How Developers (Don't) Think about Gender Privacy

Presented by Elijah Bouma-Sims, *Carnegie Mellon University*

Co-author Yasemin Acar, *The George Washington University*

Carnegie Mellon University

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Based on research conducted in Summer 2021 at the *Max Planck Institute for Security and Privacy*

1

Overview

1. Discussion of Gender Privacy Concerns
2. Overview of Our Study of Gender + Developers
   a. Reddit Study
   b. Interview Study
3. Recommendations on Handling Gender

Gender and Computing

- <u>Gender</u> is "a person's internal sense of being male, female or something else" (American Psychological Association)
- Ubiquitous in society and also computing
- Users may be required to disclose gender to use a product or service.
- Problem?

Gender and Computing

- Inadequate options = <u>Accessibility barrier</u> for people who fall outside of the binary
- Gender identity may <u>change over time</u>
- <u>Privacy dilemma</u> for individuals who don't wish to disclose their gender
- Gender data may be used to shape user experience in <u>unwanted</u>, <u>unintended</u> or <u>harmful</u> ways

An illustrative example

"Male" Icon

"Female" or "Diverse" Icon

How do developers approach gender?

1. How do developers ask users for gender?
2. How do developers use user gender data?
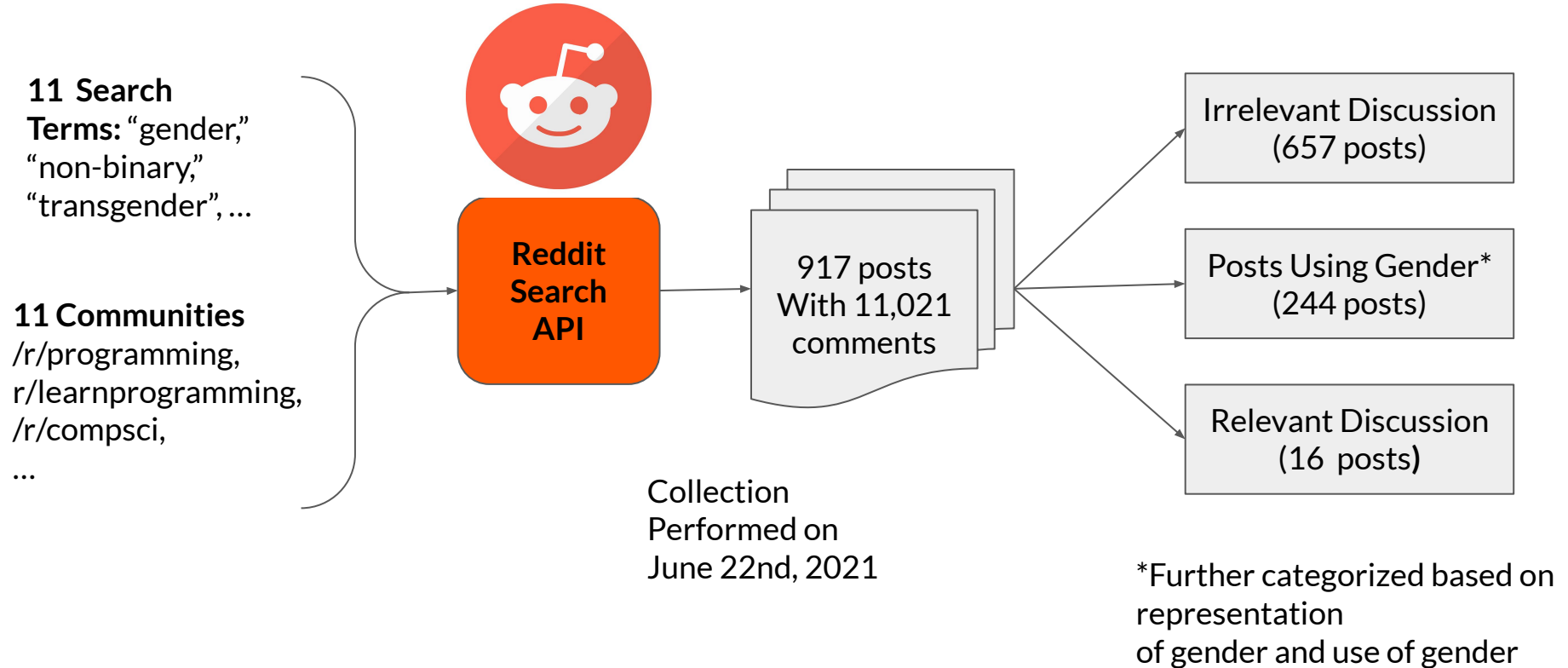3. What advice do developers receive about using gender data?

How do developers approach gender?

- Analysis of practice on Reddit
  - Discussions about using gender data
  - Code snippets which show how gender is actually used
- Semi-structured interviews with developers
  - Probe participant attitudes and approaches towards gender

The Reddit Study

Reddit Data Collection

11 Search Terms: "gender," "non-binary," "transgender", …

11 Communities
/r/programming,
r/learnprogramming,
/r/compsci,
…

Reddit Search API

917 posts
With 11,021
comments

Collection
Performed on
June 22nd, 2021

Irrelevant Discussion
(657 posts)

Posts Using Gender*
(244 posts)

Relevant Discussion
(16 posts)

*Further categorized based on representation
of gender and use of gender

Reddit: the good

- Very <u>little direct discussion</u> of gender
- In a small number of cases, <u>users asked directly for advice</u> on how to program gender fields
- These posters received inclusive recommendations

Reddit: and the bad

- Observed code <u>did not follow inclusive practices</u>
  - The most common representation of gender in code snippets was a strict binary (80 posts)
- We observed a variety of uses for using gender, although the most common identifiable use was for <u>audience demographics </u>(30 posts)
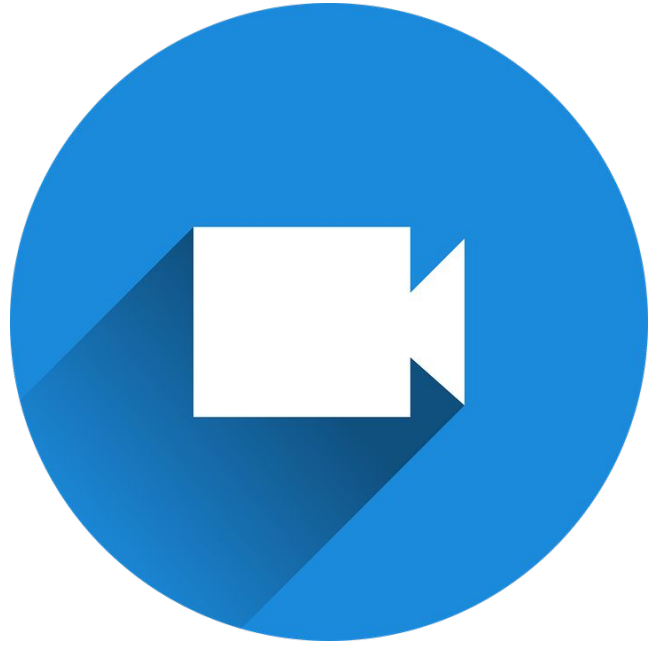
The Interview Study

The Interview Study

- <u>Broad inclusion criteria</u>: at least a year of professional experience and/or a related degree (E.g. computer science, software engineering)
- Recruitment performed on *Twitter* and the freelancing platform, *Upwork*
- Data analyzed using Qualitative coding on transcripts

Interview Results

- The plurality of our interviewees (7) showed a preference for treating <u>gender as a spectrum</u>
  - "I guess the easiest way to [collect gender] would just to be having [sic] some string input field, like `what gender do you identify as?'"
- As with Reddit, <u>audience demographic analysis</u> was the most common use identified for gender data

Interview Results

- Generally, interviewees seemed <u>unfamiliar with the topic area</u>
  - Questions often required clarification or reiteration
- One interviewee had previously worked on a social media website for trans identifying individuals
  - Platform did not directly encode gender
  - Users employed post tags to express their identity

Takeaways for Privacy Engineers

- While we saw signs of positive developments in how gender is treated, <u>deliberate standards</u> should be set
- Software should accommodate individuals who <u>cannot provide their gender</u> due to privacy concerns
- Unnecessary collection of gender data should be avoided

If it is necessary to collect gender data…

- Gender is a complex social construct which cannot be neatly summarized with a single boolean
- The <u>purpose</u> of collection should be clear
  - I.e. if there is a legal or regulatory requirement, that should be made clear
- The software should accommodate potential <u>changes over time</u>

If "male" and "female" are your only gender options, <u>you're doing it wrong.</u>

Presenter Contact:
eboumasi@andrew.cmu.edu
www.elijahboumasims.com