algorithm UNUSAFED.

exploring experiences, sensemaking, future solutions around algorithmic content harm with a focus on TikTok

Nicole Chi

nicole.chi@berkeley.edu

Keming Gao

keming.gao@berkeley.edu

Joanne Ma

joannema@berkeley.edu

SCHOOL OF INFORMATION MIMS 2022 CAPSTONE

Motivations and Summary

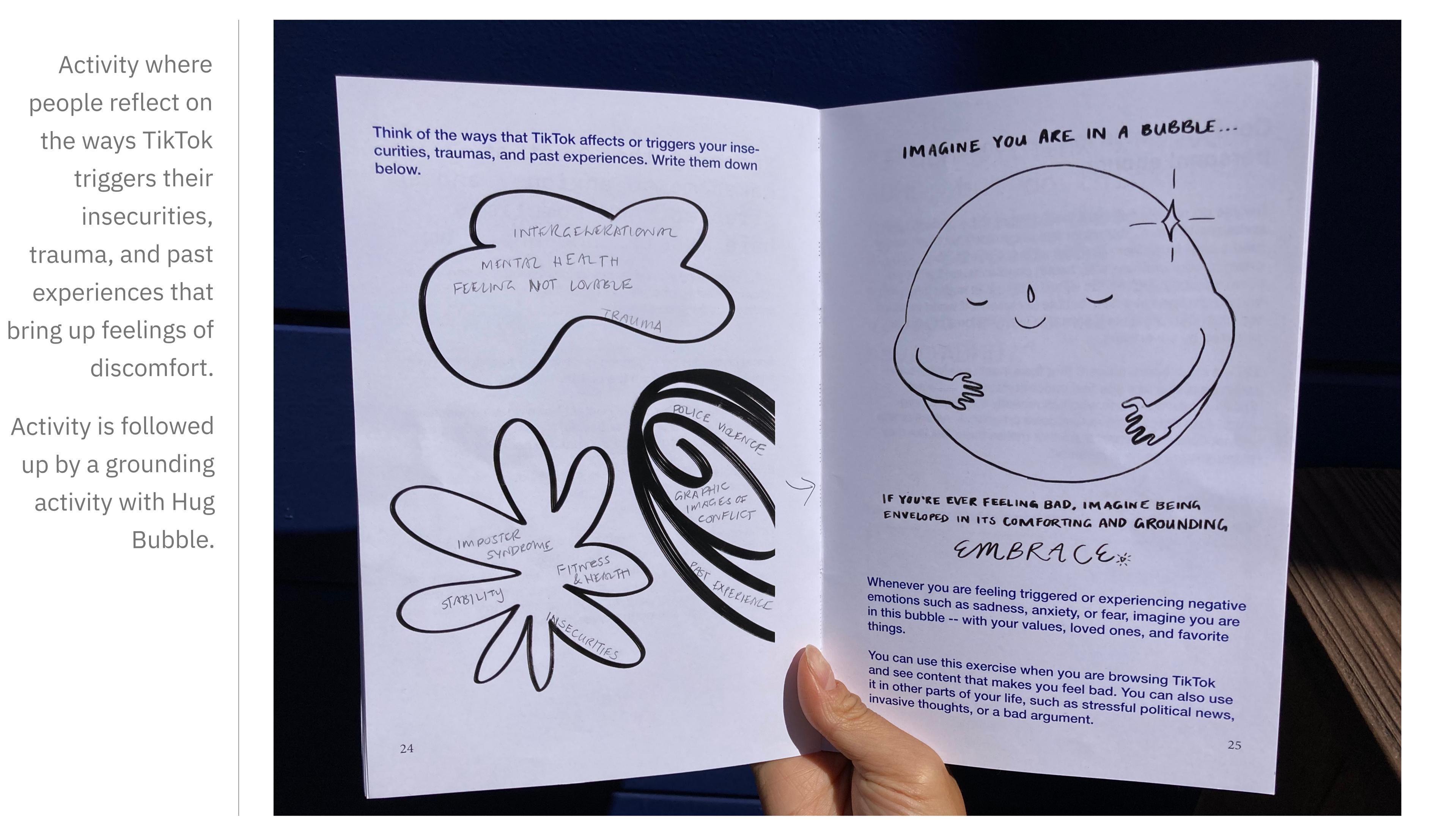
Every day, algorithms are deciding what you watch, listen to, and consume. This is the world of algorithmic content recommendations – videos or other media recommended to you automatically by a computer program based on your digital activity, identity, and interests.

- How do people experience algorithmic content harm on social media platforms like TikTok?
- What affordances make it difficult to address this type of harm?

What solutions can we create to help people resist algorithmic content harm?

Activity where people reflect on the ways TikTok triggers their insecurities,

Activity is followed activity with Hug Bubble.



Exploratory Research

We conducted expert interviews with two

the founder of a nonprofit dedicated to

2. Browser extension study

sociotechnical researchers who had published

independent algorithmic auditing, one trust &

safety expert who is currently going through ED

recovery, and one engineer on Twitter's META (ML

Ethics, Transparency, and Accountability) team.

We provided six participants with a browser

extension to record incidents where they felt

harmed or uncomfortable with algorithmically

recommended content on any platform (YouTube,

work on social media feeds and eating disorders,

1. Expert interviews

Solution 1: Educational zine

Our first solution is a zine for ML education that helps people make sense of their own experiences with algorithmic content feeds, relate to algorithmic content harm, and build healthier relationships with their TikTok algorithm.

The zine includes:

- 1. Quotes and stories from participants we interviewed through our research
- 2. Sensemaking activities that people can fill out in the zine itself
- 3. Strategies for mitigating algorithmic content harm

The entire zine is written to be as accessible to non-technical folks as possible.

By educating people about algorithmic content harm and providing immediate strategies for awareness and mitigation, we hope they can better resist, process, and be aware of these potential harms.



Solution 2: Digital toolkit

Our second solution is a digital toolkit to help users access and explore their own TikTok data.

The toolkit includes:

- 1. Guide for how to request your TikTok data
- 2. Tool for extracting view history from TikTok file
- 3. Tool to scrape the sound, creator, description, and hashtags for each video

While TikTok allows their people to request their view history (which most social media platforms do not), the process to obtain and make sense of this data presents a huge barrier.

Through this easily downloadable toolkit, we hope to encourage more people to "look back" at their entire comprehensive record on TikTok so they can reflect on what types of content they have been served over a longer span of time.



Twitter, Facebook, etc). We interviewed them to see what harms they experienced and what factors affected their feelings around that harm.

Key Finding: Algorithmic harm is hard to detect and mitigate, because its effects can be subtle and accumulate over time. User sensemaking and education is important to help people understand and address how they are harmed.

Primary Research

1. Testing information visualizations for user sensemaking

After deciding to focus on algorithmic sensemaking on TikTok through users' own data, we conducted 30-min user interviews and showed participants prototypes of what visualizations of their TikTok data might look like.

2. Understanding negative emotions, identity, and control on TikTok

We conducted 1-hour interviews with participants who self-identified as feeling negative emotions while using TikTok. Questions included asking participants to describe types of videos that made them feel negative emotions, and how accurately they felt TikTok captured their identity.

We asked participants to request their TikTok data, and personally tagged and watched hundreds of videos to better understand more context around their interviews.

"Some of this content makes me so anxious, and so stressed. Sometimes I hate myself...and who I am."

INTERVIEW PARTICIPANT

Key Findings: People's content feeds are based on their lifestyle, interests, and/or identities. Algorithmic content harms range in severity, and include harms related to: identity, trauma, accurate or inaccurate personalization, unexpected emotions, and negative societal impact. User control vastly differs as well, and some users discount their own negative feelings as fleeting or insignificant.

Our Proposed Future

The future we envision is one in which everyone is able to easily evaluate their history with their algorithms. Think Spotify Wrapped...but for any algorithmic content feed, and showing trends over time.

Imagine being able to see how social media algorithms have affected your interests and identity over months or even years.

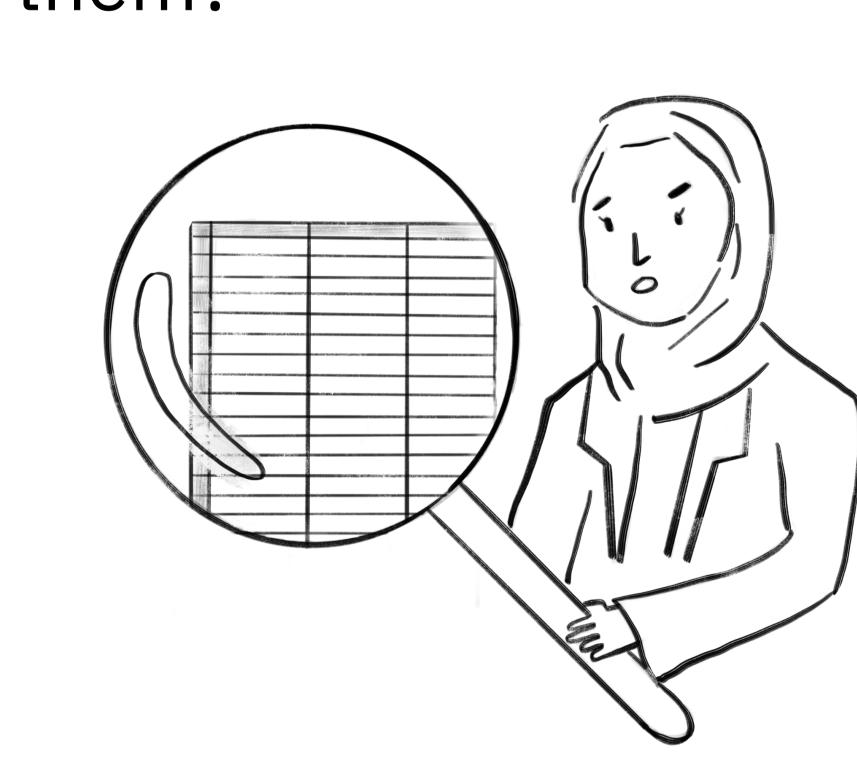
Imagine being able to get a summary of types of videos you've seen so you can identify trends that, in hindsight, weren't great for your mental health or self image. In doing so, you could reflect on sources of harm, identify what you might need to avoid or work through, and acknowledge how you've grown.

ML transparency or accountability through revealing technical details of algorithms is not necessarily useful to most users. Our ask is simple - more tools to help us proactively make sense of our own lives online.

Empowering Consumers

We started off this project being very interested in empowering consumers to do "everyday algorithmic auditing" to help surface algorithmic content harms. Through research, we gradually realized that this was impossible -- at least, for now. How can everyday users audit algorithms when they don't fully understand how they are affected by them?





In order to achieve our desired future above, we believe ML education tools like Algorithm Unwrapped are critical. Much like privacy, consumers need to understand why they should care about algorithmic

transparency before they will push for platforms and policy to give them ways to actually understand their digital lives.

Next Steps

- Publish TikTok Unwrapped: A Practitioner's Guide to share learnings about our research and toolkit that may be useful to researchers and civil society organizations
- Donate copies of TikTok Unwrapped zine to nonprofits, with a focus on youth organizations. We are currently in conversation with 10 nonprofits and organizations who are interested in collaborating and co-hosting workshops.
- Create a tool to help people visualize their downloaded TikTok data. Check out our collaborators' work for inspiration:

