Ethical frameworks

Objectives

- Introduce the course and the learning format
- Identify a range of fields in which big data has changed social behavior
- Discuss areas in which the class wants to explore in more detail
- Define frameworks for thinking about ethical behavior
- Evaluate decisionmaking in the trolley car problem based on different ethical frameworks

Assigned readings

- Chapter 2 - Introduction to Ethics. In **Ethics for the Information Age** by Michael Quinn (7th edition).
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... & Rahwan, I. (2018). The moral machine experiment. Nature, 563(7729), 59-64.
  - Try out the Moral Machine
  - Ethical guidelines for driverless cars in Germany

Optional readings

Response paper prompt

The trolley problem is a classic thought experiment in ethics. The general form of the problem is:[1]

> There is a runaway trolley barreling down the railway tracks. Ahead, on the tracks, there are five people tied up and unable to move. The trolley is headed straight for them. You are standing some distance off in the train yard, next to a lever. If you pull this lever, the trolley will switch to a different set of tracks. However, you notice that there is one person on the side track. You have two options:
>
> 1. Do nothing and allow the trolley to kill the five people on the main track.
> 2. Pull the lever, diverting the trolley onto the side track where it will kill one person.
>
> Which is the more ethical option? Or, more simply: What is the right thing to do?

In reality, balancing the trade-offs of saving one set of lives over another is a true test of human decision making:

- On the lighter side, see Star Trek: The Wrath of Kahn
- On a more serious and timely note - rationing of healthcare during the COVID-19 pandemic

This problem is especially acute in the design of fully-autonomous cars which do not require any human input while the car is in motion. What should the vehicle do in the event of a potentially fatal collision which is unavoidable, but also an event where the software can make a number of decisions to influence the particular outcome? For example, should the software prioritize the lives of the passengers over the lives of pedestrians or passengers in another vehicle? If faced with the choice of killing a child versus killing an elderly person, how should the computer prioritize one life over another?

An important aspect of autonomous vehicles (and algorithms more generally) is that they are all designed by humans, and so human values and preferences play a crucial role in the design of artificial decision making systems. The Moral Machine is an online simulation of the types of decisions an autonomous vehicle may have to make, and allows the researchers to explore how actual humans would make decisions under

[1] Source: Wikipedia

these circumstances. One could then take the results of this experiment to inform how policymakers craft regulations to govern the design of the AI systems controlling autonomous vehicles.[2]

In an essay no longer than 1000 words, assume you are a policymaker charged with drafting a set of ethical rules for autonomous vehicles to obey in the event of a potentially fatal crash. What should be the priorities or assumptions of the algorithm? What factors should be weighted most heavily when designing this set of rules? Should they be universal, or do you expect the rules to vary by society? Be sure to assess your proposed rules using at least three of the ethical theories outlined in **Ethics for the Informational Age**.

[2]See the attached readings for an example from Germany of ethical rules for automated vehicles