AI Ethics Workshop

Agenda

1. Setting the scene (15-20 minutes)
2. Workshop activity - Sweet Summer Child Score (45 minutes)
3. Reflection and discussion (15 minutes)

AI Ethics

Isn’t really a thing…

AI Ethics: Seven Traps

https://freedom-to-tinker.com/2019/03/25/ai-ethics-seven-traps/

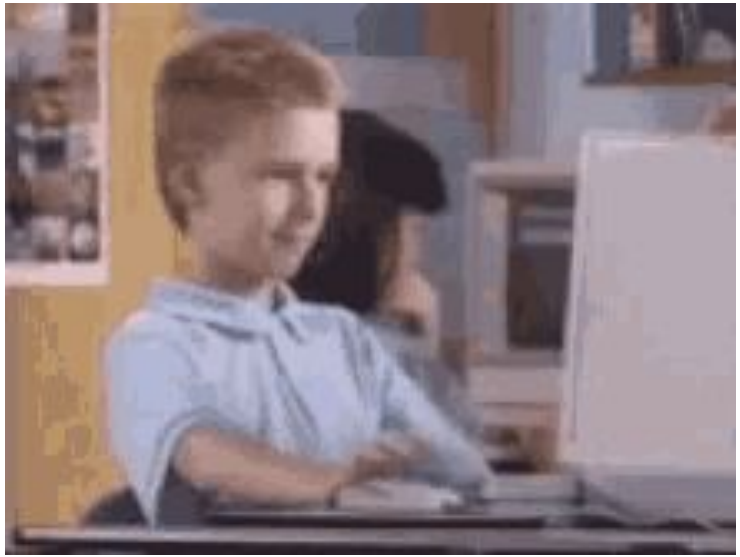
The reductionism trap

“Doing the morally right thing is essentially the same as acting in a fair way. (or: transparent, or egalitarian, or <substitute any other value>). So ethics is the same as fairness (or transparency, or equality, etc.). If we’re being fair, then we’re being ethical.”

That doesn’t mean we’re done here tho 😅

So you wanna build some tech

And you want it not to propagate existing inequalities, social bias and discrimination?



What are the jobs to do?

1. Name and formalise our personal ideologies
2. Externalise and discuss our moral intuitions with our team mates (get better at holding space for difference)
3. Capability building and advocating - in your team or with a client
4. Scan the ecosystem, determine your system’s context
5. Map the risks of your socio-technical system
6. Rank & prioritise those risks
7. Negotiate implementation trade-offs
8. Develop strategies to be alerted to failure cases
9. Develop strategies to mitigate the risk of failure cases

What kinds of tools are available?

This is a good news story - there are SO MANY more different kinds of tools, worksheets, and libraries available.

Now have the challenge of deciding which one to use 😅

Ethical frameworks

Where do you sit?

- Virtue Ethics
- Deontology (Kant)
- Consequentialism / Utilitarianism
- Ethical Pluralism
- And many more…

Practicing ethics

Getting your moral intuitions out of your gut, practice discussing with friends

- Ethics Litmus Tests
- ??

Design speculation

Imagining and mapping possible harms

- Black Mirror Brainstorm
- Ethics Explorer Toolkit - OMIDYAR Network
- IDEO AI Ethics Cards
- Consequence Scanning by doteveryone

Data Governance

Developing strategies and rules for appropriate data use

- ODI - Data Ethics Canvas
- DEON checklist
- Data Governance guidelines from UNSW

ML Explainability libraries

Peek inside the black box

- LIME
- SHAP
- Shapash
- Towards Data Science
  ^ good explainer & intro

ML Fairness metrics

“One rule, to rule them all”

- Counterfactual fairness
- Group fairness / statistical parity
- Equal false positive rate
- Equal false negative rate

Model monitoring

Detecting model drift

- Kosa.ai
- All the big cloud providers

HAX

Human AI Interaction

- https://www.microsoft.com/en-us/haxtoolkit/
- https://pair.withgoogle.com/guidebook

Want more?

List of links — tools, activities, reading, free courses (from me)
https://github.com/summerscope/mapping-fair-ml

Fair ML Reading Group (run by me, newbies welcome!)
https://github.com/summerscope/fair-ml-reading-group > paper history

https://groups.io/g/fair-ml > subscribe to listserv

AI Ethics Bookclub - great reading list

https://docs.google.com/spreadsheets/d/1Ex8MFdb6tjmYXKXlpHmNtO3tJ35aOul0lCXiWBU38n4/edit?usp=sharing

Want more?

Practical Data Ethics from Fast AI
https://ethics.fast.ai/

Intro to AI Ethics from Kaggle
https://www.kaggle.com/learn/intro-to-ai-ethics

Responsible Tech Playbook from Thoughtworks
https://www.thoughtworks.com/content/dam/thoughtworks/documents/e-book/tw_ebook_responsible_tech_playbook_2021.pdf

On my bookshelf

Questions?

Activity time


Sweet Summer Child Score

What we'll do now

- Explain activity for Sweet Summer Child Score
- Answer any questions before we start
- Break up into groups of 2 or 3, in breakout rooms
- Work through the quiz
- Calculate your Score
- Come back to the main zoom - each group present score & feedback
- Reflection (5 minutes)
- Final discussion & questions

Sweet Summer Child Score

1. https://github.com/summerscope/summerchildpy - options are python (bash), colab notebook, R app, PDF
2. Answer each question - mark the answer your group selects
3. At the end, you calculate the score by getting the points assigned to your answers, adding them all together, and multiplying by the multiplier (first question)
4. What are your recommendations? If you had to improve the system, what would you choose?

Robodebt scenario

- Time travel back to 2015
- You’re working with a mid-level government employee, working on digital innovations for Centrelink
- They send you a proposal on **“Improving the efficiency of recovering Centerlink overpayments”**, and ask you for your feedback
- What do you tell them?

Notes on scenario

- Try to reverse-engineer the specifications of the unbuilt solution
- Where you don't know, make your best guess
- For example, you can discover roughly how many people would be processed through this system (total cohort = all folks who have been on centrelink income support - not pension)
- You'd **do know** the design implementation (debt notice, web portal, etc)
- You **don't** know the consequences
- https://en.wikipedia.org/wiki/Robodebt_scheme
- https://www.crikey.com.au/2020/06/03/what-is-robodebt-what-happens-if-you-are-overpaid-by-centrelink/

Sharing our scores!

Group 1:

Group 2:

Group 3:

Group discussion

Reflection

Take 2-3 minutes to jot down some notes and responses to the session:

- Have you experienced scenarios at work where you had a bad gut feeling that you struggled to explain?
- What skills or capabilities would you like to develop for yourself?
- What are you interested to learn more about?
- What questions did this session raise for you?

Sweet Summer Child Score feedback

If you have any questions or feedback on this quiz, I'd love to hear from you! Please email me directly at

summerscope@gmail.com

PRs or suggestions on the repos are welcome, too 😍

Thanks for your time ✨

Laura Summers
laura@debias.ai
www.debias.ai
www.twitter.com/summerscope