

Computer Science Ethics



Guest lecturer: Becca Smith

Why is ethics important in
CS?

Goals for today

- Some intro to moral philosophy (woo!)
- Learn importance of ethics in CS
- Tackle some interesting cases
- Discuss AI ethics / algorithmic justice

Branches of Ethics:

1. Metaethics: what is morality?
2. Normative ethics: how do we decide what to do?
3. Applied ethics: what do we do now?

What is ethics? AKA moral philosophy

Definition: the study of what is morally right or wrong

- Often relies on assumption of moral objectivity
(there exists a right and a wrong, even if we do not know what it is)
- Yet some believe in moral relativism

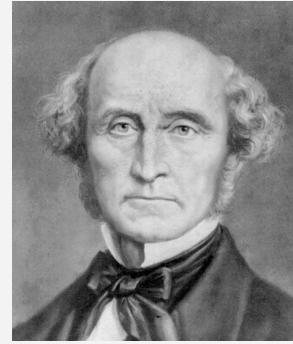
“It is not possible to produce a set of rules purporting to describe what a man should do in every conceivable set of circumstances.”

-Alan Turing

Ethical theories

- Many have attempted to create a unified account of our ethical obligations
 - Utilitarianism
 - Deontology
 - Virtue ethics

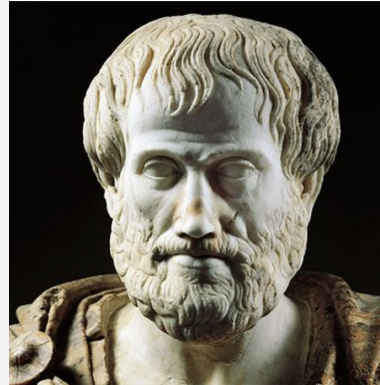
John Stuart Mill



Immanuel Kant

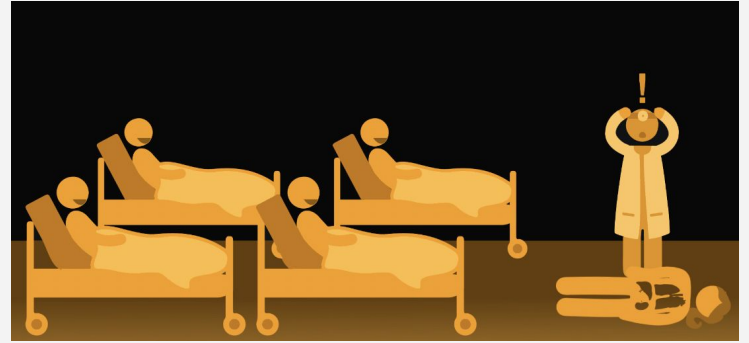


Aristotle



Utilitarianism

(a type of consequentialism, meaning an action is judged based on its consequences)



Ex: could you kill someone if their organs would save 10 people

The best moral action is the one that maximizes utility

- Everyone's utility is worth the same
- Utility = happiness? Money? Welfare? Something else??
- Expected or actual utility?

The most popular ideology for a while! Until...

Deontology

(Pretty much Kantianism;
he had a lot to say)

Moral actions should be judged by their intention, not outcome

- All humans should be treated as ends, not a means (big focus on human dignity)
- You should act as if your actions are universalizable
 - You should not do something that it would not make sense to do it all the time (lying, cheating)



Ex: Is it immoral to work at a big company that exploits workers?

De

WHEN SOMEONE SAYS TO CONSIDER THE CONSEQUENCES OF YOUR ACTION

tianism;
y)

Moral act outcome

- All (big)
- You

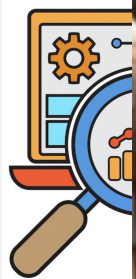
tion, not

a means

rsalizable

at it would

ime (lying,



"We don't do that here"

Virtue Ethics

comeback season

Morally good actions reflect morally good people

- We should act as virtuous people would act

Ok.. but how do virtuous people act?

- Aristotle liked traits of integrity, loyalty, wisdom, courage

ETHICS 101

Professor Chiwetel Ejiofor

SOCRATES
470 BCE - 399 BCE

Plato
428 BCE - 348 BCE

ARISTOTLE
Student of Plato
384 BCE - 322 BCE

- Founder of Western Philosophy - Dialogs
- Critical Thinking

INDIVIDUAL PERSON

**The
Good
Place**

What does this have to do with CIS?

Bringing the social and ethical responsibilities of computing to the forefront

The inaugural SERC Symposium convened experts from multiple disciplines to explore the challenges and opportunities that arise with the broad applicability of computing in many aspects of society.

Terri Park | MIT Schwarzman College of Computing

June 8, 2023

Cambridge Launches AI Research Ethics Policy

March 13, 2023 08:01 PM Eastern Daylight Time

Education

Teaching Responsible Computer Science

Scholars from around the country explore the best approaches to embedding ethics into CS curricula at a recent Stanford event.

Mar 28, 2023 | Nikki Goth Itoi [Twitter](#) [Facebook](#) [YouTube](#) [LinkedIn](#) [Instagram](#)

Global Forum on the Ethics of Artificial Intelligence 2024

The 2nd Global Forum on the Ethics of AI: Changing the Landscape of AI Governance will be organized by Slovenia, under the patronage of UNESCO, on 5 and 6 February 2024.

 HARVARD UNIVERSITY

Embedded EthiCS™ @ Harvard Bringing ethical reasoning into the computer science curriculum

What does this have to do with CIS?

Because CS is everywhere, and affects people!
It has the power to:

- Replace jobs
- Make decisions more (or less) fairly
- Use our data
- Be artificially intelligent
- Change social interactions
- Save lives
- Impact the environment
- And so much more!

Is illegal hacking to expose government corruption morally permissible?

A Kantian might say: illegal hacking cannot be universalized, so no

A utilitarian might say: the consequence helps promote accountability in government and could stop the corruption, so yes

A virtue ethicist might say: a person who hacks illegally does not have integrity, or maybe a person who fights government corruption is heroic

What do you say?

Is programming a potentially addictive game morally wrong?

A Kantian might say: you might be taking advantage of addictive behaviors for profit, so no

A utilitarian might say: it depends on if it leads to addiction, if it makes more people addicted then yes, if not then no

A virtue ethicist might say: there might be a virtue to scientific exploration, but also compassion

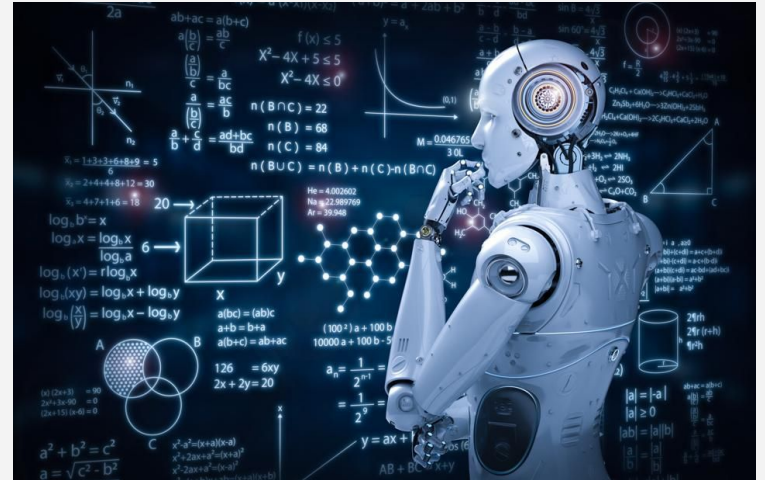
What do you say?

AI Ethics

Machine Learning

Using existing data to make predictions

- You saw this already in large language models
- Language data used to predict next words



More applications

More applications

- **Hiring**

More applications

- **Hiring**
- **College admissions (~50%)**

More applications

- **Hiring**
- **College admissions (~50%)**
- **Prison risk assessment**

How does this work?

1) Train model on data

Class President	GPA: 3.0/4.0	Debate Team	...
+0.9	-0.2	+0.8	...

How does this work?

2) Evaluate / tune model

Class President	GPA: 3.0/4.0	Debate Team	...
+0.9	-0.2	+0.8	...

Are these fair scores? Should we adjust?

How does this work?

3) Use on new data

Applicant A:

GPA: 3.8/4.0	+0.9
Model UN President	+0.8
Varsity Basketball Captain	+0.6

Total = 2.3
(say threshold = 2)
 $2.3 > 2$

Admit applicant!

A problem

A problem

If our model is trained on biased data, it will continue to produce biased data.

Some stats

- 89% of children from wealthy families attend college compared to 51% of children from low-income families
- Over 88% of people in executive leadership positions at top Fortune 500 companies are white
- Black Americans are incarcerated at more than 5 times the rate of white Americans

What happens?

- With objective machine learning algorithms and biased data, the model will give biased points to demographic features

Race: white	++
----------------	----

“female”	--
----------	----

What happens?

A ProPublica study analyzed the chance of false positive risk assessments across black and white defendants

These contingency tables reveal that the algorithm is more likely to misclassify a black defendant as higher risk than a white defendant. Black defendants who do not recidivate were nearly twice as likely to be classified by COMPAS as higher risk compared to their white counterparts (45 percent vs. 23 percent).

Better alternatives?

What if we remove demographic data?

Better alternatives?

What if we remove demographic data?

- There are too many “proxies”
- High school attended, hometown, and activities can often serve as proxies for race or socioeconomic status
- It wouldn't make sense get rid of all the context, then we lose accuracy

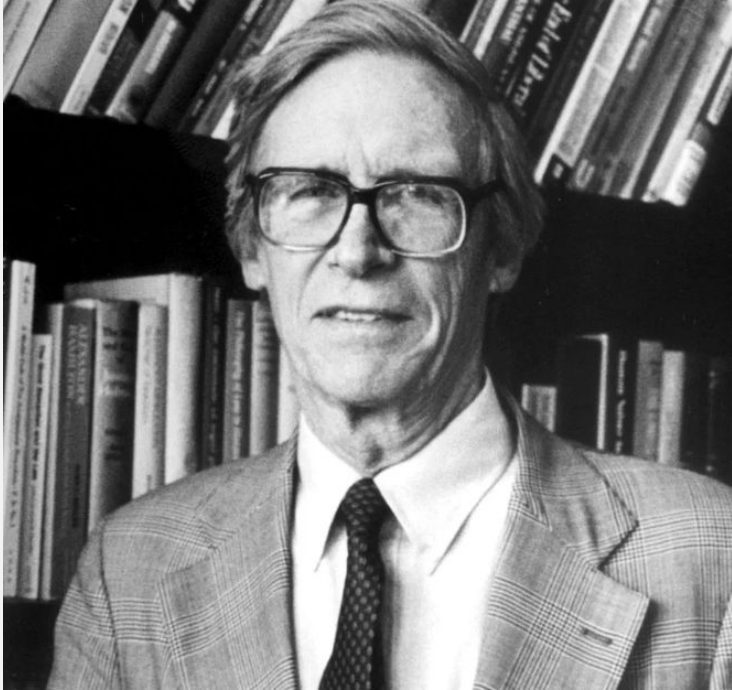
What's the solution?

- Some definition, benchmark, or method to deem models “fair”
- “Algorithmic fairness” is a hot area of research!

What is fairness / justice?

- Rawlsian
- Egalitarian
- Libertarianism
- Something else??

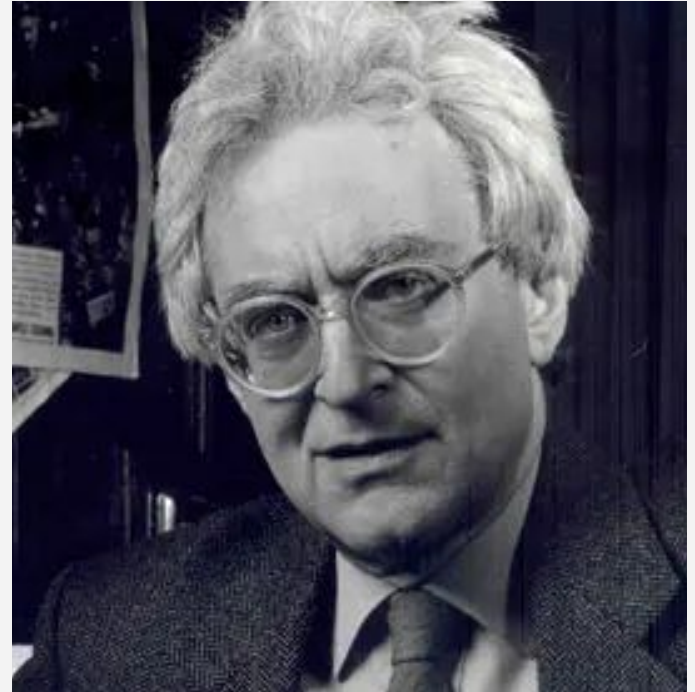
Rawlsian Fairness / Justice as Fairness



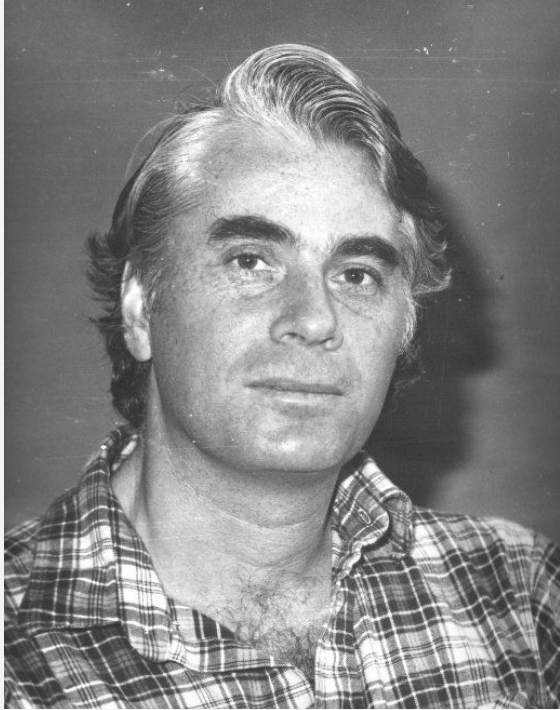
- We would agree to principles of justice behind a “veil of ignorance”
- This would lead to:
 - Equality of opportunity
 - Difference principle: the most fair outcome benefits the position of the least well off

Egalitarianism

- Everyone deserves equal resources regardless of starting point
- Redistribution necessary
- *Why Not Socialism?*
by G.A. Cohen



Libertarianism



- Forced patterns distributions cannot happen without some invasion of individual liberty

And there are more!



- Capabilities approach
- Racial contract
- Global justice
- etc.

In other words, these kinds of more qualitative decisions and judgments—which type of fairness notion to use, when the reduction in accuracy is worth the gain in fairness, and many others— must remain firmly in the domain of human decision-making. The science part can begin only once society makes these difficult choices. Science can shed light on the pros and cons of different definitions, as we'll see, but it can't decide on right and wrong.

Kearns, Michael, and Aaron Roth. *The Ethical Algorithm : The Science of Socially Aware Algorithm Design*, Oxford University Press, Incorporated, 2019. *ProQuest Ebook Central*, <http://ebookcentral.proquest.com/lib/upenn-ebooks/detail.action?docID=5905172>.

QUESTIONS?

(citations in speaker notes)