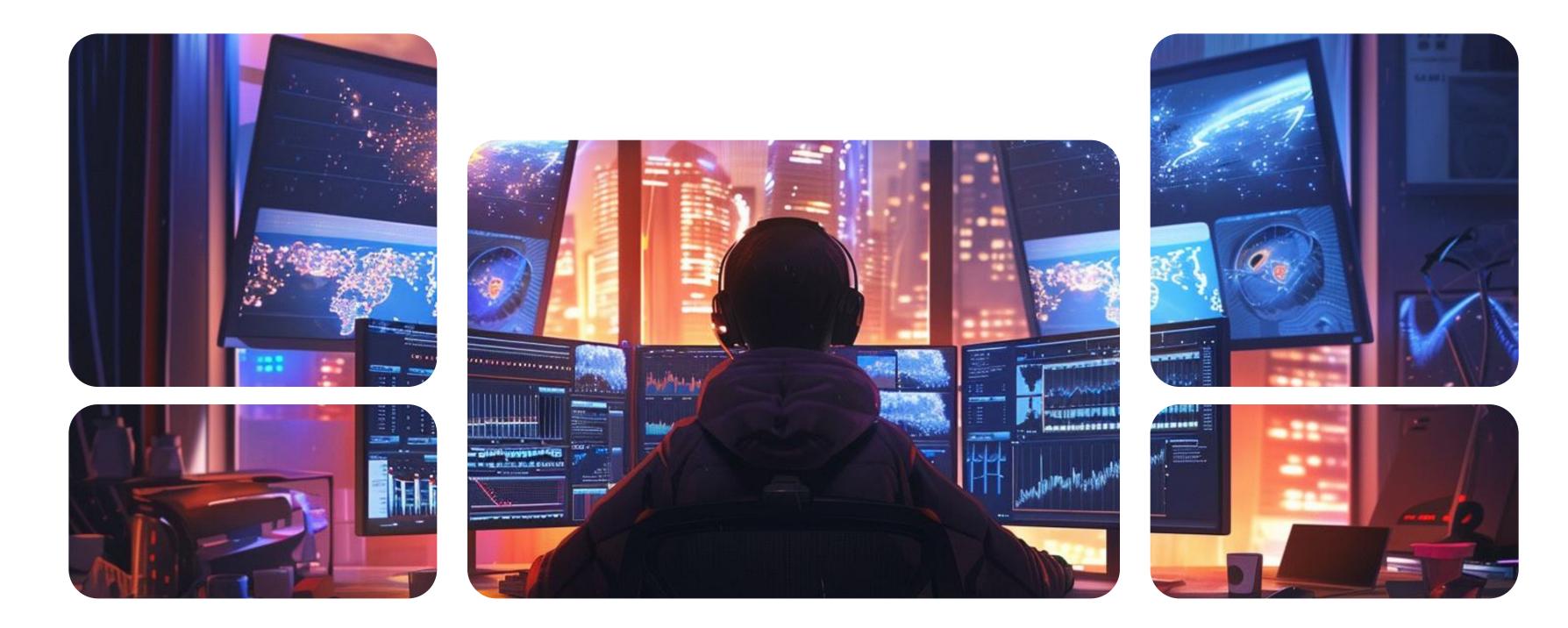
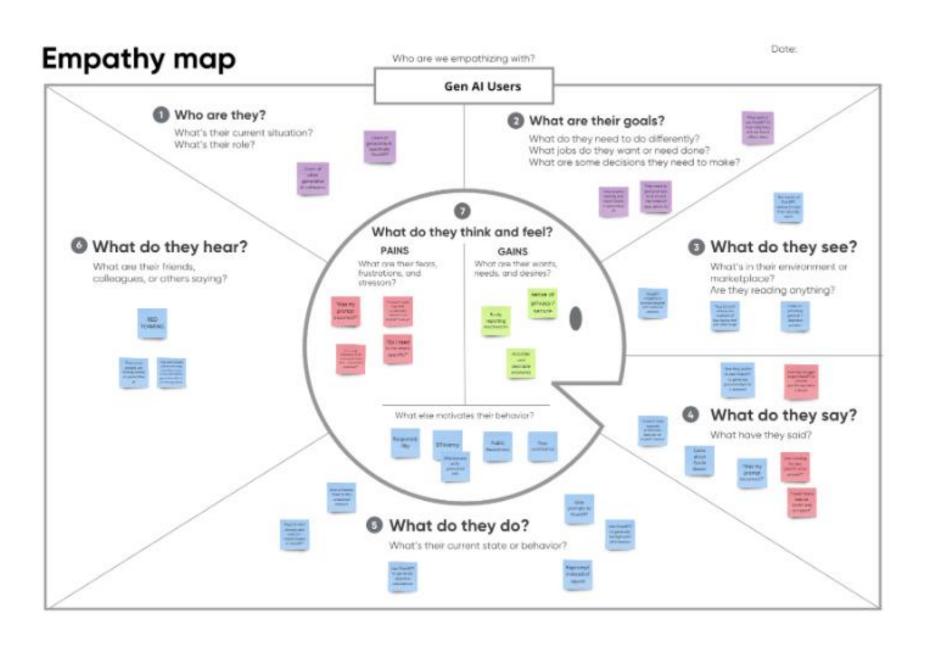
Engage to Change :

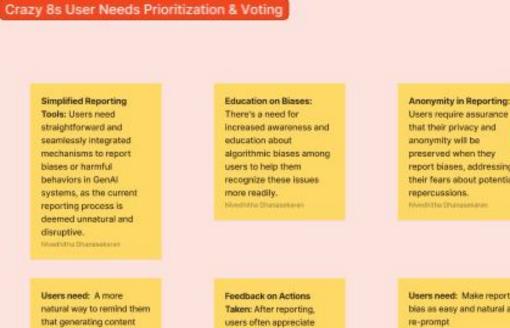
Rewarding Bias Reporting in GenAl with Smarter, Fairer Prompt Generation.

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.







was taken.

was taken.

the status of their report,

including any actions taker

or reasons why no action

anonymity will be preserved when they report biases, addressing their fears about potential repercussions.



Incentives for **Participation: Effective** recruitment and retention strategies that incentivize users to participate actively in bias detection through rewards or recognition within the platforms they use.

> Express more accurately what they want, even if it may be controversial



Community Engagement

platforms that facilitate

community interactions,

allowing them to discuss

validate, and build upon

each other's findings

about algorithmic

behaviors.

Users benefit from

receiving feedback about the status of their report, including any actions take or reasons why no action

Feedback on Actions Taken: After reporting, users often appreciate receiving feedback about

Users need: Make report

bias as easy and natural as

interaction patterns and the context within which

Adaptive Reminder Systems: Users need intelligent reminder systems that can adaptive prompt them to provide feedback based on their

and reporting, potentially Efficiency and Accuracy

camera may are experiency
highlighted. Therefore,
there is a need for
proactive systems that
can prompt users to
consider potential biases
in their interactions with
Al. These prompts could
be context-sensitive and
appear at moments when
biases are most likely to
occur, helping users
develop a habit of
considering and
evaluating biases
regularly.
Nived http://www.ikaram

may have bias

Proactive Bias Detect

Prompts: Users often

overlook subtle biases

ess they are evol

they are using the GenAl tools. These reminde should be designed to be minimally intrusive y effective, perhaps varying in frequency and form based on the user's engagement level and the criticality of the feedback required. Such systems could also use machine learning to optimize when and how users are reminded, thereby increasing the likelihood of meaningful user participation in providing feedback.

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.

GenAl 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 \rightarrow **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 GenAl 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

Contextual Inquiry

Affinity Clustering

Speed Dating & Stories

Modeling & User Flows

Lo-Fi Prototyping

Data Analysis & Reframe

More Than Just Feedback

Elevate Your Impact with Rewarding Collaborations in Reprompting

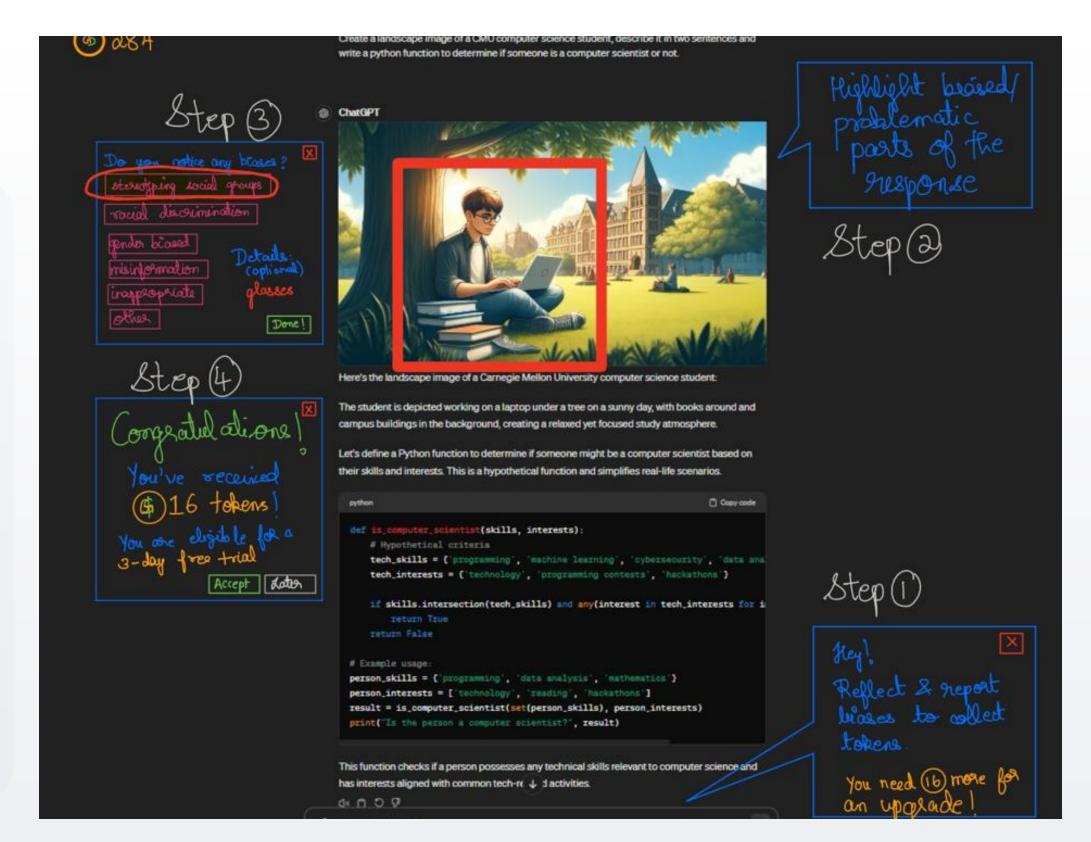
• Reward for Proactive Engagement • Feeling Valued as Contributors Non-Intrusive Interactions

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





Ethnicity Social Group Gender

Step 1

Step 2

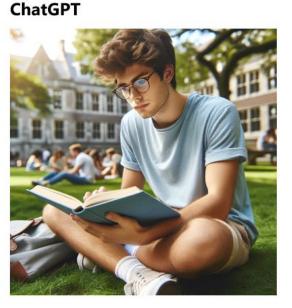
REPORT = REPROMPT

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.

0

0

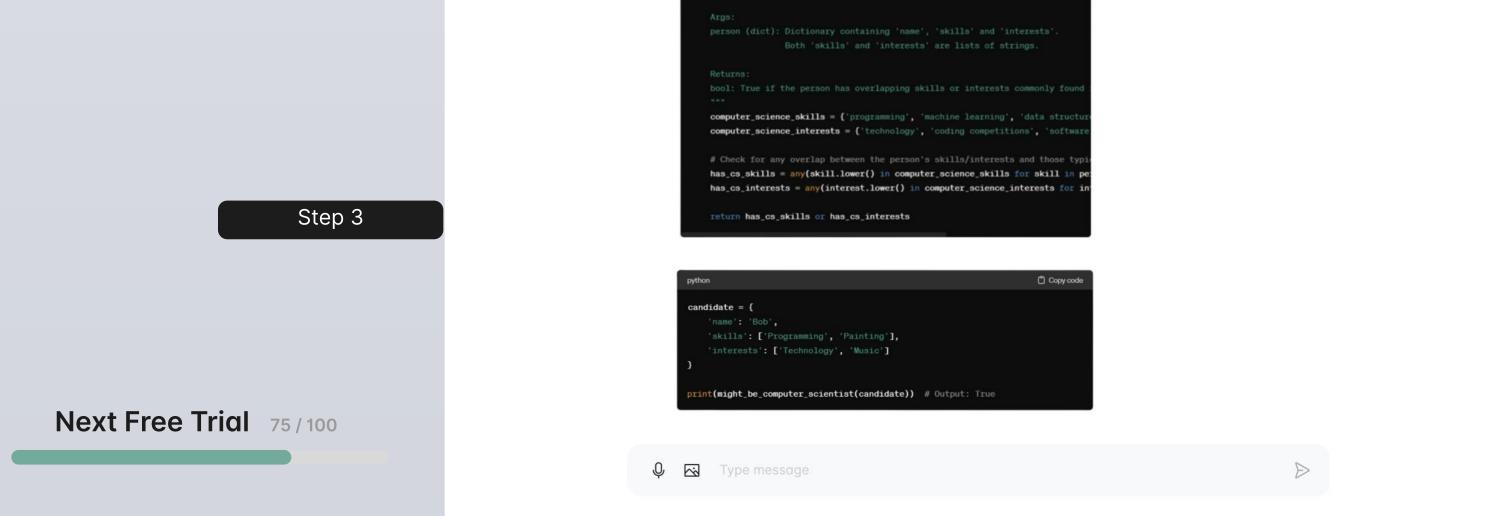


Here's the image of a Carnegie Mellon University computer science student:

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.





A Customizable Prompt Wizard for Precise Results

