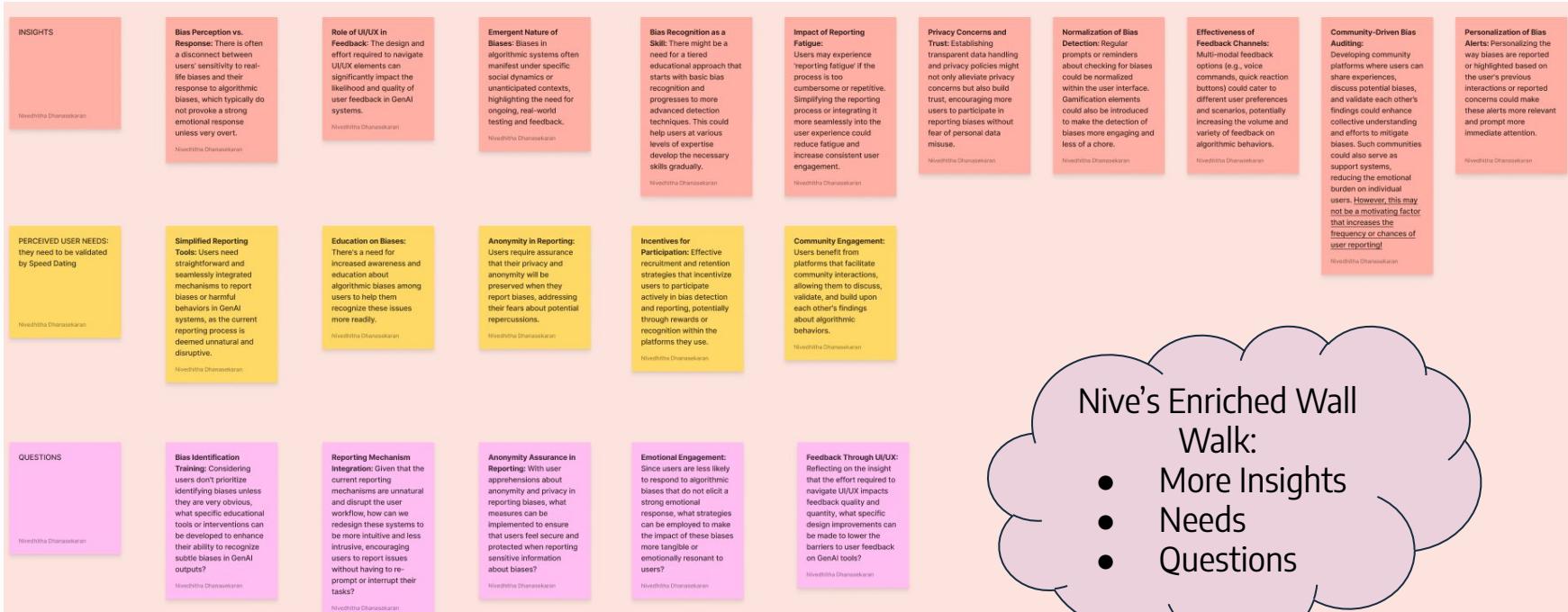
Team Nasa GenAI Vanguard

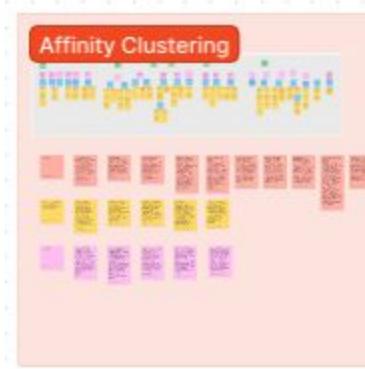
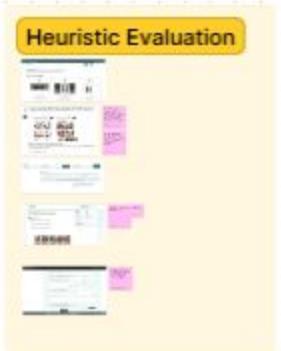
01



Synthesize by Walking the Wall

<https://www.figma.com/file/BGVIKWsaz6MlvXN4bZBwRq/Team-D1-Team-Contract-Spring24?type=whiteboard&node-id=0-1&t=VA3PT9n4LPQK9Jiq-0>

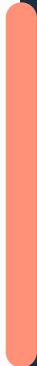




Download

<https://www.figma.com/file/BGVIKWsaz6MlvXN4bZBwRq/Team-D1-Team-Contract-Spring24?type=whiteboard&node-id=0-1&t=VA3PT9n4LPQK9Jiq-0>

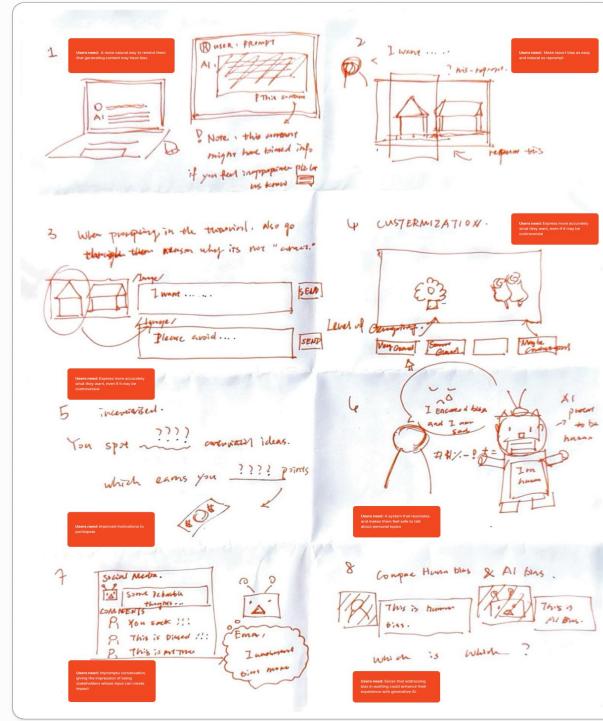
02



Crazy 8s



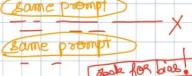
Alec's Crazy 8s_



Amanda's Crazy 8s_

Nive's Crazy 8s_

HMW: What are some motivating strategies that will nudge bias reporting into users' natural flow of interaction with GenAI?

	
Some prompt	Tokens:
Some prompt	> 100 → GPT4o > 50 → 15 free ... GPT4o prompt
think for bias!	I: 80 current
① Callout reminders	② Collect tokens to upgrade subscription
 Good bias report	 multiple wrong prompts
\$1 per bias report	 circle tokens
③ Monetary reward	④ Circle the bias! \$
 find the biased picture	use #report
 Security - Are you human?	 P Li X
⑤ Re-captcha style for free versions	⑥ Post directly to social media w/ a hashtag + OpenAI platform developers
 bias 1 tokens	Try problematic prompts? 
bias 2 tokens	genotypes phenotypes What did you observe?
⑦ Public discussion forum-Piazza style - Participation leaderboard	⑧ Focusing News - Prompt Suggestions & tagging

Sofia's Crazy 8s_

How might we...
make it easier for users to report AI-gen. bias?

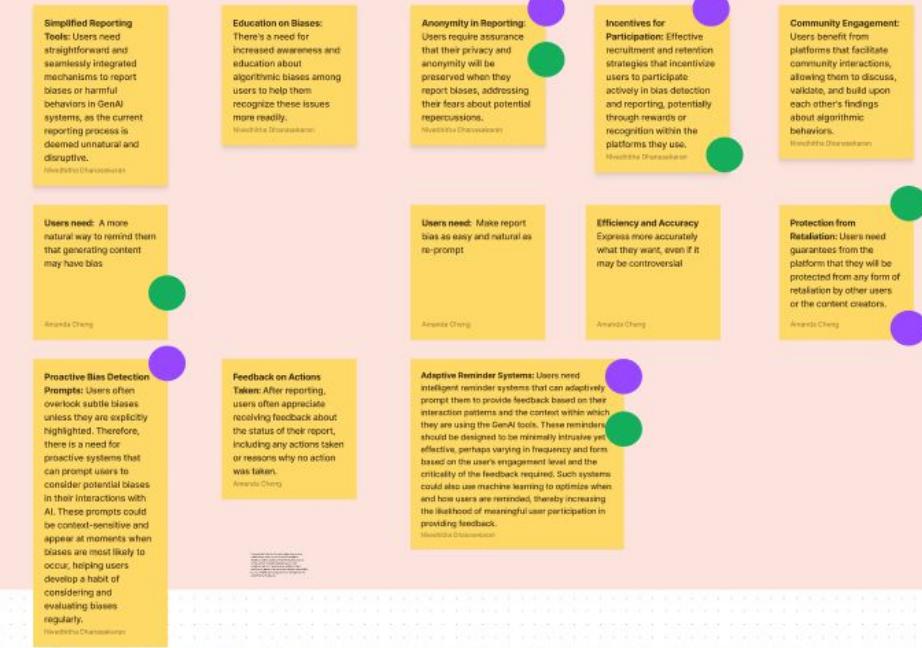
ready to use feedback forms	Simple Q's
	
incentives	influence of peers
 \$\$\$	
Better AI as a result	no resources until done
	
controlled environment	Threaten them
	OMG!

03



Crazy 8s User Needs Prioritization

Crazy 8s User Needs Prioritization & Voting



<https://www.figma.com/file/BGVIKWsaz6MlvXN4bZBwRq/Team-D1-Team-Contract-Spring24?type=whiteboard&node-id=0-1&t=VA3PT9n4LPQK9Jiq-0>

04

Storyboards

User Needs #1 - Amanda's Storyboards

Explores bias boundaries efficiently, generating accurate content that naturally aligns with their needs without extra effort.

Leading Questions:

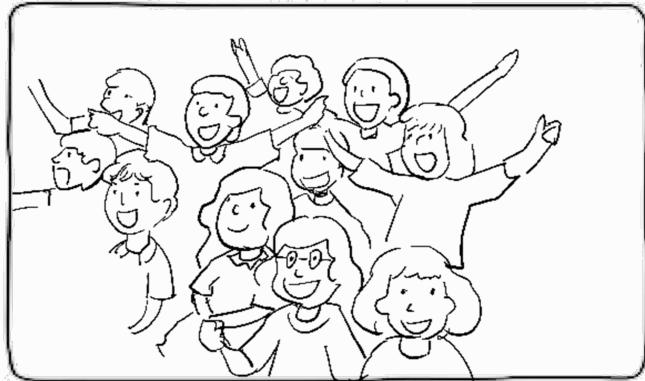
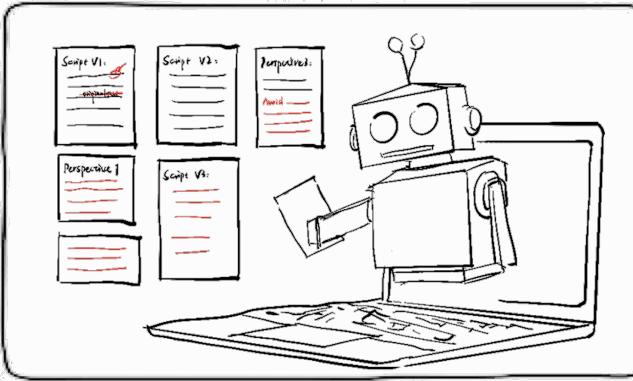
Do you sometimes intentionally push the boundaries of GenAI to generate controversial opinions with it for fun?

Do you usually share this content with people and how?

Would you be willing to try this approach if it could audit potential AI biases and actually make a difference?

Scenario #1 - A [Storyboard Description]

Risky Level



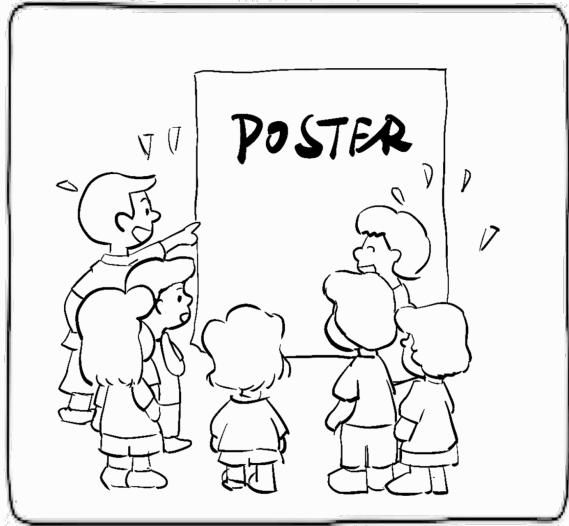
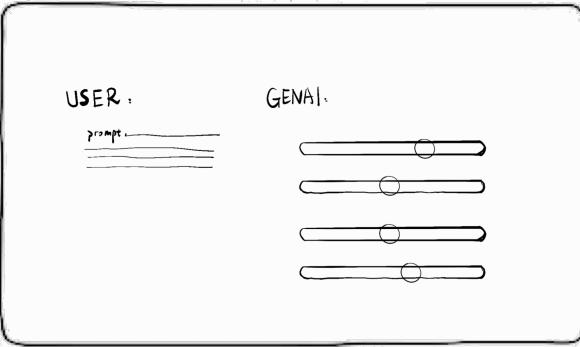
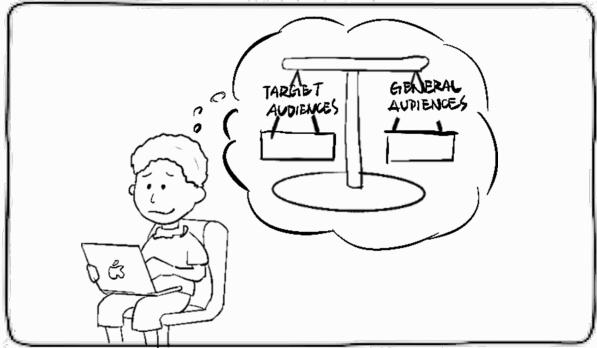
An activist utilizes generative AI to refine and conduct a word check on the script for a public speech. This person was concerned that the wording of the original draft wanted convincing neutrality,

The user ask GenAI to perform a "perspective expansion," supplying him /her with a broad range of viewpoints to enable him to craft the script as neutrally as possible

The Ai not only providing the wording correction but also multiple trending perspective and merge that into the script for perfection

Scenario #1 - B [Storyboard Description]

Risky Level



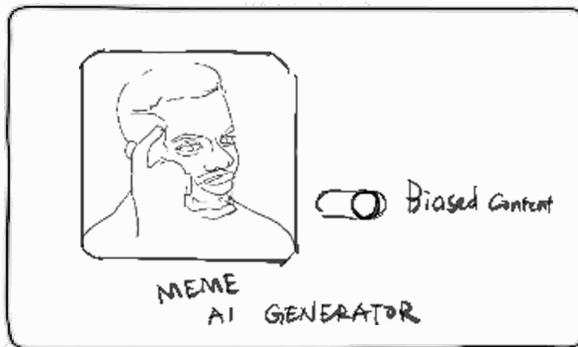
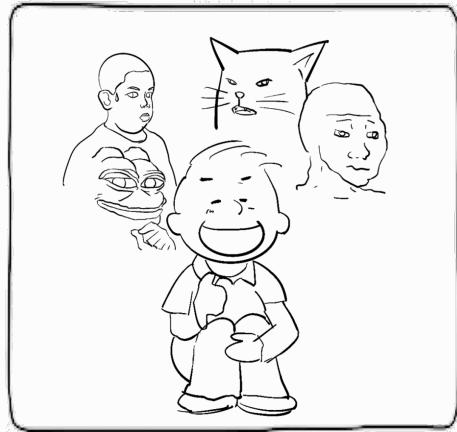
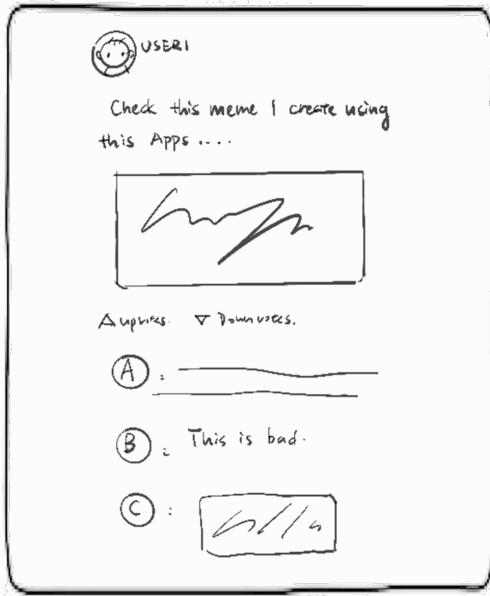
Users are using generative AI to generate posters and posts for a public campaign for their organization, and they want the AI-generated content to appeal to the target audience without underegonize other groups

The AI tool comes with customizable settings that allow users to define the target audience or specific stakeholders, as well as to fine-tune the focus intensity of the generated content. Should the initial outputs not meet expectations, users have the flexibility to modify these levels according to their requirements.

The materials receive highly positive feedback. The blog post initiates a lively debate.

Scenario #1 - C [Storyboard Description]

Risky Level



An activist utilizes generative AI to refine and conduct a word check on the script for a public speech. This person was concerned that the wording of the original draft wanted convincing neutrality,

The user ask GenAI to perform a "perspective expansion," supplying him /her with a broad range of viewpoints to enable him to craft the script as neutrally as possible

The Ai not only providing the wording correction but also multiple trending perspective and merge that into the script for perfection

User Needs #2 - Nive's Storyboards

Motivating Reminder strategy to help users spend more time reflecting on the AI response to their prompt for a more natural & low-effort reporting mechanism.

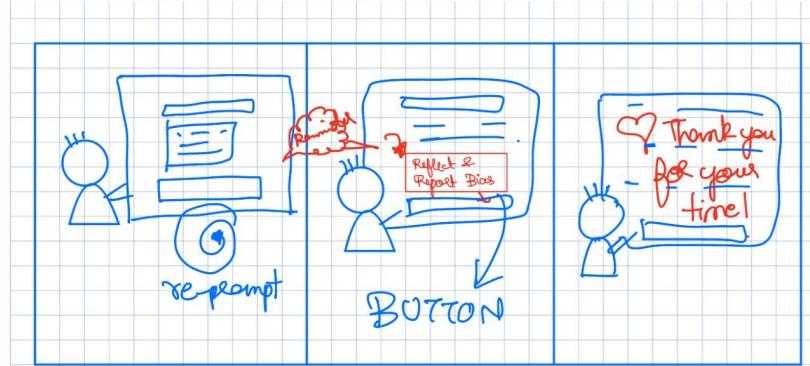
Why is it important? (Researcher's Perspective)

- Current reporting mechanism is unnatural and doesn't fit into the natural workflow of users as they typically resort to re-prompting as an immediate solution to unexpected or unsatisfactory GenAI responses, sometimes even before the generation process is complete by interrupting the flow instead of looking for features to report this behavior.
- The reminder strategy and effort required to provide feedback through UI/UX elements on different GenAI tools determines likelihood of getting feedback from the users.
- Lastly, this reminder strategy alone will not motivate users to actually report biases. The following storyboards were designed to understand what factors would retain user attention and motivate them to complete the bias reporting process even if disrupts their natural flow of interaction with the system.
- Through the Speed Dating sessions we hope to verify our understanding and establish the criticality of this need for users. That is we want to VALIDATE the our PERCEPTION of user needs with those that are ACTUALLY ESSENTIAL to the users!

Safe (Nive #1): Feedback after Every Response

- **User Need:** Seamless integration of a reminder mechanism that prompts users to reflect on the generated response to enable user bias reporting in AI interactions naturally.
- **Lead Question:** Do you spend some time to reflect and assess the response of the GenAI tool to your prompt?
- **Discussion Points:**
 - Would an automatically appearing reminder callout to check for biases after each interaction be noticeable and not intrusive?
 - Would you like some information in the reminder that gives you directions on what types of biases to look for?
 - Would you simply silence the reminder or are you more likely to report biases now?

Safe



Nasty is using an AI tool to generate images of doctors and write a brief description of their duties. He keeps re-prompts the system to get "better" results w/o spending too much time on the output.

He receives a reminder to take time to reflect on the AI response & report any biases encouted.

He finishes the report & receives a "thank you" note from the system.

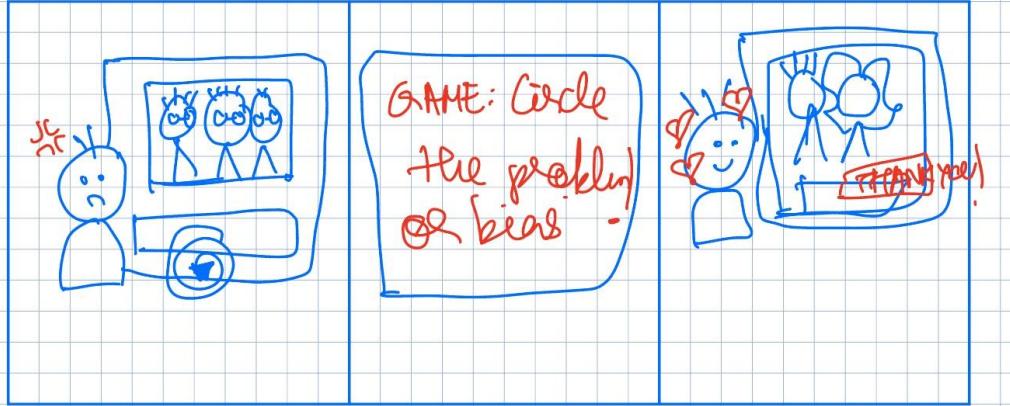
Risky Level



Slightly Risky (Nive #2): Gamified Bias Tag Game!

- **User Need:** Encourage deeper engagement with AI response analysis and bias reporting.
- **Lead Question:** What would motivate you to critically analyze AI responses and report potential biases?
- **Discussion Points:**
 - Would interaction with the system be better than filling out a form?
 - Are you more likely to engage with a game-style reporting mechanism?
 - Would a leaderboard or monetary benefit motivate you to spend more time and expend more energy in analyzing the generated responses for potential biases?

Slightly Risky



Marty is unhappy about the generated images & wants better results so he keeps re-prompting

A pop-up sheets out after some time & asks him to find the problem/bias in the response.

He gets better results & a thank you for reporting the issue.

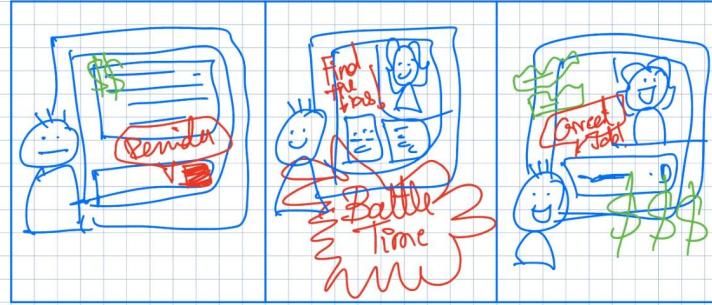
Risky Level



Scenario 3: Monetary Compensation Interactions

- **User Need:** Transform bias reporting into a highly engaging social and monetary activity.
- **Lead Question:** Would monetary compensation activities with daily reminders motivate you to reflect on responses to report on biases?
- **Discussion Points:**
 - What if bias reporting involved interacting with a public forum for tagging biased responses?
 - If you were compensated for your efforts, would you expend time and energy to report biases?
 - Would you like to schedule time to do this daily with reminder from the system?

Very Out-there!



Marty is unhappy that he has to pay to use Neostar & CheftGTT! He gets a reminder about the new report to make money faster!

He engages in a series of online interactions on a public forum to identify biases on current user ops.

At the end of the series, he receives compensation for his work!

Risky Level



User Needs #3 - Alec's Storyboards

Explore methods for AI to provide objective answers in order to support users in everyday tasks.

Leading Question:

Have you encountered any difficulties with not getting what you want from generative AI software (like ChatGPT)?

Discussion Points:

What kinds of methods do you do now, and how effective are those?

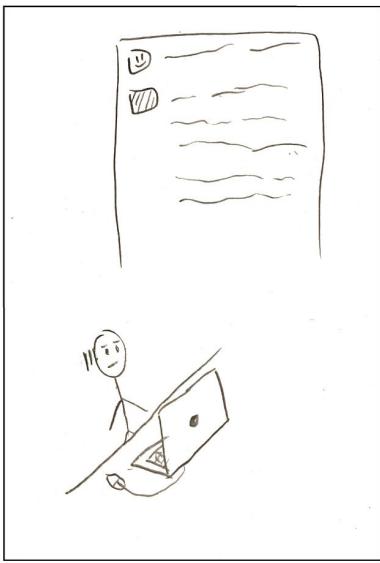
Would you be willing to pursue new methods of improving your results?

What exactly would you want from a new, alternative solution, that your current methods might not be able to offer?

Does this scenario (and solution) look reasonable to you? How likely do you think you would encounter this same situation in your own life?

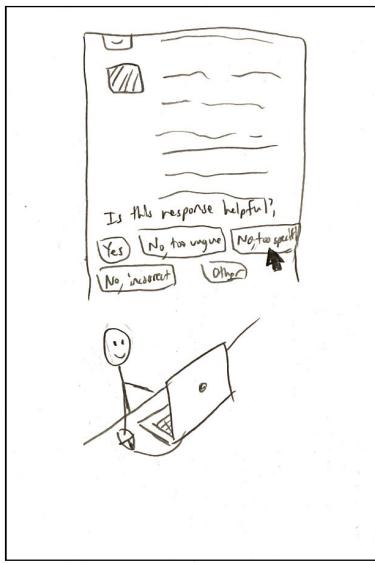
Scenario #1 (Safe) – Post-Generation Satisfaction Queries

Risky Level



CONTEXTS

As Bob tries to use ChatGPT for his background research, he finds that the result he's getting today aren't exactly satisfactory—they're just too specific for his needs right now.



SOLUTIONS

Fortunately, ChatGPT automatically generates queries for users satisfaction after every prompt input. Bob can let the AI know his current misgivings by selecting the appropriate button.

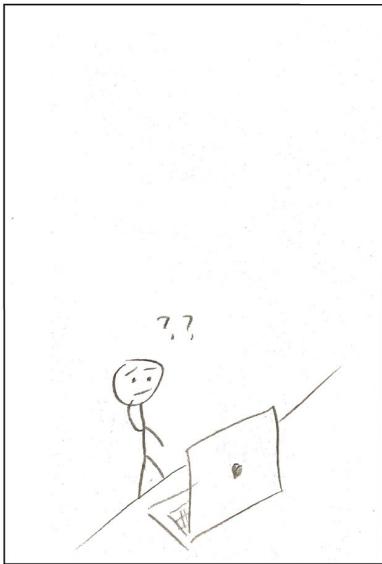


OUTCOMES

After Bob gives his feedback, the AI generates a new result based on it, and takes this into consideration for the future. Bob now receives better results—in both the present and future.

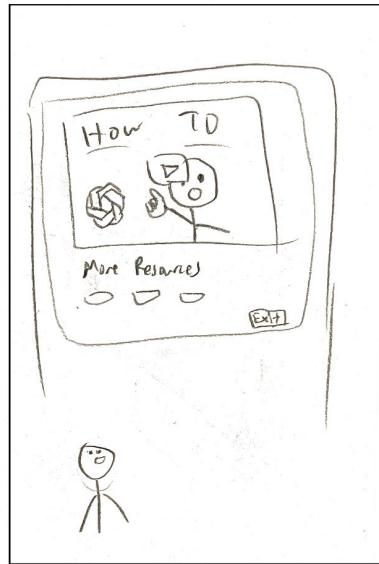
Scenario #2 (Risky) - “How-To” Informational Videos

Risky Level



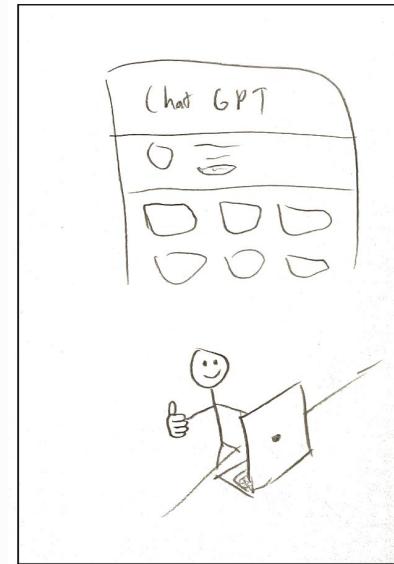
CONTEXTS

John is new to using ChatGPT, and isn't sure where to get started. He's a little intimidated by how generative AI can give you radically different answers based on small differences in prompts.



SOLUTIONS

Fortunately for John, as a first-time user, the website automatically provides him with a user guide detailing the basics of ChatGPT—and additional links should he require them later on.

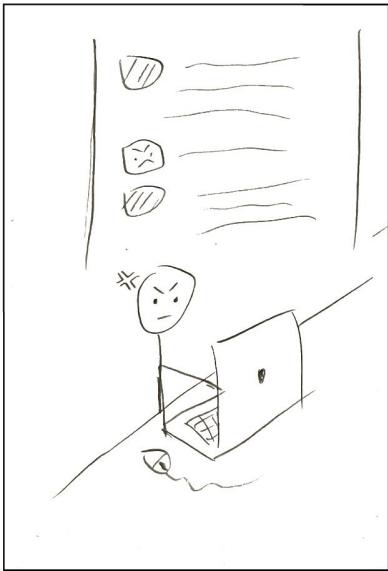


OUTCOMES

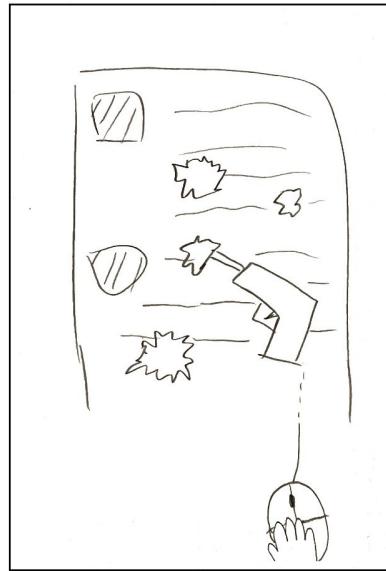
This expedites John's understanding of ChatGPT significantly, and whereas he might've been slow to start before, he's well on his way to using AI to assist his work.

Scenario #3 (Very Risky) - “Kill This Response”

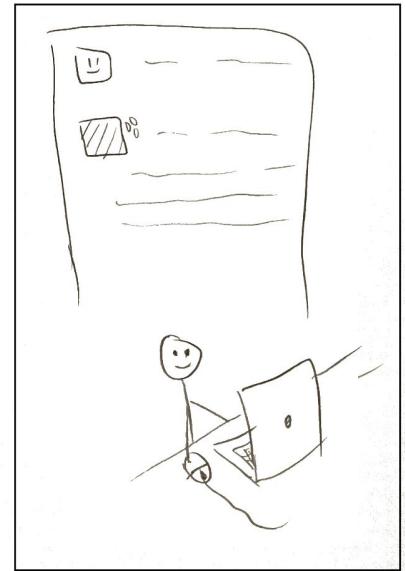
Risky Level



CONTEXTS



SOLUTIONS



OUTCOMES

Sam is getting repeatedly frustrated with ChatGPT today. Even as he's tried everything, nothing the AI provides is in line with what he wants. He's getting ready to throw his laptop out of anger.

Trashing a computer is expensive—but ChatGPT’s “Kill This Response” minigame isn’t. Sam can relieve some stress by destroying bad results, and the AI will avoid results like those later.

Having both curated ChatGPT’s outputs for the future and improved his mood with one short game, Sam is ready to focus up and get back to work.

User Needs #4 - Sofia's Storyboards

Users need to understand the importance of recognizing and reporting AI biases, including engaging in dialogues with dedicated communities and platforms.

Leading Questions:

How would you prefer to report biases in AI-generated content in a way that feels safe and respected?

How would you feel about engaging in a community dialogue to address biases in AI content?

How would you react if an AI threatened to delete your computer's data unless you completed a feedback survey on its biases?

Scenario #1 (Safe) - Easy to Access Button

Risky Level



A user scrolling through their social media feed and noticing a biased article recommended by the platform's AI.

The platform provides a simple "Report Bias" button next to each recommendation.

After reporting, the user receives a thank you message, and the feedback is used to improve AI recommendations.

Scenario #2 (Progressively Riskiers) – Discussing in a Forum

Risky Level



A user finds a biased recommendation and sees an option to "Discuss with Community."

Clicking the option opens a moderated forum where users discuss biases in recommendations.

The discussion leads to community-driven recommendations for AI improvement.

Scenario #3 (Out There) – Threatening the User

Risky Level



PDF Viewer



A user encounters a pop-up from their AI assistant warning that their computer's data will be progressively deleted unless they complete a feedback survey on AI biases within 24 hours.

The user initially thinks it's a joke but soon realizes the AI begins to 'simulate' data deletion by hiding files temporarily, increasing the urgency to comply.



After completing the survey, the AI reveals it was a bluff designed to underscore the importance of user feedback on biases. All 'deleted' data is restored, leaving the user relieved but deeply contemplative about the lengths needed to ensure AI ethics and bias correction.

05

Speed Dating

- Overall Findings & Insights
- Common misunderstandings
- Needs & designs validated by users
- New design opportunities revealed by users

The Process of Speed Dating

(Alec)

The Process

- Each member conducted at least one participant
- Participants came from varied backgrounds, but all had some experience with using AI systems (ChatGPT, etc.)
- Studies were conducted in real-time, either in-person or remotely
- Storyboards were presented in a randomized order for each participant

General Findings and Insights

(Nive, Alec and Amanda)

- **Natural Reporting Mechanisms:**
 - There is a clear need for more natural, less intrusive reporting mechanisms within GenAI systems to encourage consistent user feedback.
- **Educational Enhancement:**
 - Users generally lack awareness and the necessary skills to identify biases in GenAI outputs, indicating the need for targeted educational efforts.
- **Privacy and Anonymity Concerns:**
 - Privacy and anonymity are major concerns for users when reporting biases, affecting their willingness to engage in such activities.
- **Emotional Concerns:**
 - The emotional impact of biases on users is typically low unless the biases are overt, which affects the frequency and urgency of reporting.
 - Users concerned that the constant concern about bias can be demotivating, making the task of content creation feel laborious and fraught with potential pitfalls.

General Findings and Insights

(Nive, Alec and Amanda)

- **Design Features:**
 - The design of UI/UX significantly influences how and when users provide feedback, with more intuitive designs potentially leading to increased engagement.
- **Concerns Over Accuracy and Fairness:** Users are often concerned about the accuracy and fairness of AI-generated content. They seek assurances that the AI does not perpetuate stereotypes or propagate misleading information, especially in sensitive areas like news, education, and social issues.

General Findings and Insights (Nive)

1. **Natural Reporting Mechanisms:** There is a clear need for more natural, less intrusive reporting mechanisms within GenAI systems to encourage consistent user feedback.
2. **Educational Enhancement:** Users generally lack awareness and the necessary skills to identify biases in GenAI outputs, indicating the need for targeted educational efforts.
3. **Privacy and Anonymity Concerns:** Privacy and anonymity are major concerns for users when reporting biases, affecting their willingness to engage in such activities.
4. **Emotional Connection to Biases:** The emotional impact of biases on users is typically low unless the biases are overt, which affects the frequency and urgency of reporting.
5. **UI/UX Design for Feedback:** The design of UI/UX significantly influences how and when users provide feedback, with more intuitive designs potentially leading to increased engagement.

Common Misunderstandings

(Alec)

1. While building off of the initial prompt we were given at the beginning of this course, we tried our best to build features under the assumption that **users would be intrinsically motivated to report biases**. This was decidedly not the case—many of our interviewees only used AI for the purpose of **quickly getting objective information and nothing else**.
2. Building off of the above point, convenience was a key factor for the users that we interviewed. **Users would not interact with a feature that required them to do extra work**—it would be difficult to for them to engage with something that they believed would disrupt their usual flow.
3. We previously believed that users would be eager to report biases under the assumption that it would be **the primary method of improving the quality of their prompt outputs**. This was not the case, as many users simply ran multiple similar prompts if they did not get what they wanted. One user even claimed that it would be the user's fault if the system produced unsatisfactory results.

What were your common misundersta- ndings? (Nive)

1. **Bias Perception:** Users often do not perceive subtle biases unless they are explicitly pointed out or affect them directly, leading to underreporting.
2. **Impact of Reporting Mechanisms:** There is a misunderstanding about how cumbersome or interruptive reporting mechanisms discourage feedback, with designers possibly assuming that users will report issues regardless of the difficulty.
3. **Privacy Expectations:** There is a gap in current reporting mechanism where companies do not anonymize user information when reporting biases, which is not the case, as many are hesitant due to anonymity concerns.
4. **User Engagement Strategies:** There is a misconception that users will naturally engage with bias detection without additional incentives or gamification elements.
5. **Feedback Fatigue:** Developers may underestimate the fatigue users experience when asked to repeatedly engage with complex reporting tools or processes

Validated Needs/ Designs

(Alec)

1. Users want features to be **convenient** and unintrusive to their workflow. They were more than willing to try out novel features, but were unlikely to use them if said features were too out-of-the-way.
2. When reporting controversial content, users have specific needs related to **maintaining their privacy and security**, particularly when there's a risk that their report might be made public.
3. Users generally agreed that **community involvement/discussions** were effective and proven methods of improving the development of AI, leading to better outputs.

What needs/design ideas did users validate? (Nive)

1. **Simplified Reporting Tools:** Users validated the need for reporting mechanisms to be streamlined and integrated seamlessly into their regular interactions with GenAI systems.
2. **Incentives for Reporting:** Users supported the idea that incentives, whether social recognition or material rewards, could motivate more consistent engagement with bias reporting.
3. **Community Platforms for Shared Learning:** There wasn't a clear validation for the creation of community platforms where users can collaboratively discuss, identify, and report biases.
4. **Multi-modal Feedback Options:** Users affirmed the value of having multiple modes of feedback (e.g., text, voice) to suit different contexts and personal preferences.
5. **Regular Education and Reminders:** There is a validated need for ongoing education about biases and regular reminders to check for biases during interactions with AI systems.

Potential Design Opportunities (Nive)

Based on our user research so far and collected evidence, we think the following design approaches would be good avenues for developing more narrow and specific solutions to address the project prompt.

Each of these strategies aims to increase user engagement and participation in bias reporting by making the process more rewarding, interactive, and seamlessly integrated into everyday use of GenAI tools.

By carefully implementing these approaches, the system we hope to harness the collective vigilance of its user base to enhance overall accuracy and fairness.

1. **Token Collection Strategy:** Earn tokens for bias reports valued by usefulness (1-5 tokens); redeem 200 tokens for 15 free GPT 4.0 prompts, or 500 for a week-long GPT 4.0 upgrade.
2. **Callout Reminder Strategy:** Utilize a pop-up tool for users to critically evaluate AI responses, enabling tagging of biases in images or highlighting in text.
3. **Monetary Reward for Bias Reporting:** Offer financial incentives for routine bias reporting or during major incidents, with a simple button or hashtag for direct reporting to social media.
4. **Recaptcha-Style Engagement Checks:** Implement intermittent, non-intrusive prompts to verify user engagement and foster ongoing attention to detail

Token Collection Strategy:

Develop a token-based incentive system where users earn a form of digital currency—tokens—for reporting biases in GenAI outputs. The token allocation could range from 1 to 5 tokens per report, depending on the report's usefulness as evaluated by the developer team or through automated relevance measures. This system would not only motivate users to participate but also deepen their engagement with the platform. For example, after accumulating 200 tokens, a user might receive 15 free prompts on GPT-4.0. Collecting 500 tokens could unlock a week-long upgrade to GPT-4.0, incentivizing continuous and thoughtful participation.

Callout Reminder Strategy:

Implement a proactive tool that prompts users to critically evaluate their responses, particularly useful when users repeatedly refine the same prompt. This tool could pop up and allow users to circle or quickly tag biases in image responses. For text-based interactions, the tool could enable users to highlight biased sections of text. This visual tagging aids in simplifying the feedback process and encourages users to actively think about the content generated by GenAI, making bias detection part of their routine interactions with the system.

Monetary Reward System for Bias Reporting:

Introduce a monetary compensation system for users who actively report biases, especially during critical incidents. For example, if a significant bias incident garners media attention—such as an AI system repeatedly generating culturally inappropriate content—users could be motivated to engage more actively. A simplified reporting mechanism, such as a dedicated button or a hashtag-driven reporting feature within the platform or on social media, could facilitate this. This strategy not only encourages widespread user participation during crucial times but also helps gather extensive feedback for post-mortem analysis, enhancing the AI's accuracy and cultural sensitivity.

Recaptcha-Style Engagement Checks:

Incorporate a lightweight, Recaptcha-style verification system that occasionally prompts users to verify they are human by responding to repetitive prompts used in their GenAI interactions. This feature should be balanced carefully to ensure it doesn't become overly frequent, which might risk alienating users. The goal would be to subtly encourage users to review their inputs and outputs, reinforcing attention to detail without significantly disrupting the user experience.

More Design Avenues

(Nive, Amanda)

- **Gamification of Bias Detection:** Introducing game-like elements to make the detection of biases more engaging and less of a chore could be a novel approach to increasing user participation.
- **Personalized Bias Alerts:** Designing systems that can learn from users' past interactions to personalize alerts about potential biases could make these alerts more relevant and prompt quicker actions.
- **Anonymity-Enhancing Features:** Developing features that guarantee anonymity when reporting biases could address privacy concerns and encourage more users to participate.
- **Advanced Educational Tools:** There is an opportunity to develop more sophisticated educational tools that guide users from basic to advanced levels of understanding biases.
- **Community-Driven Validation Mechanisms:** Creating mechanisms within community platforms that allow for peer validation and discussion of reported biases could enhance the accuracy and credibility of the feedback process.

Contributions

- 01 - Walking the Wall
 - Nive
- 02 - Crazy 8's
 - Individual Crazy 8's - All
 - Voting - Nive, Amanda
- 03 - Crazy 8's - User Needs
 - User Need Prioritization & Summary - Nive
- 04 - Storyboards
 - Individual Storyboards: All
 - Presentation theme: Amanda, Alec
- 05 - Speed Dating
 - Findings & Insights - Alec, Nive
 - Misunderstandings - Alec, Nive
 - Need/Designs - Alec, Nive
 - Potential Designs - Nive, Amanda