Computer Science Ethics

Guest lecturers: Becca Smith & Bhrajit Thakur
Pollev
(Why is ethics important in CS)
Goals for today

- Some intro to moral philosophy (woo!)
- Learn importance of ethics in CIS
- Tackle some interesting cases
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
Utilitarianism
(a type of consequentialism, meaning an action is judged based on its consequences)

The best moral action is the one that maximizes utility
- Everyone’s utility is worth the same
- Expected or actual utility?

The most popular ideology for a while! Until…

Ex: could you kill someone if their organs would save 10 people
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 selling data immoral, since it uses humans as a means to help companies make predictions?
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
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

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.

Cambridge Launches AI Research Ethics Policy

March 13, 2023 08:01 PM Eastern Daylight Time

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
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? [Poll ev]
You’d be surprised by how many people are collecting data on you…

Social Media

Government

Websites

Google / Emails

Your friends

Amazon

Ed Discussion

How does that make you feel?
Data Collection

Problems with collecting data: breaking anonymity & leaks
1990s a MA Insurance Commission released hospital visit summary of state employees
- To protect patient information, they removed obvious identifiers like name, address, and SSN, and left others like zip-code, birthdate, and sex
- Do you think this was enough to protect anonymity? NO

2006 Netflix prize competition:
- Wanted to improve their recommendation algorithm
- Released data of 100m ratings of ~500,000 users (anonymized → only identifiable by their unique numeric ID)
- PhD student Arvind Narayan connected public IMDB profiles with the Netflix dataset, exposing thousands of profiles
Leaks

Barking up the wrong data tree: even pets aren’t safe from a data breach

ChatGPT can leak training data, violate privacy, says Google's DeepMind
What is AI?
AKA artificial intelligence

Definition: “a field, which combines computer science and robust datasets, to enable problem-solving” (IBM)

- Algorithms you’ve programmed in this course have been written by **you**. You could **always** reason why there was a certain output given an output
- A programmer using AI may derive the **process** of coming up with an AI model (e.g. how to find patterns in data) but the model itself isn’t derived by the programmer (Ethical Algorithm - Kearns & Roth)
Ex: Let’s say you want to write a program where users can take a personality quiz to find their personality type...

**Human Intelligence:**

(Step 0*: Research personality quizzes but you probably remember them)
Step 1: Write personality categories & questions
Step 2: Assume correlations between answers & personality types
Step 3: Aggregate user answers & compare user answers to predefined assumptions

**Artificial Intelligence:**

Step 0: Research existing personality quizzes (categories, types of Qs, & scoring)
Step 1: Write personality categories using natural language processing
Step 2: Assume correlations between answers & personality types from patterns
Step 3: Aggregate user answers & compare with predefined assumptions using malleable neural network
Step 4: Learn continually from user responses & improve accuracy of model
Consequences of AI design:

Because of how it’s designed, AI can often codify biases through the programmer malintent, error, or underlying bias in the dataset.

When a car breaks down, you can often find tell-tale signs of how it broke down and potentially who broke it. When AI fails…

1. Hard to tell how it failed
2. Hard to tell who is accountable
3. Hard to tell how to fix it
Biased Applications of AI

- **Image generation & search results**
  - Output reductionist stereotypes

- **Targeted ads:**
  - LGBTQ+ targeted ads: LGBTQ+ people described them as stereotyping and tokenizing

4.3.2 *Queer people want more transparency and autonomy.*

>[It’s scary] how much we’re already monitored, and how much [ad systems] already know about me. […] [I would prefer] increasing the amount of diversity within advertisements, without necessarily needing to know for sure if somebody identifies a certain way. (P15)
Biased Applications of AI cont.

- **Hiring Decisions:**
  - AI tools trained on data that itself is reflective of existing institutional & systemic biases.

- **Police Sentencing:**
  - Risk assessments: used to inform bail amounts or levels of freedom
  - ProPublica found Northpointe’s “risk assessment” discriminated against black defendants

### Prediction Fails Differently for Black Defendants

<table>
<thead>
<tr>
<th></th>
<th>WHITE</th>
<th>AFRICAN AMERICAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeled Higher Risk, But Didn't Re-Offend</td>
<td>23.5%</td>
<td>44.9%</td>
</tr>
<tr>
<td>Labeled Lower Risk, Yet Did Re-Offend</td>
<td>47.7%</td>
<td>28.0%</td>
</tr>
</tbody>
</table>
Are there positive uses of Data collection/AI?

Yes!

1. Climate Change AI
2. Combating World Hunger
3. Identifying fake news
4. Assessing medical imagery
5. Create tools for people with disabilities

But does that make the use of AI **ethical**?
Solutions

● change the ways in which these models are created to prioritize equality
  Google AI loans visualization
● diversify datasets
● regulation & enforcement against AI discrimination & for data protection
  ○ combatting “ethics washing” (a.k.a. solely nominal actions to enforce ethics)
● auditing algorithms
Is collecting & analyzing data to construct AI models morally permissible?

A Kantian might say: no because it does not respect the dignity and autonomy of individuals and uses them merely as tools/data for achieving an objective (consent is key)

A utilitarian might say: yes if these activities lead to the greatest overall happiness or utility (such as advancing medical research, improving safety, etc.) even if it can create some inequalities, they could be considered morally permissible.

A virtue ethicist might say: yes if the model was created with good intentions (honest, noble, integritous)  

What do you say? [Poll ev]
Thank you all! (and ty harry for letting us take over cis 1100 for the day)

If you’re interested in getting involved with AI fairness at penn, feel free to email lena318@seas.upenn.edu (she’s going to be a PhD at harvard in AI fairness next fall, and she was a cis 1100 head ta o-o)

If you’re interested in classes, CIS 4230 - ethical algorithm design & CIS 4210 - human computer interaction (both professors whose research was cited today & doing super cool research) (all citations in the speaker notes)

If you’d like to follow up with us, Becca and Bhrajit’s info is on the staff website :))