

ZEERAK TALAT | ITU | 05/03/2019

DATA SCIENCE, NATURAL LANGUAGE PROCESSING, AND UNSINKABLE SHIPS

WE NEED TO TALK...

1. Who even needs ethics?
2. How do you even think about ethics?
3. Does this even relate to real life?
4. Can we even ethics in real life?



**“GOD HIMSELF COULD NOT
SINK THIS SHIP!”**

Unnamed deckhand on the Titanic [Contested]

ROSE BE LIKE I NOMINATE JACK



FOR THE ICE BUCKET CHALLENGE



TAKEAWAYS FROM TODAY

- We all have blindspots
- Making mistakes is human
 - But Twitter *will* drag you
- Find out which questions to ask yourself.
E.g.
 - What might I be missing?
 - What might be hard for me to fully grasp?
- How can it be helpful to think about my task/problem?



*“What’s he searching for? - The ride!
The ultimate ride.”*

- Lori Petty, “Point Break”

2 MINUTE TASK

1. *Go to Google Images and search for "Doctor". What do you see?*

2. *Search for "Nurse".*

What are the differences between the two results?

3. *Think about, and try to find other pairs of queries that generate disparities in results.*

Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms

The controversial tech used to detect problems before they happen

**FACEBOOK'S HATE SPEECH
POLICIES CENSOR
MARGINALIZED USERS**

**Why we need to solve the issue of
gender bias before AI makes it worse**

**Predictive policing is a scam that
perpetuates systemic bias**



CONSEQUENTIALISM

"I have a duty to speak the truth as I see it and share not just my triumphs, not just the things that felt good, but the pain, the intense, often unmitigated pain."


Audre Lorde



UTILITARIANISM

"I was determined to play my horn against all odds, and I had to sacrifice a whole lot of pleasure to do so."

- Louis Armstrong



What did I see? Well that brings up an interesting philosophical question:

Is it okay to snitch to the police...

...on the police?

CONSEQUENTIALISM & UTILITARIANISM

Discussion

What are the consequences of a company choosing to address disparities in their models? What are the consequences of not doing so?

Is a system which only works for a single group ethical?

Is it ethical to make a system work better for one group, if that means that it doesn't work as well for another?



THE TROLLEY PROBLEM



VIRTUE ETHICS

“Everybody is a hero in their own story if you just look.”

Maeve Binchy



VIRTUE ETHICS

Discussion

A) Estimated 20,000 tech workers walked out at Google in response to allegations of protection of sexual harassers.

B) 3,000 Google employees signed an open letter asking CEO to back out of a military project (Project Maven).

Are A and B comparable? What does the difference in numbers say about the ethics governing the two actions?



DEONTOLOGY

“On Wednesdays we wear Pink”

- Amanda Seyfried, “Mean Girls”

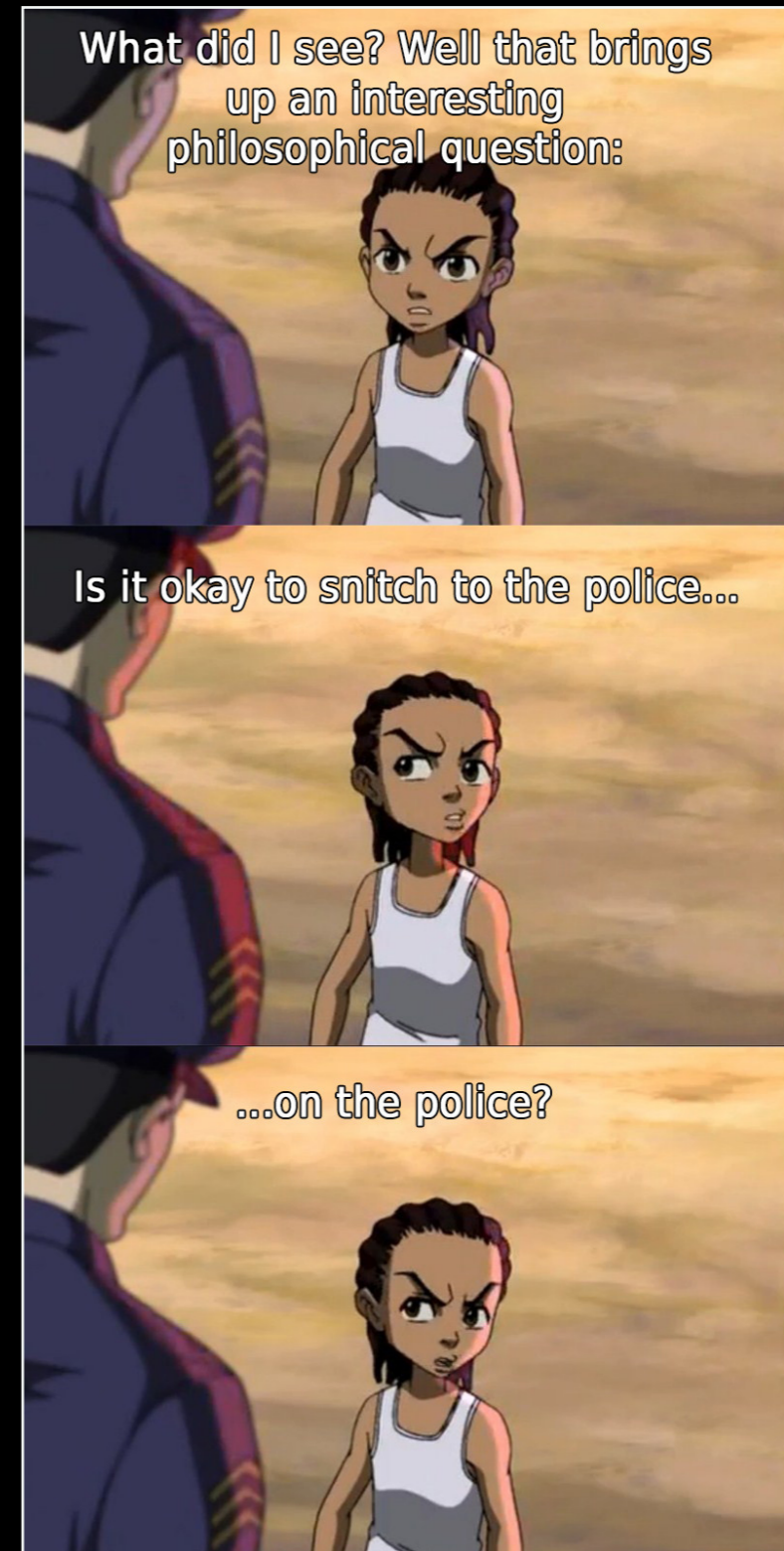


DEONTOLOGY

Discussion

Returning to the question of king Leopold II, would it be ethical to kill him at all? And what if it is after committing the crime of genocide?

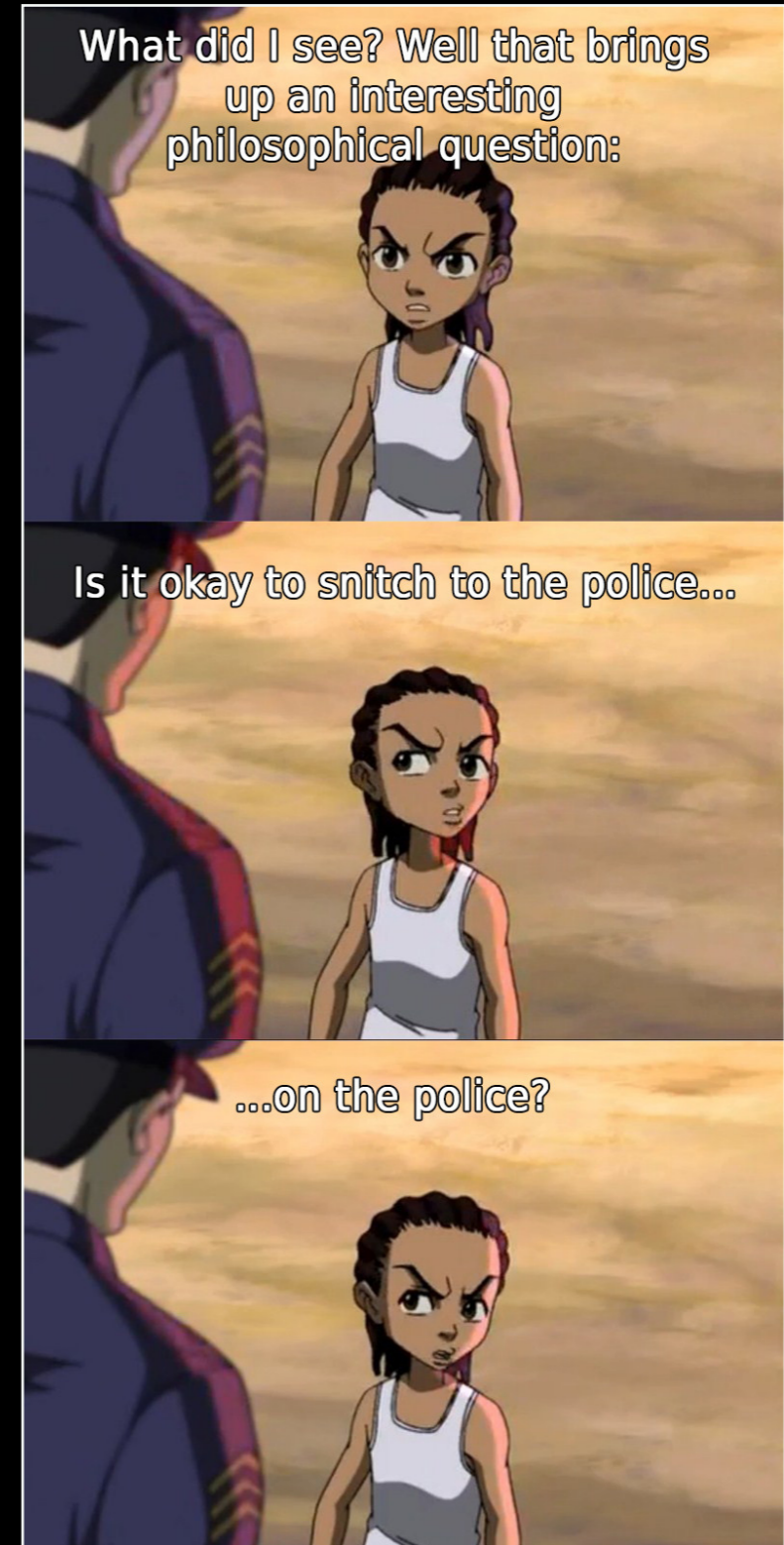
If we must never make classifiers that perform worse, is it ethical to make classifiers that are better improve classification for one group but not another?



SOCIAL CONTRACT THEORY

"We will never have true civilization until we have learned to recognize the rights of others."

Will Rogers



SOCIAL CONTRACT THEORY

Discussion

Is it moral for an organisation to decide whether discrimination is acceptable and what kind of discrimination is acceptable?



ANTI-DISCRIMINATION & ANTI-SUBORDINATION

“Treating different things the same can generate as much inequality as treating the same things differently”

Kimberlé Crenshaw



ETHICS IN MACHINE LEARNING

"The curse of imagination is picturing the world as it should be."

Morley



ETHICS IN MACHINE LEARNING

Equalised Odds - Accuracy on all groups should be the same.

$$Pr\{\hat{Y} = 1 | A = 0, Y = y\} = Pr\{Y = 1 | A = 1, Y = y\}, y \in \{0, 1\}$$

ETHICS IN MACHINE LEARNING

Demographic Parity - Decision is independent of protected attribute

$$Pr\{\hat{Y} = 1 | A = 0\} = Pr\{\hat{Y} = 1 | A = 1\}$$

ETHICS IN NLP

- ▶ Brand new field
- ▶ Questions are posed; few are answered
- ▶ Very limited number of groups considered



ORIGINAL POSITION

