Philosophical Crises

How Technology is Demanding Answers from Age-Old Questions

NIKHIL GARG

STANFORD SPLASH

4/9/2016

Ideas do not rule the world. But ***it is because the world has ideas*** *(and because it constantly produces them)* ***that it is not passively ruled by those who are its leaders*** *or those who would like to teach it, once and for all, what it must think.*

-- Michelle Foucault

Overview

1. Ethical and Moral Decision-making by Machines
2. Artificial Intelligence and Philosophy of Mind

Ethical and Moral Decision-making by Machines

Motivating questions

1. What morals and ethics should we teach machines?
2. How do we teach machines morals and ethics?

What morals should we teach machines?

Humans disagree about basic questions

Let’s demonstrate: Trolley Problem (Philippa Foot, 1967)

- Basic scenario
- Variation 1: Person’s about to fall anyway
- Variation 2: Falling person is evil, people on the tracks are children
- Variation 3: It’s your family on the ground
- Variation 4: You can jump in front of the tracks to save the others

Application in self-driving cars!

What should a self-driving car do in an unavoidable accident?

- Protect the people inside the car?
- Protect the most number of people?
- Protect the car itself?

Accident decision models *can* integrate these ethical decisions – should it?

- Who should make the decision?
- Should the vehicle's occupants be able to overrule the decision?

Other related questions and thought experiments

Predictive analytics in Criminal Justice

- If we can statistically determine where crime is likely to occur through machine learning techniques, what should we do with the predictions?
- How do we know whether the algorithm is discriminatory?

Genome Editing

- Should we edit out diseases in human embryos?
- Should we make non-medical changes?

Artificial Intelligence and Philosophy of Mind

Philosophy of Mind

How do you know others feel the same pain you do?

How would you describe what ‘red’ looks like?

How do you know others see the same red you do?

How do you know others experience anything?

How do you know that…

- we aren’t all mindless zombies?
- we aren’t all in your imagination?
- the world itself isn’t a simulation?

These are open questions: see Descartes, Leibniz, Wittgenstein, Kant, Heidegger, Strawson, Nagel, Chalmers, Kripke …

What does this have to do with technology?

How do you know your computer today isn’t conscious?

- We can describe its logic exactly
- It doesn’t act as if it has consciousness

What does it mean for a computer to act as if it had consciousness?

- Creativity, vulnerability, emotions
- Turing in 1950: Indistinguishable from a human

If a computer couldn’t be told apart from a human, does it have consciousness?

- Leibniz in 1714
- Searle in 1980

Why does this matter?

Human/Machine rights

- Should an indistinguishably ‘human’ machine be given voting rights?
- Is turning it off murder?
- And the same question for any ‘human’ right

More abstract notions

- What does it mean to be human?
- Should humans seek to protect our own species?

Conclusion

Conclusion

Many important, open problems at the intersection of technology and philosophy

- Some are openly discussed, others often swept under the rug
- Requires exposure in multiple disciplines – don’t narrow yourself too quickly!

Questions?

Nikhil Garg

nkgarg@stanford.edu

Slides available at:
gargnikhil.com/files/garg_splash2016.pdf