# AUDITBOT

*A chat-based approach to auditing bias on social media*

**Team Be Free:** John Jongyeon Chae • Priya Jain • Lisa Leung • Yumi Sato • Martina Tan

## PROBLEM

How might we improve end-users' confidence and motivation to act on incidences of AI algorithmic bias?

## METHODS

- Semi-Structured Interviews
- Think Aloud Studies
- Contextual Inquiry
- Affinity Diagramming
- Stakeholder/Empathy Maps
- Speed Dating

## INSIGHTS

### Anonymity

> There's a guise of anonymity that comes with being on the Internet ... I might talk about something I normally wouldn't talk about because there is less social stigma online.

### Simplicity

> I don't know if there's an option for 'report this ad' ... I would just 'thumbs it down' because I don't know how else to go about reporting it.

### Transparency

> Sometimes, it's just like 'ok so the report has been made, and someone has looked at it.' ... I want to know that work is being done on that front and that we're moving along in the process of review.

## USER NEEDS

### Anonymity

People are actively aware of their actions and how it will be perceived by others. We found ***users feel more comfortable speaking their mind anonymously*** because they don't have to worry about their identity being exposed or negatively perceived.

### Simplicity

***Users prefer easily discovered and traversed reporting structures upon viewing harmful content.*** Without simple reporting structures, they are unable to take action in time, if at all, and communicate their needs to the platform.

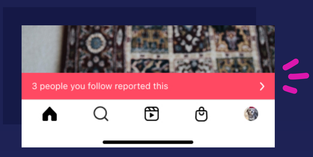
### Transparency

***Users perceive that they will not be heard unless many others are saying the same thing.*** Users feel that companies prioritize overarching business goals over individuals' feelings, and users feel insufficiently informed about what would happen after they report content.

## OUR SOLUTION

### Low-Pressure Social Proof

See an anonymized indication of reports made by personal connections.




### Tap to Opt In on any Post

Click on the AuditBot icon under the post to initiate a conversation at your pace.

### All Your Auditing in One Place

AuditBot opens in Direct Messages and guides users to discuss and consider potential actions regarding the post. Revisit the chat anytime to follow up on posts you have reported.


## PROPOSED AUDITBOT FEATURES

**Prompts for User Action**

- Do you have any initial thoughts about this post?
- Do you want to report this post or explore other options?

**Suggestions for Support**

- Here are organizations you can explore to learn more about bias.
- Here's how you can start a discussion about this with friends.

**Resources for Learning**

- I can help you understand why others might think this is biased.
- I can help you understand Instagram's reviewing system for reportedly biased content.

**Updates on Reported Content**

- This post has received 12 more reports since you visited it. Here's a recap of where this leaves us in the review process.
- This post has been removed. Here's how we came to the decision to take it down.

## THANKS FOR VIEWING!

Team Be Free - John, Priya, Lisa, Yumi, and Martina
Module 2.5, Deliverable A: Elevator Pitch
UCRE Fall 2021

## Intro - hit them hard and grab their attention

I'm [name], and I'm representing team Be Free. We're here today to introduce our innovative solution for auditing instances of algorithmic bias on Instagram's social media platform. With the introduction of a chatbot to Instagram's user interface—named AuditBot for the purpose of auditing—we're proposing a holistic solution that will help social media end-users assess and audit potentially biased posts, enabling the co-creation of value between Instagram and its end-users.

## Dev steps and how it drove the solution

We approached the issue of auditing algorithmic bias through social proof to motivate and improve confidence of end-users in the auditing process.

From there, we used 6 different methods to user-test potential end-users, and discovered 3 overarching insights: the user's desire for simplicity, transparency, and anonymity.

End-users desire simplicity for easy discovery and traversal through the auditing process—essentially deformalizing traditional reporting; they also desire transparency from the platform on their audit's progress. Furthermore, end-users benefit from a layer of anonymity and social-proof-driven motivation—enough anonymity that they will not have to fear social consequences, but enough visibility to validate their desire to audit.

## Reiterating how we met those needs

Our solution will motivate users and improve their confidence in auditing harmful posts. It meets the user's need to have a layer of anonymity between themselves and their connections by including the AuditBot function within Instagram's user interface. The AuditBot chat promotes transparency between the platform and its end-users by providing a direct line of communication.

In summary, using AuditBot within Instagram will co-create value between end-users and the platform by meeting end-users' needs and by providing the platform with crucial data to drive future decisions. Now, we'll open the floor to you for any questions, and for viewing our poster in-depth.