Disclaimer

- This talk will be recorded. If you want to remain anonymous, please join the meeting from an incognito window and keep your video off.
- Clarifications can be asked for right away. There is a QA session at the end of the talk for questions that are more technical or philosophical.

Fairness

Alan Chan

High-level goal

Be more critical of “fairness” claims

What's going on?

FAccT 2021
Toronto, Canada

Google's solution to accidental algorithmic racism: ban gorillas

Google's 'immediate action' over AI labelling of black people as gorillas was simply to block the word, along with chimpanzee and monkey, reports suggest

▲ A silverback high mountain gorilla, which you'll no longer be able to label satisfactorily on Google Photos. Photograph: Thomas Mukoya/Reuters

ProPublica

How We Analyzed the COMPAS Recidivism Algorithm

Through a public records request, ProPublica obtained two years worth of COMPAS scores from the Broward County Sheriff's Office in Florida.

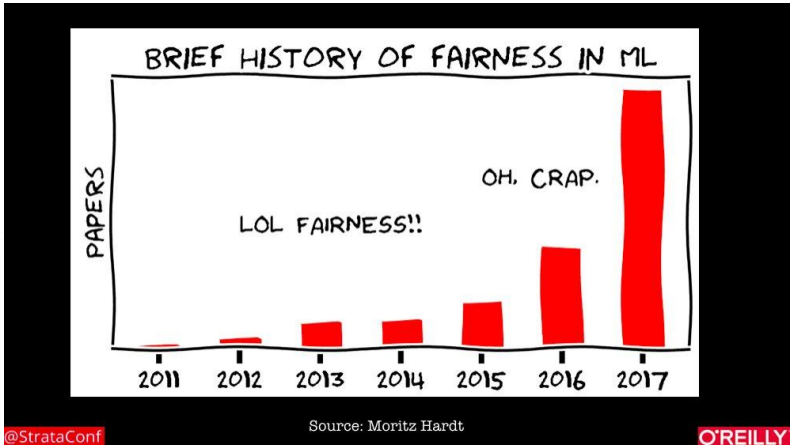
May 23, 2016

HireVue

Fairness and machine learning

Limitations and Opportunities

Solon Barocas, Moritz Hardt, Arvind Narayanan



My position

1. ML can’t (shouldn’t) solve all problems
2. ML can make things worse
3. The harms might not always outweigh the benefits

What is fairness in ML?

- Decision **D**
- Protected attributes **A**
- Other features **X**

$$\Pr(D = 1 \mid X = x, A = a) = \Pr(D = 1 \mid X = x)$$

Examples

Three Aspects of Fair ML

1. Supervised learning
2. Ignoring protected attributes
3. Power is omitted

What else is missing?

Scenario 1

- Racialized groups are underrepresented in positions of power
- More representative hiring benefits the person hired
- More diverse representation has downstream effects: e.g., role model effect, better policies

Fairness is long-term

Scenario 2

- Affirmative action can help redress and prevent harms
- AA explicitly uses protected attributes

Possible objection

But that's discriminatory!

Substantive Equality

Canadian Charter of Rights and Freedoms

Equality Rights

Equality before and under law and equal protection and benefit of law

15. (1) Every individual is equal before and under the law and has the right to the equal protection and equal benefit of the law without discrimination and, in particular, without discrimination based on race, national or ethnic origin, colour, religion, sex, age or mental or physical disability.

Affirmative action programs

(2) Subsection (1) does not preclude any law, program or activity that has as its object the amelioration of conditions of disadvantaged individuals or groups including those that are disadvantaged because of race, national or ethnic origin, colour, religion, sex, age or mental or physical disability. (84)

Fairness isn’t just formal equality

Scenario 3

- A company develops completely “fair” facial recognition tech
- Who chooses that definition of fairness?

Fairness involves power

Summary

- Long-term consequences
- Substantive equality
- Who has power?

Closing questions

- How do we design systems that take into account broader contexts and distributions of power?
- What role can reinforcement learning play in fair, dynamic interventions?
- Who defines objective functions, and how should they be defined?

Works

https://arxiv.org/abs/1808.00023

https://www.cs.cornell.edu/~red/fairness_equality_power.pdf

https://dl.acm.org/doi/abs/10.1145/3351095.3372878

https://arxiv.org/abs/2001.09773

https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199656967.001.0001/acprof-9780199656967

http://proceedings.mlr.press/v81/binns18a

https://www.philosophy.rutgers.edu/joomlatools-files/docman-files/4ElizabethAnderson.pdf