Recommender System Trust and Safety
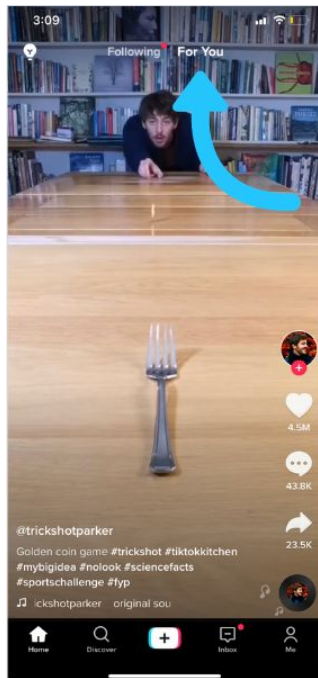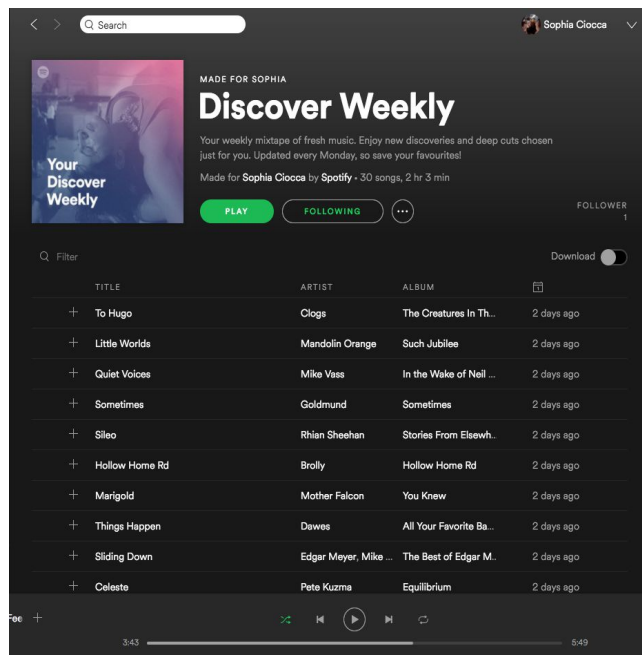
Murad-Vishva-Thomas

Recommender Systems are everywhere

Taken from LaterBlog

Taken from Medium

Taken from Twitter

Captivity

- Developers motivation [Seaver]
- Absurd outcomes [Alkhatib et al.]

Privacy

- Presents utility to privacy tradeoff. [Mohallick et al.]
- Dangerous inferences. [Kashmir Hill, Forbes]

Research Question

How do personalized recommender systems affect human behavior with respect to trust and safety according to user perception across different platforms?
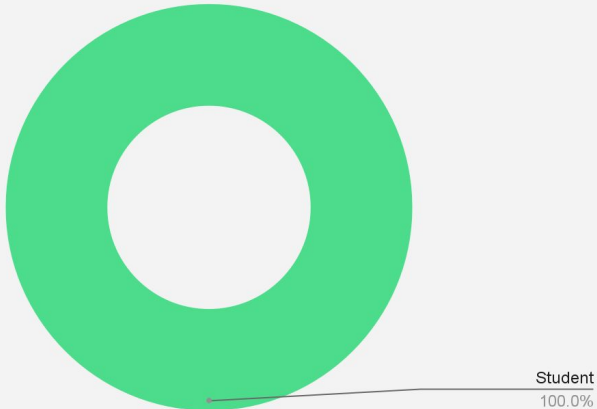
Methods/Design

- Interviews
  - Semi-structured interviews
    - Convenience sampling
    - Interviewer and observer
    - Audio only recording
  - 3 generic questions
  - 5 specific questions for each platform
  - Video calls through Zoom
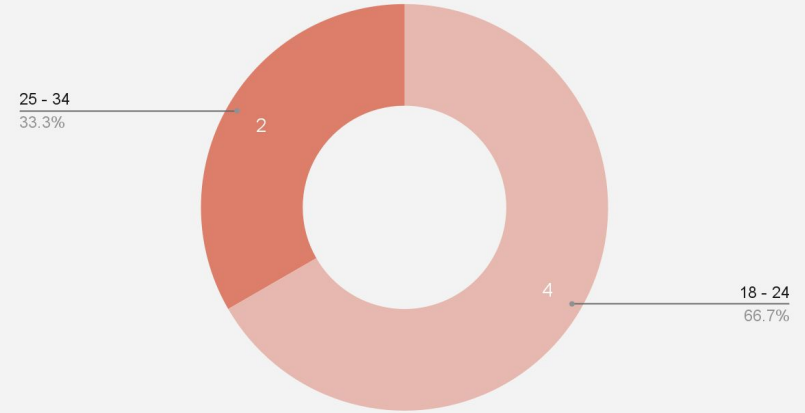
Video Call Interviews
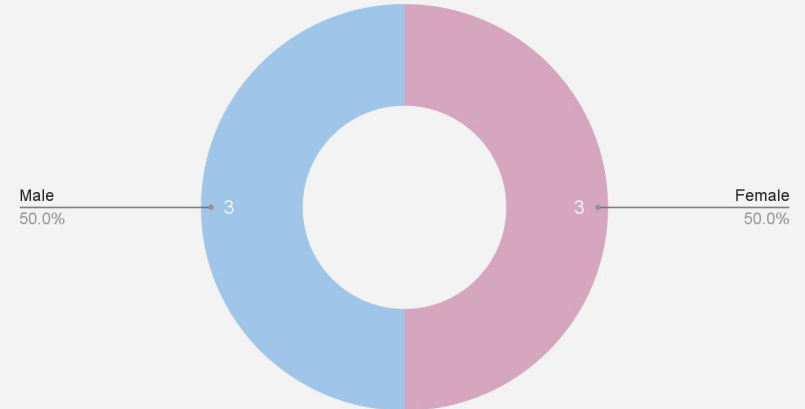
- 6 participants
  - Time between 31 and 48 minutes

Age groups

25 - 34
33.3%

2

4

18 - 24
66.7%

Occupation

Student
100.0%

Gender

Male
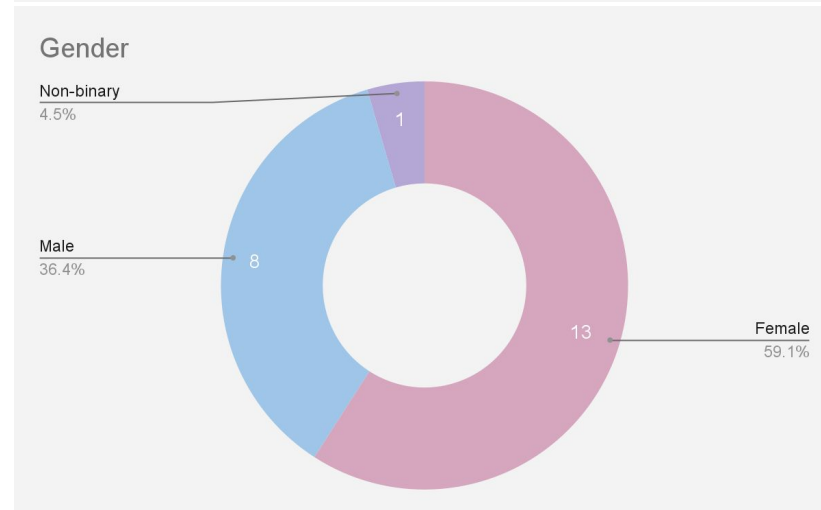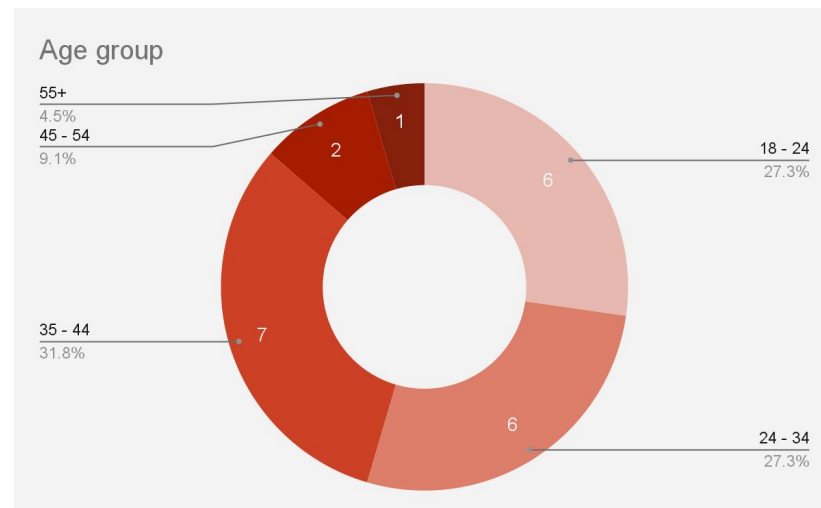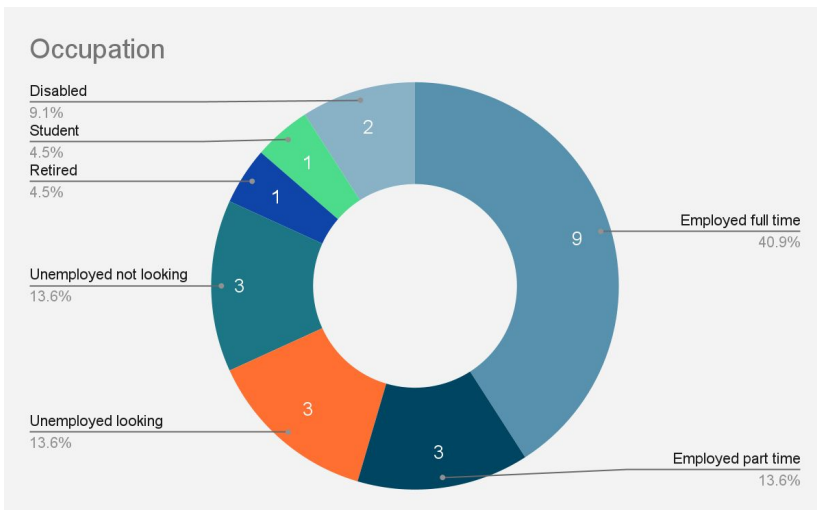50.0%

3

3

Female
50.0%

Online Questionnaire

- Qualtrics questionnaire
  - 2 generic questions
  - 1 attention question
  - 6 questions per platform
  - 4 Demographic questions
- Prolific study
  - Standard Sample
  - US only
  - Pilot run with 5 participants
  - Full run with 22 participants
  - Total spend $25 (rewards + fees)

Full Run

- 22 participants for full run
  - $0.67 reward (5 minutes) = $8.04 per hour
  - Mean Captcha score 0.98 (min 0.90)
  - 100 % acceptance
  - Average time: 5:36 /Mean time: 3:28

Age group

55+ 4.5%
45 - 54 9.1%
18 - 24 27.3%
35 - 44 31.8%
24 - 34 27.3%

Gender

Non-binary 4.5%
Male 36.4%
Female 59.1%

Occupation

Disabled 9.1%
Student 4.5%
Retired 4.5%
Unemployed not looking 13.6%
Unemployed looking 13.6%
Employed full time 40.9%
Employed part time 13.6%

Interview Results

- Company Size
- Recommendation Quality

Online Questionnaire Results

- TikTok, Instagram had much more influence on resource consumption (time, money, etc) than other social media applications.
- Participants felt there was more biased content in TikTok/Instagram/Amazon than other applications.
- Participants did not feel that their personal safety was at risk because of the data collected by the recommendation system.

Limitations

- Interview – participant demographics (Age, occupation)
- Less participants for online surveys (22)
- Lack of options for entertainment applications in surveys (Netflix/Spotify).

Conclusion

- Trust and safety perception is dependent on company size and app reputation.
- Data collection is acceptable for quality recommendations.
- Financial data and Location tracking is off limits.
- Applications for Music are most trusted except those by conglomerates.

Questions?

- Trust and safety perception is dependent on company size and app reputation.
- Data collection is acceptable for quality recommendations.
- Financial data and Location tracking is off limits.
- Applications for Music are most trusted except those by conglomerates.