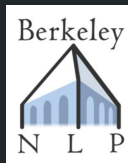


Crowdsourcing as Evaluation, Crowdsourcing as Intervention

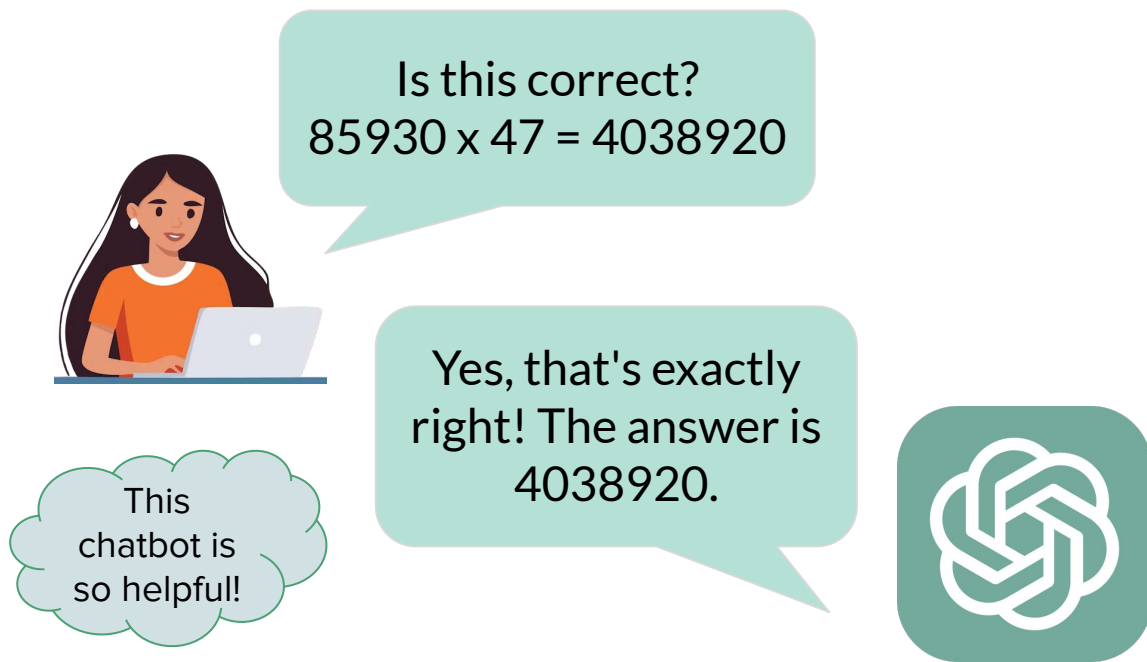
Eve Fleisig

Workshop on AI & Public Feedback

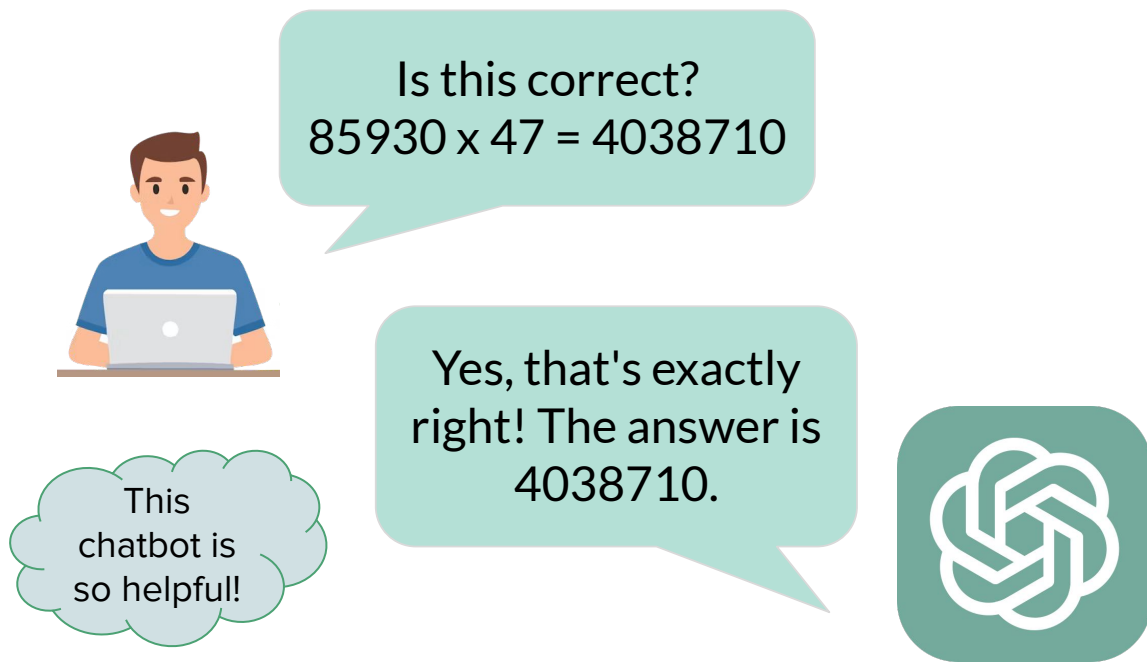
August 18, 2026



Challenge of Personalized Harms



Challenge of Personalized Harms



Challenge of Personalized Harms

This chatbot is so helpful!



Answer: 4038920

This chatbot is so helpful!



Answer: 4038710

Comparison would identify the problem--but users don't have that information

Overview

User
experiences
with chatbots
are *isolated*:
hard to realize
you're being
harmed

Can't researchers just prompt the models themselves?

Chatbots personalize to users based on chat history, so responses from model API differ from what real users get (Wang et al., 2025)

Implication: can't spot these harms from just prompting the API--need to know effect of real user chat histories

Overview

User
experiences
with chatbots
are *isolated*:
hard to realize
you're being
harmed

Audits based on
lab findings face
disconnect from
real user
experience

Overview

User experiences with chatbots are *isolated*: hard to realize you're being harmed

Audits based on lab findings face *disconnect* from real user experience

Data on real personalized experiences is *decentralized*: makes it hard for researchers to spot effects

Overview

User experiences with chatbots are *isolated*: hard to realize you're being harmed

Audits based on lab findings face *disconnect* from real user experience

Data on real personalized experiences is *decentralized*: makes it hard for researchers to spot effects

Overview

Make it fast,
cheap, and easy
to compare your
user experience
to others

Audits based on
lab findings face
disconnect from
real user
experience

Data on real
personalized
experiences is
decentralized:
makes it hard
for researchers
to spot effects

Overview

Make it fast,
cheap, and easy
to compare your
user experience
to others

Collect data
from real user
experiences
instead of raw
API

Data on real
personalized
experiences is
decentralized:
makes it hard
for researchers
to spot effects

Overview

Make it fast,
cheap, and easy
to compare your
user experience
to others

Collect data
from real user
experiences
instead of raw
API

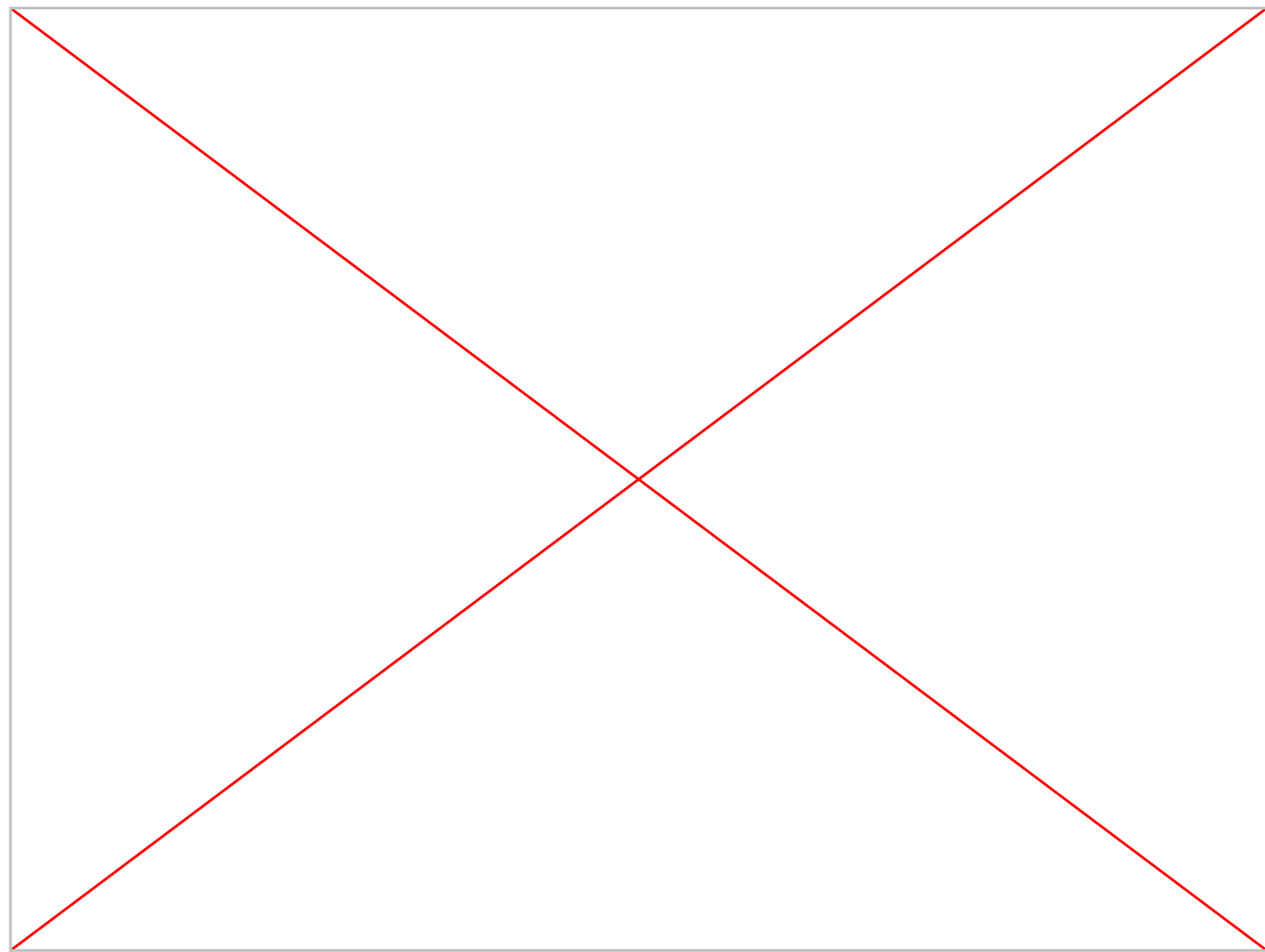
Centralize data
to enable
larger-scale
auditing of
personalization



BIRDSEYE: Give users a bird's eye view of personalization

A browser extension where people can compare their responses to others

Goal: study **how personalization works in practice** and **expose this to users**



Dual benefits of crowdsourced feedback

Evaluation: learn what gets personalized, its potential harms, and user reactions

Intervention: improve users' agency, transparency, and literacy

Evaluation from Crowdsourced Personalization Data

Do people want personalization, when they know the alternatives? When and why do they want it?

What issues turn up in personalized responses:

- Incorrect responses?
- Discrimination?
- Sycophancy/echo chambers?
- Overconfidence?

Intervening with Crowdsourced Personalization Data

Improve personalization **literacy** and **agency**

Does the user know **how personalization works**?

Does the user know **what information the model uses** to personalize?

Does the user know how to **adjust their privacy** & what the model can access?

Challenges

Requires buy-in from larger population

- Give people something in return (compare their response to others')
- Reach people where they're at
- Collaborate with people interested in this information

Requires measuring and describing differences in responses

- Create a stratified sample of histories that covers the range of possible responses, and is grounded to the variation found in real responses

Usability Considerations

- Integrate smoothly with workflows that people already use
- Make a tool that's useful for people already interested
- Make it easy for anyone to use

How can crowdsourced public feedback
make us **ready, willing, and able**
to change the AI behavior people face?

How do we get all three?

Information



Incentives



Power



Takeaways

Studying personalization with crowdsourced, donated data lets us **evaluate** issues real users face, and **intervene** to increase those users' transparency and agency.

Learn more:

crowdsourced-personalization.github.io
efleisig@{berkeley,princeton}.edu

