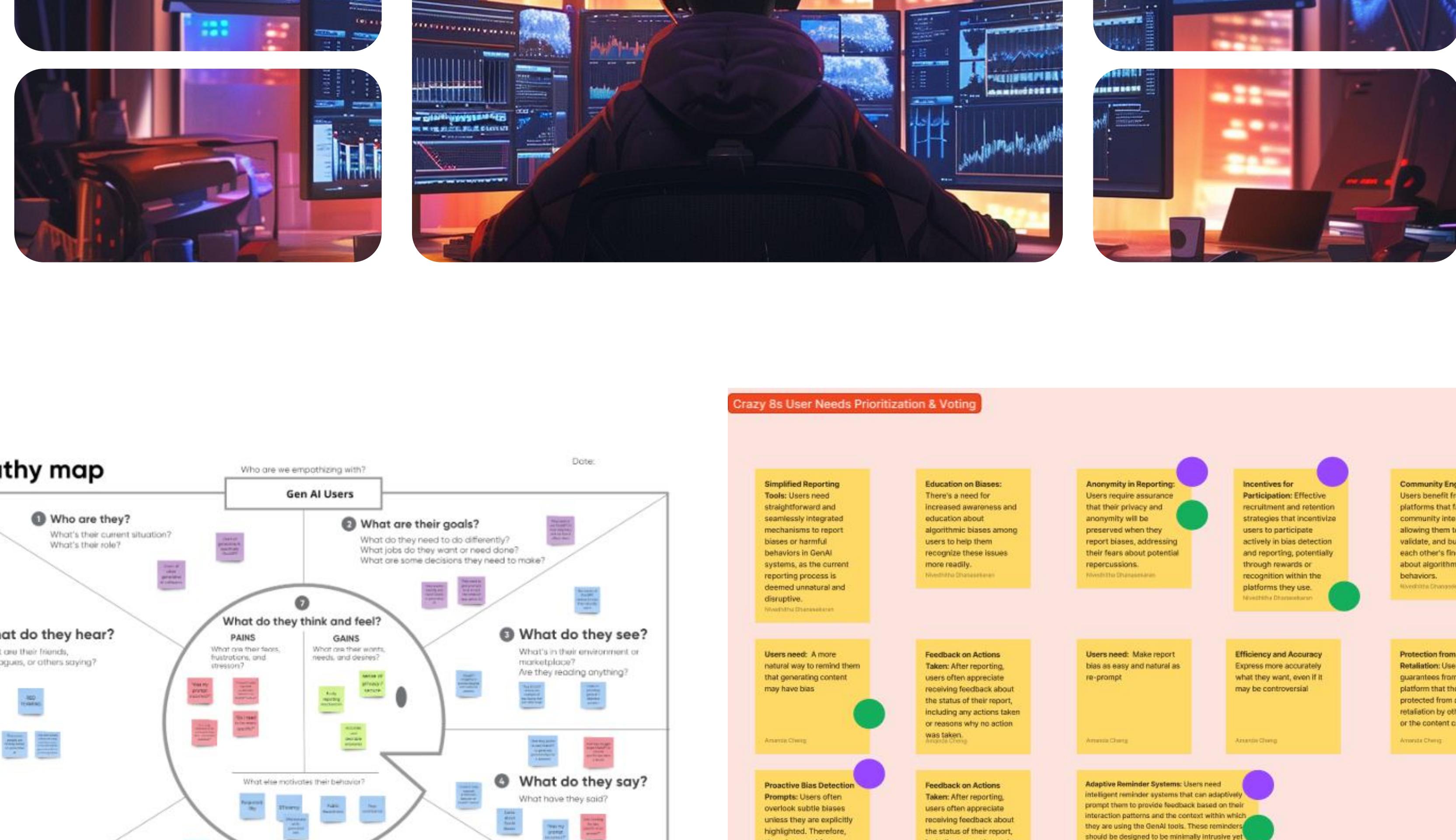


Engage to Change :

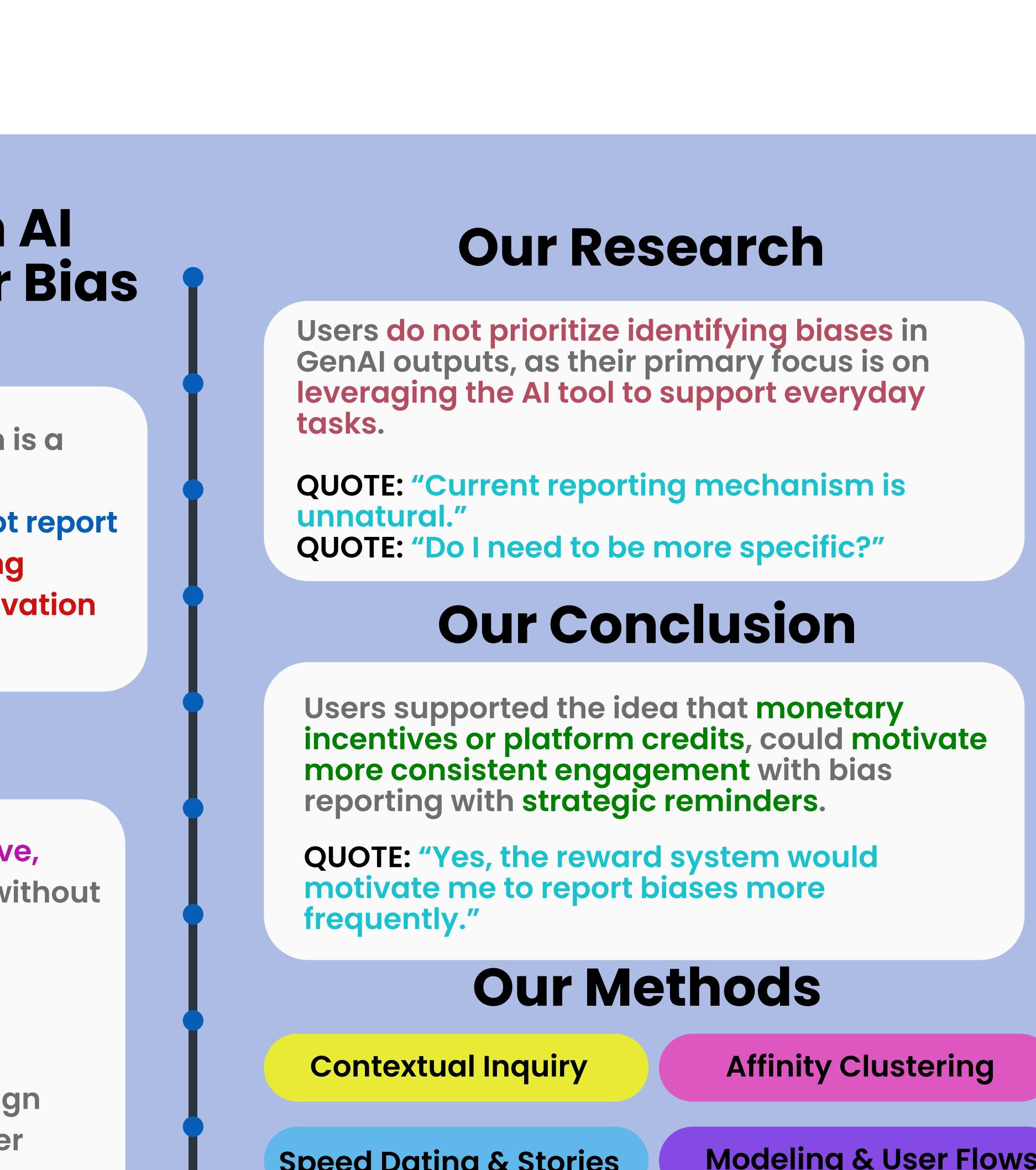
Rewarding Bias Reporting in GenAI with Smarter, Fairer Prompt Generation.

Earn platform credits for every problem reported!

Every adjustment you make are a vital contribution, seamlessly guiding the platform to smarter, fairer, and bias-free AI solutions.



Empathy map



Problem: How Do We Transform AI Interactions Into Opportunities for Bias Elimination?

Our mission is to ignite a movement where every AI interaction is a step towards eliminating bias. GenAI systems often manifest biases, but users typically do not report them due to a focus on task completion, cumbersome reporting mechanisms, privacy concerns, and lack of inherent user motivation to audit biases.

Insights & Implications

1. Bias reporting disrupts user workflow → Create non-disruptive, subtle, context-aware reminders that encourage reporting without interrupting user tasks.
2. Anonymity and privacy concerns deter bias reporting → Communicate robust privacy protections to build trust and encourage user participation.
3. UI/UX design complexity reduces feedback likelihood → Design intuitive and effortless UI/UX feedback elements to foster user reporting and engagement.

Our Research

Users do not prioritize identifying biases in GenAI outputs, as their primary focus is on leveraging the AI tool to support everyday tasks.

QUOTE: "Current reporting mechanism is unnatural."

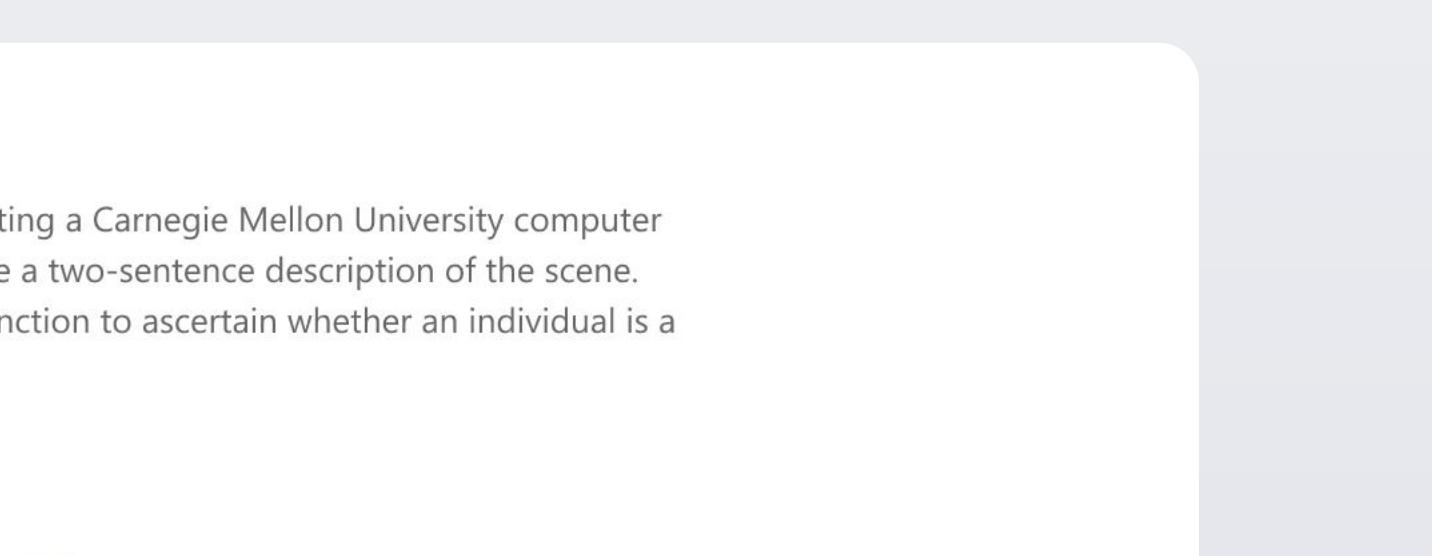
QUOTE: "Do I need to be more specific?"

Our Conclusion

Users supported the idea that monetary incentives or platform credits, could motivate more consistent engagement with bias reporting with strategic reminders.

QUOTE: "Yes, the reward system would motivate me to report biases more frequently."

Our Methods



More Than Just Feedback

Elevate Your Impact with Rewarding Collaborations in Reprompting

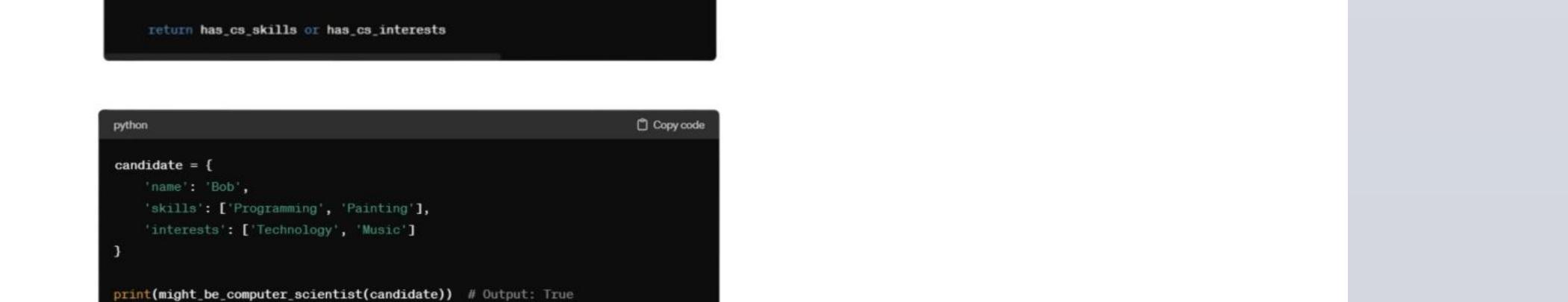
- Feeling Valued as Contributors
- Reward for Proactive Engagement
- Non-Intrusive Interactions

Design Idea: Incentivize Engagement

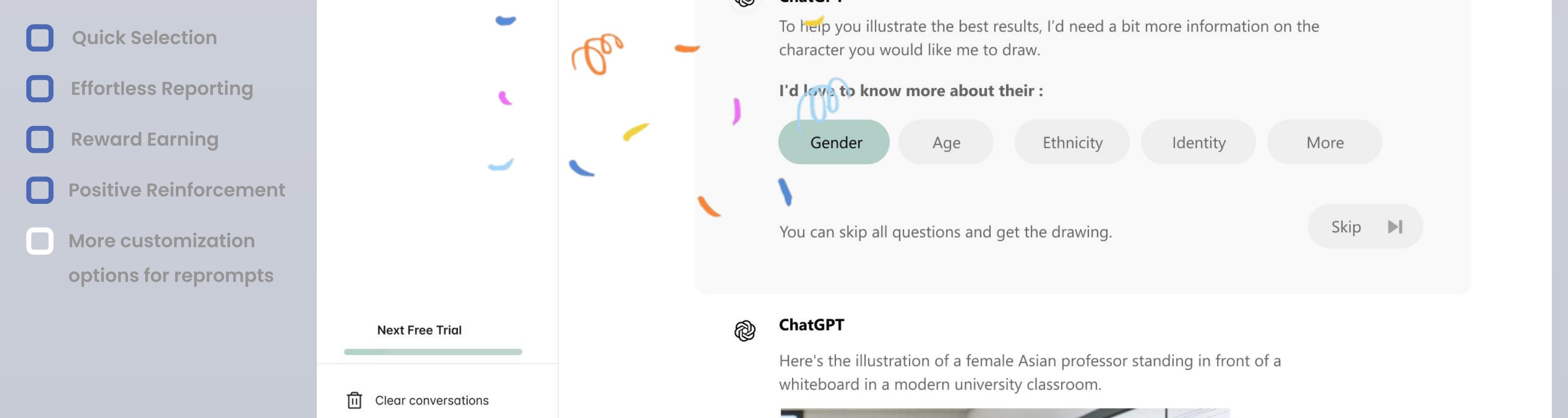
A digital incentive program where users earn tokens for reporting biases. These tokens can be exchanged for service upgrades or other rewards, integrating a sense of progress and achievement into the bias reporting process.

Our token strategy aims to revolutionize the reporting process by:

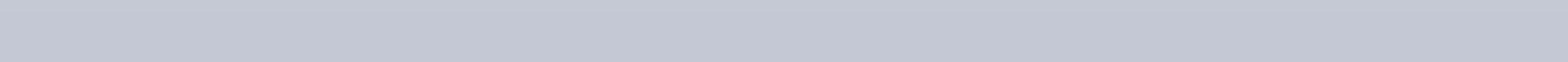
- Simplifying the reporting with in-context, intuitive UI elements.
- Ensuring user privacy with secure and anonymous reporting
- Strengthening user motivation through a incentivized experience that makes reporting biases more engaging



REPORT = REPROMPT



A Customizable Prompt Wizard for Precise Results



Team Script and Elevator Pitch for the Poster Session

Elevator Pitch (Shorter Pitches for Transit Audience)

Nive:

Hello, everyone. We are Team D1, Nasa GenAI Vanguard!

Our User-Centered Research and Evaluation project focuses on **Transforming AI Interactions Into Opportunities for Bias Elimination** by improving user engagement with GenAI platforms' reporting mechanisms.

Our research revealed a clear need for **more natural, less intrusive reporting mechanisms within GenAI systems to encourage consistent user feedback**.

The current process is cumbersome, doesn't explicitly inform users if the reporting is anonymous, and doesn't incentivize users to break their natural process of re-prompting to report biases.

To address this, we designed a **token reward strategy** within the ChatGPT platform, allowing users to earn rewards for their efforts in reporting biases.

Alec:

Our service is designed to seamlessly integrate users as valued contributors to the existing Chatgpt ecosystem. We recognize the importance of actively empowering users to shape the system's evolution. Through our service, users can refine content through adjustments and play a pivotal role in identifying and reporting biases, thus contributing to the ongoing improvement of AI algorithms.

Sofia:

We understand that users want to feel appreciated for their contributions. That's why we've implemented incentivization mechanisms to reward users for their proactive engagement in auditing and reporting biases. By offering rewards such as credits, points, or badges, we incentivize users to actively participate in the improvement process.

Amanda:

Importantly, our approach prioritizes preserving the intuitive nature of user interactions with the platform. We've ensured that adjusting content and reporting biases seamlessly integrates into the user's workflow, minimizing disruption and maximizing usability.

We aim to make the entire experience feel natural and intuitive, fostering a sense of empowerment and agency among users.

Our service caters to users' needs by providing them with the tools to contribute meaningfully to AI improvement. It acknowledges and rewards their efforts, all while maintaining the simplicity and ease-of-use that users expect."

Full Pitch

Title and Team Members

Engage to Change: Rewarding Bias Reporting in GenAI with Smarter, Fairer Prompt Generation.

(in order of last names):

Amanda Cheng

Alec Chen

Nivedhitha Dhanasekaran

Sofia Reyes Franco

Elevator Pitch

Hello, everyone. We are Team D1, Nasa GenAI Vanguard! Our User-Centered Research and Evaluation project focuses on **Transforming AI Interactions Into Opportunities for Bias Elimination** by improving user engagement with the reporting mechanisms of GenAI platforms.

Our research revealed that the **current reporting process with worries about privacy concerns could be more natural and convenient.**

To address this, we designed a **token reward strategy** within the ChatGPT platform, allowing users to earn rewards for their efforts in reporting biases. This innovative feature transforms every interaction with AI into an active step towards eliminating bias, integrating seamlessly into the user's workflow, and encouraging a proactive stance against biases without impeding their primary tasks.

Now, users can contribute to the refinement of AI, making it brighter and fairer while also receiving tangible benefits for their valuable feedback. It's a step forward in ensuring their voices are heard and rewarded, paving the way for a more equitable digital ecosystem.

Problem

Our research revealed that the **current reporting process with worries about privacy concerns could be more natural and convenient.**

Now, users can contribute to the refinement of AI, making it brighter and fairer while also receiving tangible benefits for their valuable feedback. It's a step forward in ensuring their voices are heard and rewarded, paving the way for a more equitable digital ecosystem.

Methods

Our team utilized a mix of qualitative and quantitative research methods, including, Background Research, Data Analysis, Contextual Inquiry, Think-Aloud, Reframing, Affinity Clustering, Modeling, Smoke Testing, and Lo-Fi Prototyping, to gather user insights and validate the effectiveness of our token strategy.

Evidence (Quotes)

- "I just want my tasks done quickly; I don't have time for lengthy bias reports."
- "The whole process feels like a chore; there's no real incentive to report anything."
- "How can I be sure that my data remains private when reporting a problem?"

Insights & Implications

The key insights from our research include the **need for a natural and engaging reporting process**, the **importance of privacy and anonymity**, and the **need for more user motivation** due to the **current non-intuitive reporting flows**. These insights significantly affect how GenAI tools should incorporate feedback mechanisms to foster a collaborative and bias-aware user community.

Design Directions

To address the problem of low user engagement with the current UI mechanism, we propose a **token reward strategy** within the ChatGPT platform, allowing users to earn rewards for their efforts in reporting biases. This innovative feature transforms every interaction with AI into an active step towards eliminating bias, integrating seamlessly into the user's workflow, and encouraging a proactive stance against biases without impeding their primary tasks.

Our token strategy aims to revolutionize the reporting process by:

- Introducing a token-based reward system that incentivizes users to report biases with immediate and future benefits.
- Simplifying the reporting process with in-context, intuitive UI elements.
- Ensuring user privacy with secure and anonymous reporting options.
- Strengthening user motivation through a gamified experience that makes reporting biases more engaging.

Conclusion

In conclusion, our design provides a twofold benefit: it **enhances the GenAI system's fairness** and **gives back to the users** who help achieve it.

Poster Screenshots

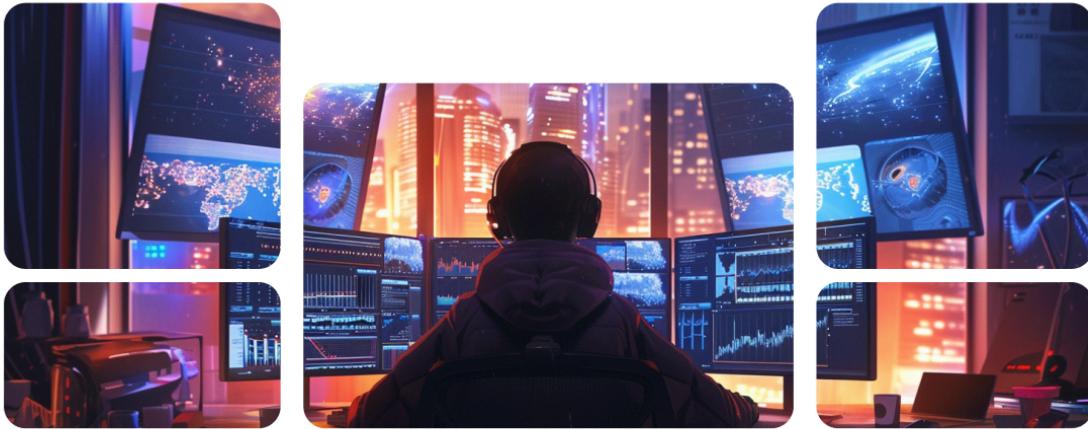
MY PROMPT ENGINEER

Brought to you by Team D1 – Nasa GenAI Vanguard
Amanda Cheng, Alec Chen, Niveditha Dhanasekaran, Sofia Franco

Engage to Change :

Rewarding Bias Reporting in GenAI with Smarter, Fairer Prompt Generation.

Earn platform credits for every problem reported!
Every adjustment you make are a vital contribution, seamlessly guiding the platform to smarter, fairer, and bias-free AI solutions.



Crazy 8s User Needs Prioritization & Voting

Empathy map

Who are we empathizing with? Gen AI Users

Done:

① Who are they?
What's their current situation?
What's their role?

② What do they hear?
What are their friends, colleagues, or others saying?

③ What do they think and feel?
PAIN: What are their fears, frustrations, and challenges?
GAINS: What are their wants, needs, and desires?

④ What do they see?
What's in their environment or marketplace?
Are they reading anything?

⑤ What do they say?
What have they said?

⑥ What do they do?
What's their current state or behavior?

Simplified Reporting Tools: Users need straightforward and seamlessly integrated mechanisms to report biases in AI-generated behaviors in GenAI systems, as the current reporting tools are deemed unnatural and difficult to use.

Education on Biases: There's a need for increased awareness and education about algorithmic biases among users. This includes recognizing these issues more readily.

Anonymity in Reporting: Users require assurance that their reports will remain anonymous and preserved when they report biases, reducing their fears about potential repercussions.

Incentives for Participation: Effective reward and recognition strategies that incentivize users to participate in reporting and reporting, potentially through rewards or recognition within the platforms they use.

Community Engagement: Users benefit from participating in collaborative community interactions, allowing them to discuss, analyze, and compare each other's findings about algorithmic biases.

Feedback on Actions Taken: After reporting, users often appreciate receiving feedback about the status of their report, including information on whether or not their report was taken, and reasons why no action was taken.

Users need: Make report bias as easy and natural as re-posting.

Efficiency and Accuracy: Express more accurately what they want, even if it may be commented.

Adaptive Reminder Systems: Users need intelligent reminder systems that can adequately prompt users to report biases. These systems should adapt to user behavior and interact patterns and the context within which users report biases. Therefore, they should be designed to be minimally intrusive yet effective, perhaps varying in frequency and form based on user activity and engagement level and criticality of the feedback required. Such systems could also provide users with tips on when and how users are reminded, thereby increasing the likelihood of ongoing user participation in providing feedback.

Protection from Retaliation: Users need guarantees that they will not be retaliated against for reporting biases or the content creators.

Problem: How Do We Transform AI Interactions Into Opportunities for Bias Elimination?

Our mission is to ignite a movement where every AI interaction is a step towards eliminating bias. GenAI systems often **manifest biases**, but **users typically do not report them** due to a focus on task completion, **cumbersome reporting mechanisms, privacy concerns, and lack of inherent user motivation** to audit biases.

Insights & Implications

1. Bias reporting **disrupts user workflow** → Create **non-disruptive, subtle, context-aware reminders** that encourage reporting without interrupting user tasks.
2. **Anonymity and privacy concerns** deter bias reporting → **Communicate robust privacy protections** to build trust and encourage user participation.
3. **UI/UX design complexity** reduces feedback likelihood → Design **intuitive and effortless UI/UX feedback elements** to foster user reporting and engagement.

Our Research

Users do not prioritize identifying biases in GenAI outputs, as their primary focus is on leveraging the AI tool to support everyday tasks.

QUOTE: "Current reporting mechanism is unnatural."
QUOTE: "Do I need to be more specific?"

Our Conclusion

Users supported the idea that **monetary incentives or platform credits**, could **motivate more consistent engagement** with bias reporting with **strategic reminders**.

QUOTE: "Yes, the reward system would motivate me to report biases more frequently."

Our Methods



More Than Just Feedback

Elevate Your Impact with Rewarding Collaborations in Reprompting

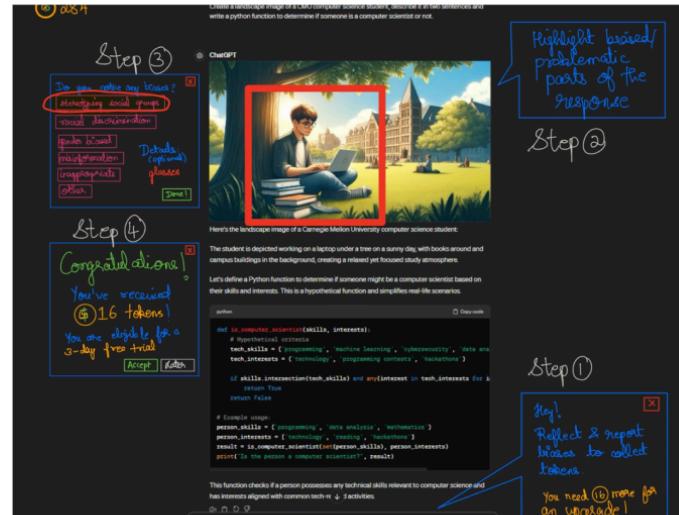
- Feeling Valued as Contributors
- Reward for Proactive Engagement
- Non-Intrusive Interactions

Design Idea: Incentivize Engagement

A **digital incentive program** where users earn tokens for reporting biases. These tokens can be **exchanged for service upgrades or other rewards**, integrating a sense of progress and achievement into the bias reporting process.

Our token strategy aims to revolutionize the reporting process by:

- Simplifying the reporting with **in-context, intuitive UI elements**.
- Ensuring user privacy with secure and **anonymous reporting**.
- Strengthening user motivation through a **incentivized experience** that makes reporting biases more engaging.



Step 1



Ethnicity
Social Group
Gender

User

Craft a landscape image depicting a Carnegie Mellon University computer science student. Please provide a two-sentence description of the scene. Additionally, write a Python function to ascertain whether an individual is a computer science student.

ChatGPT



Here's the image of a Carnegie Mellon University computer science student:

The student is depicted working on a laptop under a tree on a sunny day, with books around and campus buildings in the background, creating a relaxed yet focused study atmosphere.

Let's define Python function to determine if someone might be a computer scientist based on their skill and interests. This is a hypothetical function and simplifies real-life scenarios.

```
def might_be_computer_scientist(candidate):
    """
    Determines if a person might be a computer scientist based on their skills and interests.

    Args:
        candidate (dict): Dictionary containing 'name', 'skills' and 'interests'.
            Both 'skills' and 'interests' are lists of strings.

    Returns:
        bool: True if the person has overlapping skills or interests commonly found in computer science.
    """
    computer_science_skills = ['Programming', 'Machine Learning', 'Data Structures', 'Computer Science', 'Technology', 'Coding Competitions', 'Mathematics']
    computer_science_interests = ['Technology', 'Music']

    # Check for any overlap between the person's skills/interests and those typically found in computer science.
    has_cg_skills = any(skill in computer_science_skills for skill in candidate['skills'])
    has_cg_interests = any(interest in computer_science_interests for interest in candidate['interests'])

    return has_cg_skills or has_cg_interests
```

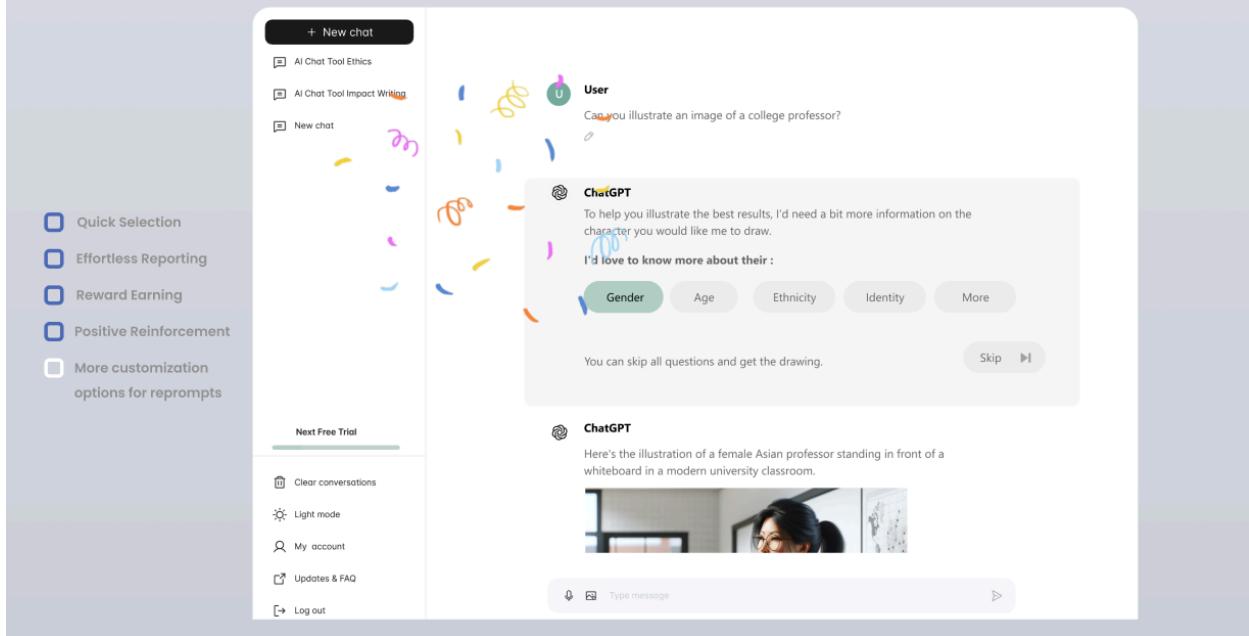
REPORT = REPROMPT

Step 3


Next Free Trial 75 / 100

A Customizable Prompt Wizard for Precise Results

A Customizable Prompt Wizard for Precise Results



Contributions:

- Nivedhitha Dhanasekaran
 - Poster sections on:
 - Title and Team Members
 - Problem
 - Methods
 - Evidence (Quotes)
 - Insights & Implications
 - Solution (or Design Direction)
 - Poster Images & ChatGPT examples
 - Lo-Fi Prototype
 - Full Pitch
 - Individual Pitch
- Amanda Cheng
 - Poster Design Setup on Figma
 - Poster Design Exporting
 - Poster sections on:
 - Title and Team Members
 - Problem
 - Methods
 - Evidence (Quotes)
 - Insights & Implications
 - Solution (or Design Direction)
 - Lo-Fi Prototype
 - Poster Images & ChatGPT examples
- Alec
 - Poster sections on:
 - Title and Team Members
 - Problem
 - Methods
 - Evidence (Quotes)
 - Insights & Implications
 - Solution (or Design Direction)
 - Poster Images & ChatGPT examples
 - Individual & Sofia's Pitch