Leveraging requirements from the crowd for more inclusive software

Dr Kelly Blincoe

Human Aspects of Software Engineering Lab

University of Auckland, New Zealand

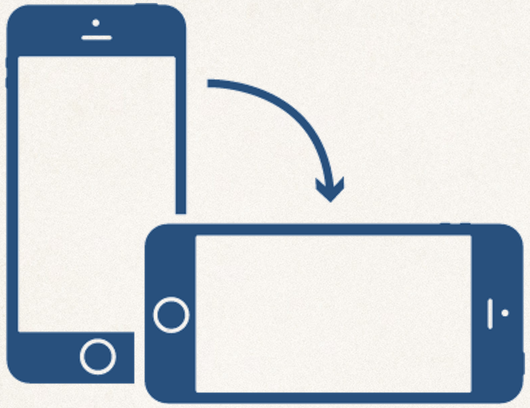
@KellyBlincoe

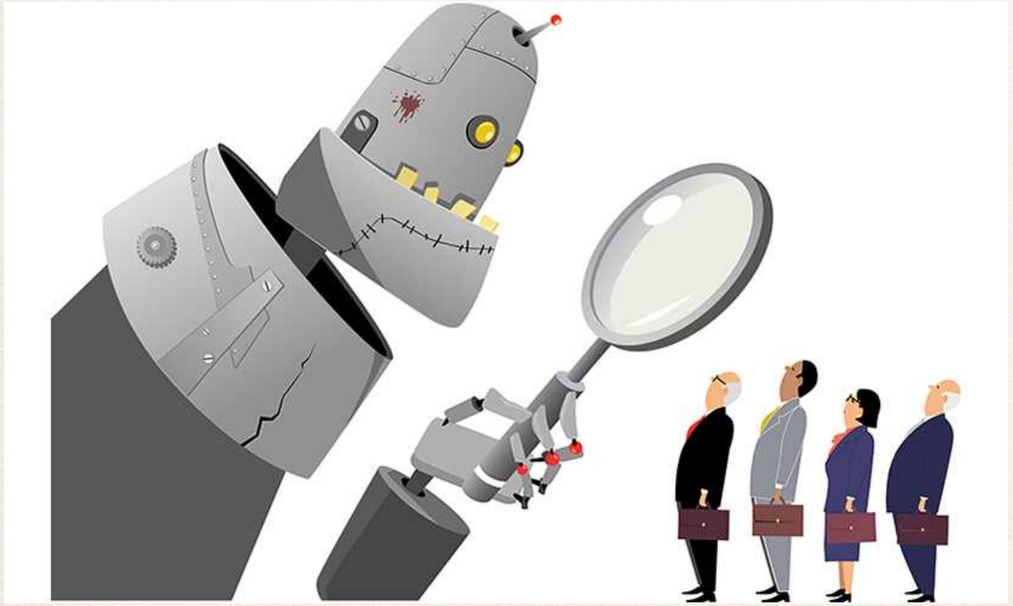
2021 Workshop on Human Centric Software Engineering and Cyber Security (HCSE&CS)

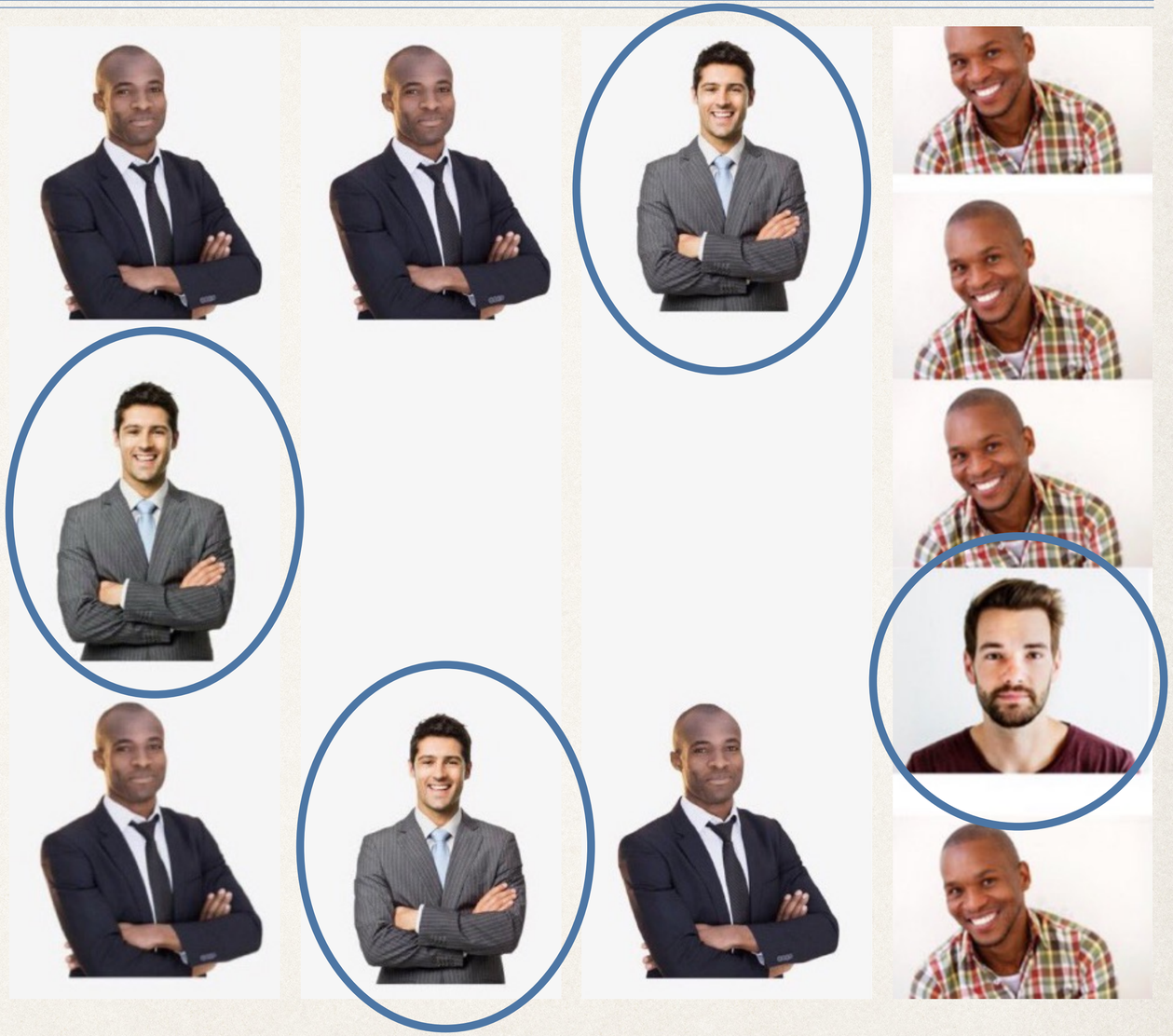
Importance of Software

Software is an integral part of society that impacts our daily lives

Software can exclude and discriminate







Images: https://www.businessinsurance.com/article/00010101/NEWS06/912324509/Amazon-scraps-secret-AI-recruiting-tool-that-showed-bias-against-women; https://ruinmyweek.com/trending/twitter-photo-preview-bias/

Why does software exclude and discriminate?

The needs of all of society are not being considered in its creation

Image: https://www.clipartkey.com/

App 4 Aug

★★★☆☆ by Customer40523662

The app is great, but first of all you should fix search box, because you can't search certain users. Secondly you should add reply box so people could answer the questions, and make it so eveyone could see. Looking forward!

@twitter u guys really need to add the direct return to home button in Twitter APP.

#suggestion #app #MuchNeededChange #ThanksInAdvance

Image: https://www.clipartkey.com/

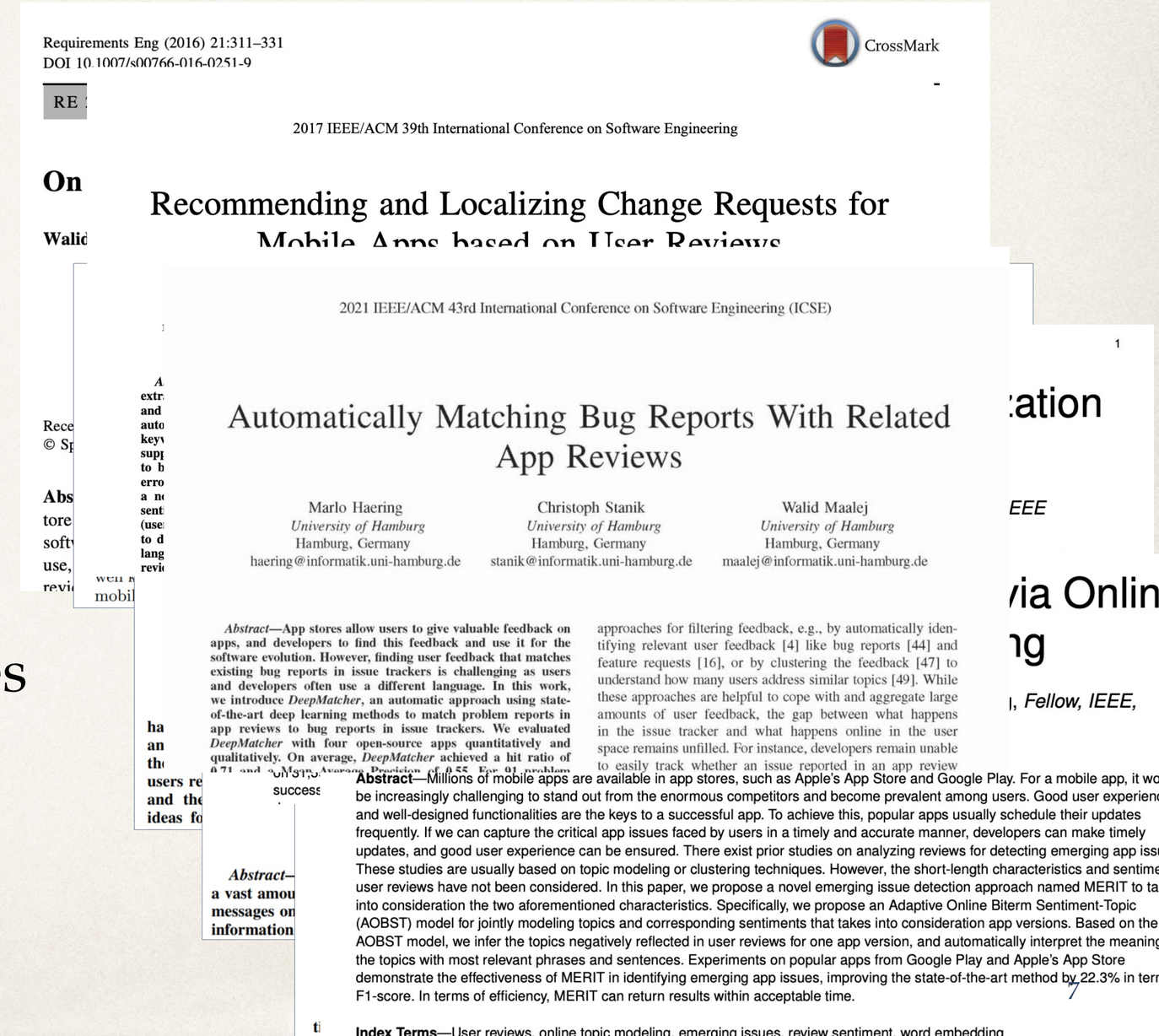
Online user feedback

- Feedback on App Stores contains requirement relevant information (Pagano and Maalej, 2013)
  - user experiences
  - bug reports
  - feature requests
- Similar information available on
  - Twitter (Guzman et al., 2016)
  - Forums (Tizard et al., 2019)

James Tizard, Hechen Wang, Lydia Yohannes, and **Kelly Blincoe**. 2019. Can a conversation paint a picture? mining requirements in software forums. In 2019 IEEE 27th International Requirements Engineering Conference (RE), 17–27.

Methods to automatically extract requirements

- Huge amount of online feedback
- Recent research investigating methods to automatically
  - Classify feedback
  - Cluster related feedback
  - Match feedback to known issues



Requirements Eng (2016) 21:311–331
DOI 10.1007/s00766-016-0251-9

2017 IEEE/ACM 39th International Conference on Software Engineering

Recommending and Localizing Change Requests for Mobile Apps based on User Reviews

2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)

Automatically Matching Bug Reports With Related App Reviews

Marlo Haering
University of Hamburg
Hamburg, Germany
haering@informatik.uni-hamburg.de

Christoph Stanik
University of Hamburg
Hamburg, Germany
stanik@informatik.uni-hamburg.de

Walid Maalej
University of Hamburg
Hamburg, Germany
maalej@informatik.uni-hamburg.de

***Abstract*—App stores allow users to give valuable feedback on apps, and developers to find this feedback and use it for the software evolution. However, finding user feedback that matches existing bug reports in issue trackers is challenging as users and developers often use a different language. In this work, we introduce *DeepMatcher*, an automatic approach using state-of-the-art deep learning methods to match problem reports in app reviews to bug reports in issue trackers. We evaluated *DeepMatcher* with four open-source apps quantitatively and qualitatively. On average, *DeepMatcher* achieved a hit ratio of**

approaches for filtering feedback, e.g., by automatically identifying relevant user feedback [4] like bug reports [44] and feature requests [16], or by clustering the feedback [47] to understand how many users address similar topics [49]. While these approaches are helpful to cope with and aggregate large amounts of user feedback, the gap between what happens in the issue tracker and what happens online in the user space remains unfilled. For instance, developers remain unable to easily track whether an issue reported in an app review

Abstract—Millions of mobile apps are available in app stores, such as Apple's App Store and Google Play. For a mobile app, it wo
be increasingly challenging to stand out from the enormous competitors and become prevalent among users. Good user experien
and well-designed functionalities are the keys to a successful app. To achieve this, popular apps usually schedule their updates
frequently. If we can capture the critical app issues faced by users in a timely and accurate manner, developers can make timely
updates, and good user experience can be ensured. There exist prior studies on analyzing reviews for detecting emerging app iss
These studies are usually based on topic modeling or clustering techniques. However, the short-length characteristics and sentime
user reviews have not been considered. In this paper, we propose a novel emerging issue detection approach named MERIT to ta
into consideration the two aforementioned characteristics. Specifically, we propose an Adaptive Online Biterm Sentiment-Topic
(AOBST) model for jointly modeling topics and corresponding sentiments that takes into consideration app versions. Based on the
AOBST model, we infer the topics negatively reflected in user reviews for one app version, and automatically interpret the meaning
the topics with most relevant phrases and sentences. Experiments on popular apps from Google Play and Apple's App Store
demonstrate the effectiveness of MERIT in identifying emerging app issues, improving the state-of-the-art method by 22.3% in ter
F1-score. In terms of efficiency, MERIT can return results within acceptable time.

Index Terms—User reviews, online topic modeling, emerging issues, review sentiment, word embedding

Advanced Text Embeddings will improve accuracy

- Word-frequency methods (TF-IDF, Bag of Words)
- Topic models (LDA, BTM, GSDMM)
- Transformer based models (SBERT, USE, LaBSE)
  - Best at grouping similar feedback together
  - USE most performant

Peter Devine, Yun Sing Koh, and **Kelly Blincoe**. 2021. Evaluating Unsupervised Text Embeddings on Software User Feedback. In *Proceedings of the International Workshop on Artificial Intelligence and Requirements Engineering*.

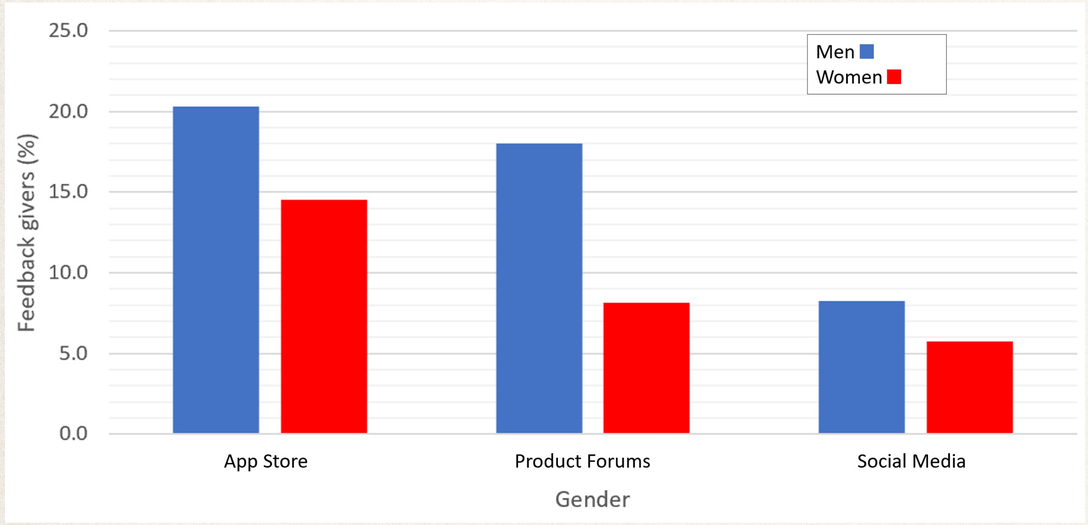
Prioritization of online user feedback

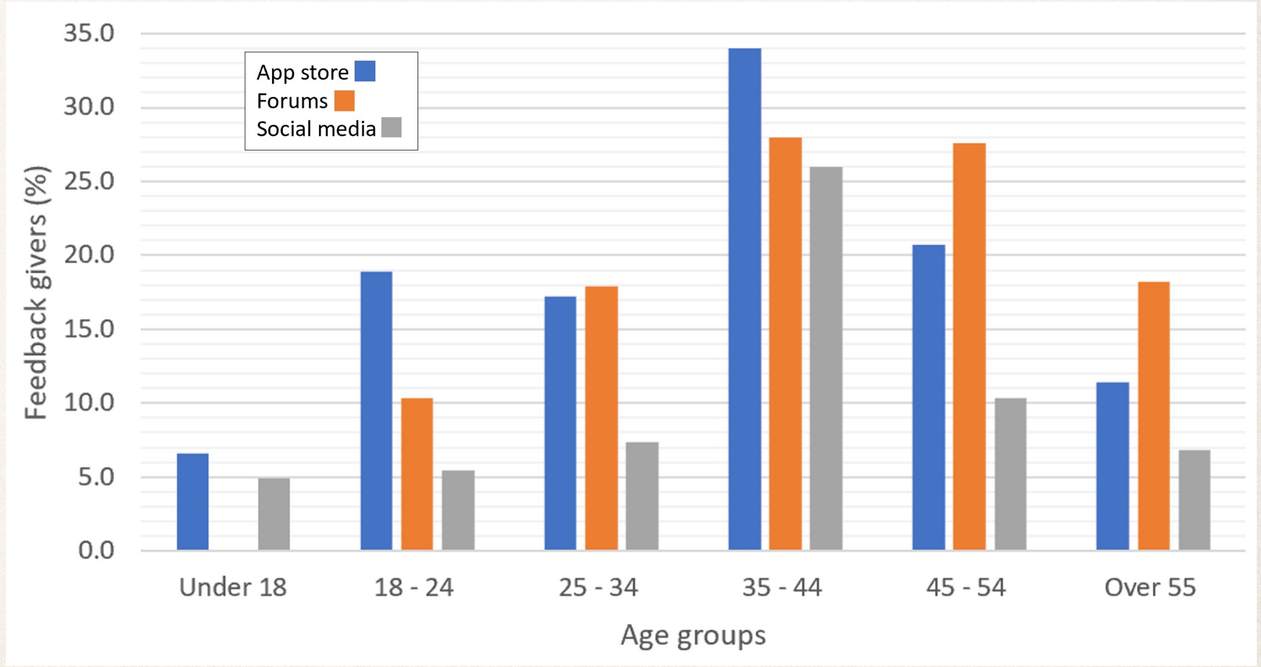
- With so much feedback online, how to prioritize what to add / fix? Prior work has considered features like:
  - Sentiment, ratings
  - Frequency, consensus
  - Retweets, likes, number of followers
- How do we know we are not simply amplifying the voices of the majority?
  - We often compare new methods with how developers prioritize, is that the right way?

Kifetew et al., 2021; Etaiwi et al., 2020; Guzman et al., 2017; Keertipati et al., 2016

Is online user feedback representative?

- Some preliminary evidence of feedback giving differences across genders and geographic locations (Guzman et al., 2018; Guzman and Rojas, 2019)
- Survey of Software Users about feedback giving





James Tizard, Tim Rietz, and **Kelly Blincoe**. 2020. Voice of the users: A demographic study of software feedback behaviour. In *2020 IEEE 28th International Requirements Engineering Conference (RE)*, 55–65.

Motivations to give feedback

App Store	(%)
1. Show appreciation	65.15
2. Influence improvement	52.02
3. Show dissatisfaction	34.85
4. Recommend to others	29.80
5. Discourage others	12.63

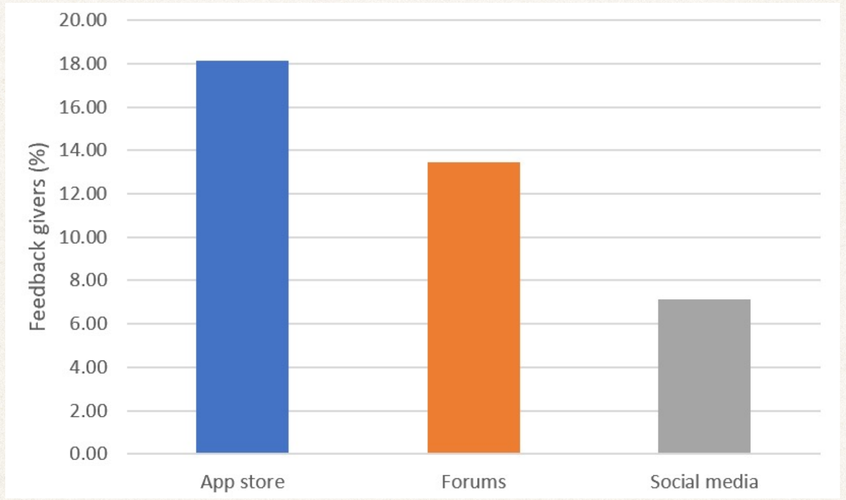
Product Forum	(%)
1. Get help	70.37
2. Influence improvement	44.29
3. Show appreciation	26.43
4. Recommend to others	17.86
5. Show dissatisfaction	16.43

Social Media	(%)
1. Show appreciation	56.76
2. Influence improvement	51.35
3. Show dissatisfaction	37.84
4. Connect or socialise	35.14
5. Recommend to others	32.43

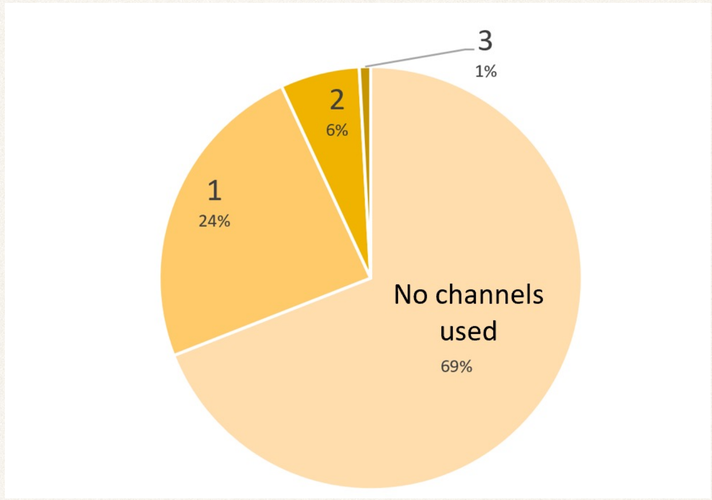
James Tizard, Tim Rietz, Xuanhui Liu, and **Kelly Blincoe**. 2021. Voice of the Users: A study of software feedback differences between Germany and China. In Proceedings of the International Workshop on Crowd-Based Requirements Engineering.

James Tizard, Tim Rietz, Xuanhui Liu, and **Kelly Blincoe**. 2021. Voice of the users: an extended study of software feedback engagement. *Requirements Engineering*: 1–23.

Where is feedback given



Written Feedback Givers (%)



Number of Feedback Channels

James Tizard, Tim Rietz, and **Kelly Blincoe**. 2020. Voice of the users: A demographic study of software feedback behaviour. In *2020 IEEE 28th International Requirements Engineering Conference (RE)*, 55–65.

James Tizard, Tim Rietz, Xuanhui Liu, and **Kelly Blincoe**. 2021. Voice of the users: an extended study of software feedback engagement. *Requirements Engineering*: 1–23.

Embedding inclusion in the feedback analysis

- Need methods that work across multiple feedback channels
- Ways to engage more diverse set of users in the feedback loops
- Prioritization techniques that consider diversity and underrepresented voices

James Tizard, Hechen Wang, Lydia Yohannes, and **Kelly Blincoe**. 2019. Can a conversation paint a picture? mining requirements in software forums. In 2019 IEEE 27th International Requirements Engineering Conference (RE), 17–27.

James Tizard, Tim Rietz, Xuanhui Liu, and **Kelly Blincoe**. 2021. Voice of the users: an extended study of software feedback engagement. *Requirements Engineering*: 1–23.

Why does software exclude and discriminate?

The needs of all of society are not being considered in its creation – because the software industry is not representative of all of society

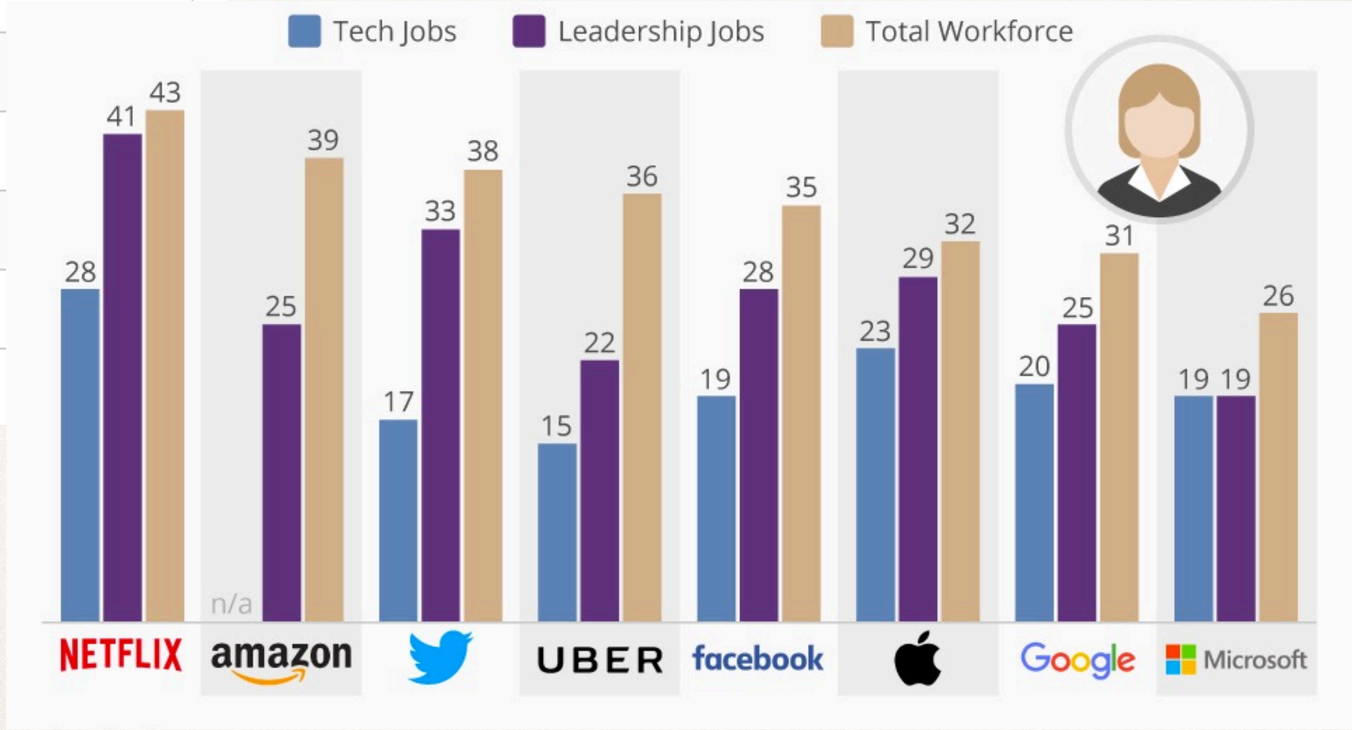
Image: https://www.clipartkey.com/

Tech's Diversity Problem

New Zealand diversity stats for tech

Ethnicity	New Zealand population	N.Z. information media and telecommunications industry proportion
White/European	70%	72%
Māori	16%	8%
Asian	15%	19%
Pacific peoples	8%	5%
Other	1%	NA

Source: New Zealand 2018 Census



Source: https://www.computerworld.com/article/3567095/it-snapshot-ethnic-diversity-in-the-tech-industry.html; : https://www.statista.com/chart/4467/female-employees-at-tech-companies/

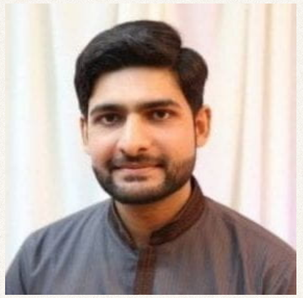
Human-centric SE research

- A call to embed inclusion in software *design*, *education*, and *research*.
  - How can we embed inclusive design into SE?
  - How can we educate the future software engineers to design for inclusion?
  - Is our own research further embedding inequities?
    - Are we actively considering the needs of marginalized groups?
    - How can we design methods and tools for a diverse and inclusive SE community?

Thanks to my collaborators

- Co-authors on this work
  - James Tizard
  - Hechen Wang
  - Peter Devine
  - Lydia Yohannes
  - Tim Rietz
  - Xuanhui Liu
  - Yun Sing Koh

Human Aspects of Software Engineering Lab












https://hasel.auckland.ac.nz

Let's discuss

- A call to embed inclusion in software *design*, *education*, and *research*.

Dr Kelly Blincoe

Human Aspects of Software Engineering Lab

University of Auckland, New Zealand

https://hasel.auckland.ac.nz
https://kblincoe.github.io
k.blincoe@auckland.ac.nz

@KellyBlincoe