Embedded Ethics
Module
Recommender System Objectives

Welcome to Embedded Ethics!

1) This is an active, participatory module – your contributions will help make it successful!

2) Our goal is not to tell you *what* to think about ethical problems, but to give you some tools for *how* to think about them.

Group Exercise

- Suppose that you are an intern at Reddit in charge of the site's recommender system.

- This is a vast oversimplification, but imagine that the algorithm works this way: users subscribe to subreddits, and see the posts that are most upvoted by users of those subreddits. Advertisements are sprinkled occasionally into the posts.

- Now imagine that Reddit has just been acquired by a billionaire who has fired half of the employees and demanded that as part of "Reddit 2.0", the algorithm must be improved to create as much engagement and profit as possible.

- List at least three changes you would make. We'll discuss them in 6-8 minutes.

Two Ethical Questions

Q1) What consequences will be produced by your program, feature or design?

- What? (changes in happiness, resources, control, freedom, etc)
- For whom? (Users, designers, other stakeholders)
- When? (Soon, in the medium future, in the far future)
- Positive and negative consequences

If the negative consequences > positive consequences, you should rethink your design decision, or try to reduce or mitigate the negative consequences….

Note: we often think that it is worse to harm someone than it is to fail to benefit them

Q2) Would your program, feature or design violate anyone's moral rights?

Ronald Dworkin (1977): Rights are "trumps": if person P has a right not to have X done to them, other people shouldn't do X to them, *even if it produces the best consequences*

If your program, feature or design violates someone's rights, that gives you a very strong reason to modify it!

Consider Philippa Foot's famous philosophy/ethics thought experiment of the traveler and the organ transplant…

A brilliant transplant surgeon has five patients, each in need of a different organ, each of whom will die without that organ. Unfortunately, there are no organs available to perform any of these five transplant operations. A healthy young traveler, just passing through the city the doctor works in, comes in for a routine checkup. In the course of doing the checkup, the doctor discovers that his organs are compatible with all five of his dying patients. Suppose further that if the young man were to disappear, no-one would suspect the doctor. (Foot, 1967)

Discussion Question

What are some other moral rights that people arguably have?

Positive Consequences

One positive consequence: recommender systems can narrow down content for users.

When there are a lot of choices, what are the alternatives to recommender systems?

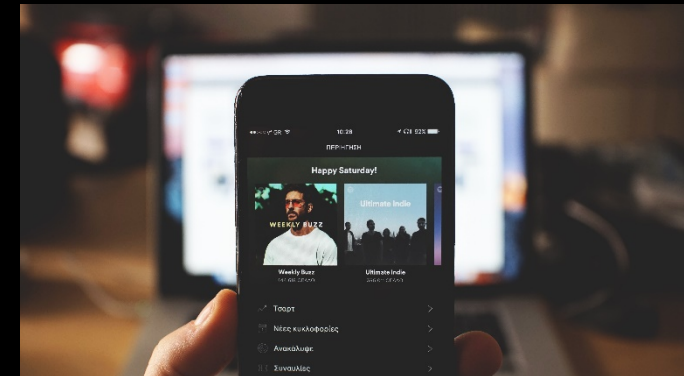Going through large amounts of content by yourself

Relying on expertise of others

Random chance

Sometimes recommender systems can also help users find content that is appropriate for them (which they couldn't find easily just by previewing it):

Collaborative Filtering

Collaborative filtering "uses the known preferences of a group of users to make recommendations or predictions of the unknown preferences for other users." (Su and Khoshgoftaar, 2009)
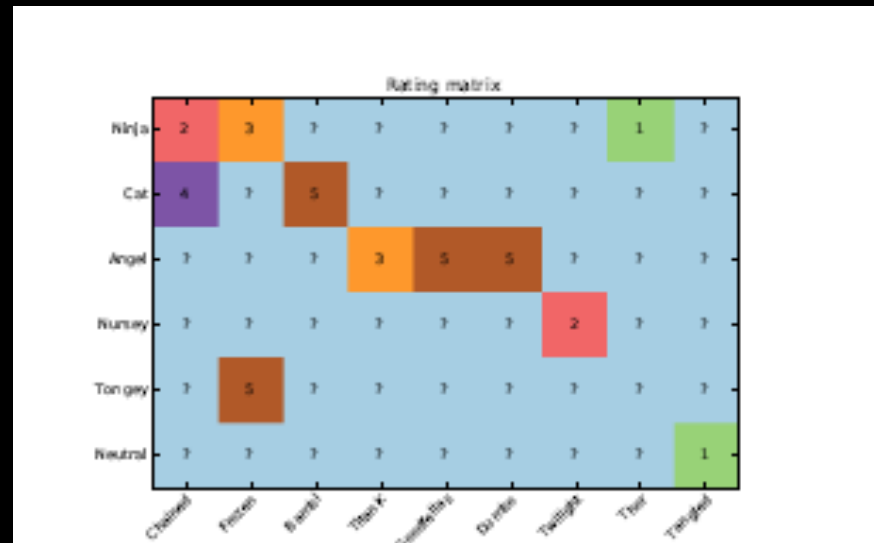
Convergence in Opinion: Many recommender systems make recommendations for users based on what other people who make similar choices have chosen.

Discussion Question

- In your opinion, if the following apps used collaborative filtering, for which one of them would that produce the worst overall consequences?

1. Duolingo Language Practice Sets
2. Tinder
3. Netflix
4. Facebook

Echo chamber: an environment where a person encounters only information or views that reflect and reinforce their own information or views.

They "may limit the exposure to diverse perspectives and favor the formation of groups of like-minded users framing and reinforcing a shared narrative." (Cinelli et al 2021)

Discussion Question

Does social convergence violate
anyone's rights?

The Facebook Papers

The Facebook Papers (2021)

In October 2021, a number of internal Facebook documents were made public by a whistleblower named Frances Haugen.

Many thought that these documents showed that Facebook was aware of many of the ethically dubious consequences of their social media platforms.

Weight Decision 12/15/2017

Component	Final Weight for 2018Q1
Like	1
Reaction, Reshare without Text	5
Non-sig Comment, Non-sig Reshare Non-sig Message, Rsvp	15
Significant Comment, Significant Reshare, Significant Message	30
Groups Multiplier (Non-friends)	0.5
Strangers Multiplier (non-friend-of-friend, small pages)	0.3

Wall Street Journal, "Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead"

- In deciding which posts to present to users, Facebook has an explicit formula describing the relative weights of certain factors.
- Facebook introduced this formula in order to drive more meaningful interactions.
- "The goal of the algorithm change was to reverse the decline in comments, and other forms of engagement, and to   . It would reward posts that garnered

, the documents show."

Discussion Question

What are the likely short-term consequences of Facebook's design decision on its users?

Discussion Question

What are the likely long-term consequences on its users?

"While the FB platform offers people the opportunity to connect, share and engage, an

, often before we can catch it and mitigate its effects," he wrote. "Political operatives and publishers tell us that they rely more on for distribution due to recent algorithmic changes that favor reshares."
(Internal Facebook Memo, quoted by the *Wall Street Journal*)

23

Discussion Question

Did Facebook's design decision
violate anyone's rights?

Some people would say that we have a right to autonomy – a right to make decisions about our lives ourselves, without manipulation by others.

Worries about autonomy have led some commentators to observe that users should have more control what they see in their feeds (Stray, "Beyond Engagement")

E.g. 'see less often' or 'hide post' functions in feeds

Discussion Question

What sort of personal controls would you want to have over your feeds in the social media platforms you use?

Group Exercise (Part 2)

- Evaluate the ethics of your impacts of your suggestions from the initial group exercise.

In this module, you have learnt:

- That recommender systems are powerful and can be valuable to individuals and society.

- To unpack the positive and negative consequences of your technical decisions

- To unpack how your technical decisions may impact people's rights

- If you have questions or thoughts, I'm happy to chat more – steven.coyne@mail.utoronto.ca

If you're interested in further exploring Ethics and Computer Science, check out these other resources at the University of Toronto.

Talks and events at:

- Schwartz Reisman Institute

- Centre for Ethics

Courses :

- Upper-year course in Computer Science: CSC 300 Computer Science and Society

- Philosophy (especially PHL256 Philosophy in the Age of the Internet, PHL271 Law and Morality, PHL275 Introduction to Ethics, PHL295 Business Ethics, PHL342 Minds and Machines, PHL377 Ethical Issues and Big Data)

- History and Philosophy of Science and Technology (especially HPS255 History and Philosophy of Artificial intelligence)

And stay tuned for more Embedded Ethics modules!

Acknowledgements

This module was created as part of an Embedded Ethics Education Initiative (E3I), a joint project between the Department of Computer Science[1] and the Schwartz Reisman Institute for Technology and Society[2], University of Toronto.

Instructional Team:
Roger Grosse, Steven Coyne, Emma McClure, Rahul Krishnan, Michael Zhang, Chandra Gummaluru

Faculty Advisors:
Diane Horton[1], David Liu[1], and Sheila McIlraith[1,2]

Department of Computer Science
Schwartz Reisman Institute for Technology and Society
University of Toronto

Computer Science
UNIVERSITY OF TORONTO

UNIVERSITY OF TORONTO

SCHWARTZ REISMAN INSTITUTE
FOR TECHNOLOGY AND SOCIETY

References

- Stray, Jonathan. "Beyond Engagement" Accessed online: https://partnershiponai.org/beyond-engagement-aligning-algorithmic-recommendations-with-prosocial-goals/

- Su, Xiaoyuan, and Taghi M. Khoshgoftaar. "A survey of collaborative filtering techniques." Advances in artificial intelligence 2009 (2009).