Robots and Morality

2/24
“We can’t make machines liable for actions,” said Zeerak Talat, an A.I. and ethics researcher at Simon Fraser University in British Columbia. “They are not unguided. There are always people directing them and using them.”

Are developers qualified to make moral robots?
Robots & Morality

Is there anyone you look up to morally or seek for moral guidance?

Do your family members or friends have the same moral values as you?
MIT Moral Machines
Would you want to make the decision for the trolley problem? Or leave it to your car to decide?

You are in a self driving car that has sudden brake failure. What option should the car decide? Continue ahead and hit 3 children or swerve to the next lane and hit 5 adults? Which would you choose?
The Principle of “Meaningful Human Control”

Principle of meaningful human control future weapons systems must preserve meaningful human control over the use of (lethal) force, that is: humans not computers and their algorithms should ultimately remain in control of, and thus morally responsible for relevant decisions about (lethal) military operations.
The Principle of “Meaningful Human Control”

1. Human operators are making informed, conscious decisions about the use of weapons.

2. Human operators have sufficient information to ensure the lawfulness of the action they are taking, given what they know about the target, the weapon, and the context for action.

3. The weapon is designed and tested, and human operators are properly trained, to ensure effective control over the use of the weapon.

Follow-up Question

Should humans always be the decider for life and death situations?

Should robots always be subordinate to human judgment and control?
I, Robot: Robot saves Spooner (Will Smith) instead of child
Deciding who to provide loans to?

- Banks struggle to get an accurate picture of a small business’s viability and creditworthiness
- Less bias?
- Conventional loans take an average of 51 days to close – 49 days on average for a purchase transaction and 51 days for a refinance
- Do the pros outweigh the cons?
  - AI bias.
  - Explainability and ethics.
Pepper at HSBC in Beverly Hills

- “it's about time the banking customer experience be just as fun and engaging”
- Jeremy Balkin, Head of Innovation for HSBC
- In Beverly Hills, where Pepper will greet customers and answer questions to "provide a memorable banking experience"
Perception of morality of a robot is affected by appearance
Robosurgeons will need to understand why the pianist might prefer the loss of her life to the loss of her hand.

The problem with Asimov’s laws is that they must be interpreted. The first rule talks about ‘harm’, but what does ‘harm’ mean? Imagine, for instance, a pianist whose dominant hand is suffering gangrene. If her hand is not amputated, she will die. But she swears she does not wish to live without her ability to play. Yet she is feverish as she says this; maybe she doesn’t really mean it.

What should a robot surgeon do with this patient? Cause the harm of removing her hand despite her protestations? Or allow the harm of her death from gangrene?

Biases because training set

Is there a certain level of bias we will accept from robots?

Can humans be considered to be completely unbiased? Is this the expectation we should have for robots?
Should humans even be the deciding factor? We’re flawed too

- Robots consider more courses of action
- Robots are capable of following rules of engagement to the letter
- Humans are more inconsistent
Too sleepy: almost like drunk driving

Should robot prevent passenger from getting in car? Should there be an override?

What’s Worse: Sleeping Drivers or Self-Driving Cars?
Emergency resource allocation and distribution

- Current problems
  - PTSD
  - Mental and emotional burnout
  - Low pay
  - High turnover
  - Undervalued
  - Lack of training provided to support behavioral health crises
Robot judges?

- The issue of judicial and implicit bias
- Would robot judges make decisions more accurately than humans?
- Streamlines service
Robot lawyers already exist

The DoNotPay app is a “personal attorney” that helps people with issues ranging from a contested parking ticket to an unfair banking fee.
A killer with no remorse?

- Robots aren’t sentient… yet
- Do all humans feel remorse for taking another human life?
- Serial killers lack empathy
- Psychopaths tend to experience no fear of punishment and a lack of emotional sensitivity and reactivity
So…

- Do you think robots can make better decisions than humans?
- How can we build moral robots?
- Do you trust your own moral decisions? Can you “parent” or “teach” a robot?