Tuesday 7/5

Welcome to INFO4601! In this class, we will use collaboratively edited documents like this one to guide group discussions and to share information with each other.

Important link with more important links:
Class Canvas page

Survey (please fill this out today!)
https://forms.gle/T5niKnNVaSyuxoYD8

Announcements:

In the News:

Agenda:

Do not edit above this line.

Questions or Things to Share?
-

LIGHTNING INTROS

Write your **first name** (as you want to be called), your **major** (e.g. CS undergrad), and **what you're most excited to talk about** in this class or **something about tech you've heard about in the news lately**.

BRAINSTORMING

What other topic(s) are you interested in talking about in this class? What do you want to make sure that we cover? You can start with the obvious, but then try to go beyond that!

GROUP DISCUSSION

Discuss the prompt: When is it ok to murder someone? What makes a person bad enough to die? How do we know that is "bad"? **Make sure you take notes & write down your name next to your group - this is how I will take attendence**!!!

READING/WATCHING FOR TONIGHT

Crash Course Videos:
- [Utilitarianism](#)
- [Deontology](#)
- [Virtue Ethics](#)

Things to think about:
- What are the key characteristics of each moral framework?
- Which moral framework(s) do you see operationalized most often in technology development?
- Which moral framework resonates most with you? Why do you think it resonates with you?

Wednesday 7/6

Announcements:

In the News:

Agenda:

Additional Resources
- "[Act and Rule Utilitarianism](#)," *Internet Encyclopedia of Philosophy*.
- "[Distributive Justice](#)," *Stanford Encyclopedia of Philosophy.*
- "[Ethics of Care](#)," *Wikipedia*.
- "[Virtue Ethics](#)," *Stanford Encyclopedia of Philosophy.*
- "[Should Batman Kill the Joker? Perspectives from Five Famous Philosophers](#)," *Comics Alliance*, 2012.
- Garg, Vaibhav, and L. Jean Camp. [Gandhigiri in cyberspace: a novel approach to information](#)
- [ethics.](#) *ACM SIGCAS Computers and Society* 42, no. 1 (2012): 9-20.
- Mhlambi, Sabelo. "[From rationality to relationality: ubuntu as an ethical and human rights framework for artificial intelligence governance.](#)" *Carr Center for Human Rights Policy Discussion Paper Series* 9 (2020).
- [Sabelo Mhlambi on what AI can learn from Ubuntu ethics](#), *People + AI Research*, 2020.
- Everett, Jim AC, Clara Colombatto, Vladimir Chituc, William J. Brady, and Molly Crockett. "[The effectiveness of moral messages on public health behavioral intentions during the COVID-19 pandemic](#)." (2020). [pre-print paper, not peer reviewed]
- [Ineffective Altruism](#) (2022)

Do not edit above this line.

ETHICAL FRAMEWORKS

Should Batman kill the Joker?
- Utilitarianism - consequences
- Deontology - action/principle/duty
- Virtue Ethics - intention
- Gandhian
- Ubuntu
- Ethics of Care

Ubuntu Ethics
- An African philosophy that emphasizes self through others

- "a set of values central among which are reciprocity, common good, peaceful relations, emphasis on human dignity, and the value of human life as well as consensus, tolerance, and mutual respect" (Ubuntu Ethics)
- From the AI and non-western philosophy reading, we have to "work together in a community" and "sharing is caring."
- Emphasizes reconciliation and restorative justice, as well as community empowerment over individualism
- Mhlambi, Sabelo. "From rationality to relationality: ubuntu as an ethical and human rights framework for artificial intelligence governance." Carr Center for Human Rights Policy Discussion Paper Series 9 (2020).
- Sabelo Mhlambi on what AI can learn from Ubuntu ethics

Ethics of Care
- Originated by Carol Gilligan, known for her book *A Different Voice*
- A feminist ethical theory that emphasizes interpersonal relationships, connectedness and interdependence, and how others are vulnerable to our actions and choices
- Bottom-up, emphasizing lived experience over abstraction
- Grounded in "voice and relationships… [it] directs our attention to the need for responsiveness in relationships (paying attention, listening, responding) and to the costs of losing connection with oneself or with others" (interview with Gilligan)

Individual Reflection

Please write down your name and the answer the prompt next to it.

Prompt: You have 1 million dollars to donate after a massive tsunami just occurred. You could give directly to families impacted using GoFundMe or you could invest in a well established green energy tech company to fund the development of technology that will slow climate change in the future.

Which would you choose? Which option would a utilitarian choose? Which option would a deontologist choose? How are you defining the "greatest good"? How are you establishing a universalizable principle?

MORAL MESSAGING

The Colorado river is a vital waterway for the southwestern states. But over the past 20 years, flow of the river has decreased by 20%. It is now time to renegotiate a 100 year old compact drafted to distribute water amongst the states that depend on it. The compact was created through consensus and has stood the test of time despite some shortcomings (e.g. not considering Native American rights to the water). Some suggest that a market based approach would be beneficial to most efficiently allocate resources. However, market based approaches run the risk of arbitrage at the cost of the public good

Draft a proposal with your group detailing how the limited water supply should be rationed according to three moral frameworks of your choice. Detail your approach, explaining what your plan is and how this moral framework helps you determine if your plan is ethical.

READING/WATCHING FOR TONIGHT

The Ethical Dilemma of Self-Driving Cars, Patrick Lin (TED-Ed, 2015)

"Uber's Self-Driving Car Didn't Know Pedestrians Could Jaywalk," *WIRED*, 2019

Things to think about:
- Based on your own feelings both before and after thinking through these readings, how do you feel about self-driving cars? Are you excited, worried, something else? Why?
- How would different ethical frameworks affect how self-driving cars function?
- Fortunately there isn't a huge issue with brakes going out on self-driving cars and them having to choose between life and death. What do you think are some of the more pressing ethical issues with self-driving cars we should be thinking about?

REFLECTION (end-of-class)
Please take a moment to leave a note about something interesting you thought about or learned in class today and/or questions that the class raised for you. (Please include your name by default, but you are also welcome to include an anonymous note.)

Thursday 7/7
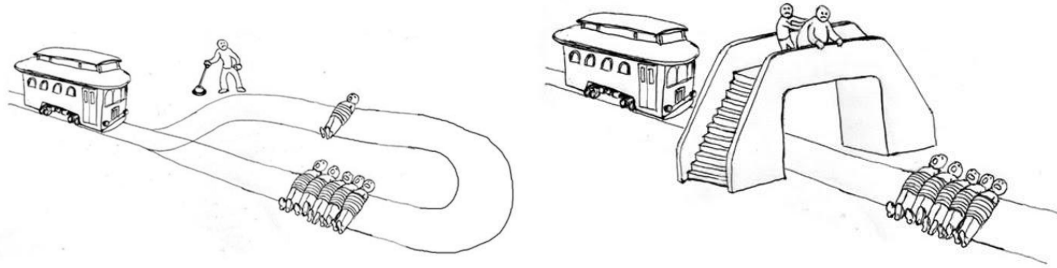
Announcements:
In the News:
Agenda:
Additional Resources

INDIVIDUAL REFLECTION
Go to moralmachine.net, read the instructions, and then click on "Start Judging." It will take you through 13 scenarios. When it asks if you would like to help understand your judgment, you can choose yes or not (though if you choose yes you'll find out a little more.)

Write down your name below. Next to your name, reflect on your experience doing the moral machine experiment. For example: Were you surprised by the outcome and how your answers compared to other people's? What did you learn? How do you think this is useful or not?

THE TROLLEY PROBLEM

Does the second scenario feel different to you? Why or why not?

WHO GETS TAKEN CARE OF?

Fink, Sheri. U.S. Civil Rights Office Rejects Rationing Medical Care Based on Disability, Age.
The New York Times. 2020.

Discuss the following:

- How do you feel about medical care rationing? If a loved one was in a situation where they needed urgent medical care would your perception of medical care rationing change?
- What are other moral frameworks that might be more appropriate than utilitarianism in the case of medical care rationing?

HATE SPEECH OR CENSORSHIP?

You're designing a hate speech detection algorithm to be deployed on a social media platform. You have to decide how to calibrate it, such that it is likely to have more false negatives (legitimate hate speech getting through when it shouldn't) or false positives (content being wrongfully flagged as hate speech when it isn't). Considering who is likely to be harmed by each decision: what do you do?

READING/WATCHING FOR TONIGHT

HIDDEN PENTAGON RECORDS REVEAL PATTERNS OF FAILURE IN DEADLY AIRSTRIKES

Police Are Looking to Algorithms to Predict Domestic Violence

Things to think about:

- How do you feel about medical care rationing? If a loved one was in a situation where they needed urgent medical care would your perception of medical care rationing change?
- What are other moral frameworks that might be more appropriate than utilitarianism in the case of medical care rationing?

REFLECTION (end-of-class)

Please take a moment to leave a note about something interesting you thought about or learned in class today and/or questions that the class raised for you. (Please include your name by default, but you are also welcome to include an anonymous note.)

Friday 7/8

Announcements:

In the News:
-

Agenda:
-

Additional Resources

INDIVIDUAL REFLECTION

What do you think Intel & Invidia's responsibility is in the role of malicious uses of their technology? Write your name down below and your response next to it.

Jeremiah Smith: In this case, I don't think that Intel and Invidia should be blamed for selling their

RESPONSIBILITY AND TECH PROVIDERS

The project maven controversy caused Google to cancel its Pentagon contract. Reflect on the following in your groups:
1. Do you think that Google would have been complicit in possible malicious uses of its technology if it engaged in the contract?
2. How do you feel about Google's decision to pull out of the contract? Was it the right or wrong thing to do? If you were a Google employee, what would your opinion be?

ALLOCATING RESPONSIBLITY

Discuss with your group:

(1) Who do you think is more responsible for the civilian deaths discussed in the [NYTimes Civilian Casuality Files](), the drone pilots or the makers of the AI technology used to identify targets?
(2) How should tech be developed to pre-emptively mitigate human failings? Can technology be developed that can do this?

FINAL PROJECT

You can find the full instructions for the final project here.

Some important points:
- The final project will be an **individual** project, you will be graded **solely** on your efforts. Here
- The final project is due on **AUGUST 5TH, 2022 AT 5PM.** Late work **will not be accepted and you will receive a zero for any work turned in late**.
- The final project is worth 40% of your grade
- The final project is composed of three smaller deliverables, a presentation, and a final report
  - Deliverable 1 due on JULY 11TH @ 11:59PM
  - Deliverable 2 due on JULY 18TH @ 11:59PM
  - Deliverable 3 due on JULY 25TH @ 11:59PM
  - Presentation due on AUGUST 3RD IN CLASS
  - Final Report 1 due on AUGUST 5TH @ 5:00PM
- After each deliverable, you will receive feedback from me at least 48 hours prior to when your next deliverable is due
- Each deliverable is worth 100 points
- You will be giving a final presentation on August 3rd in class on your project

For the final project, you will be asked to identify a controversial technology that was developed within the past decade. <mark>Your job will be analyze *Is the creation and dissemination of this technology ethical?* Additionally, **you will develop a strategy to mitigate possible adverse outcomes if the technology were to be deployed**</mark>.

Deliverable 1 - Identify your topic & sources:
- Submit a write up on the technology controversy you'd like to study summarizing what the technology is, who developed it, why you want to study it, and what makes it a controversy. You should also submit a short annotated bibliography of articles that discuss the technological controversy that you'll draw upon in your analysis. *Your write up should be at minimum 300 words, not including the short annotated bibliography.*

Deliverable 2 - Analyze your controversy w/moral frameworks:

- Consider what utilitarianism, deontology, and virtue ethics would suggest about the ethical standing of this technology and its creators. Write each of these analyses (including explaining each framework), and then consider them as a whole and what you actually think, drawing from and combining these frameworks or making a decision entirely on your own. You may, but do not have to, consider other ethical theories we learned about in class as part of the final synthesis. As part of your analyses, you may also note additional information that you would want to know in order to make better decisions.
- Remember that there are no right or wrong answers. You will be graded on your understanding of and ability to explain these different ethical frameworks, and your ability to think through how they might apply to a specific situation.
- As a rule of thumb, *your answer will likely be around 800 words* -- 100 to explain and 100 to analyze for each ethical theory, and another 200 to synthesize and provide your own opinion. If you consult any additional sources (which you are welcome to but do not have to do), be sure to include a list of references after your response.

Deliverable 3 - Develop a Mitigation Strategy:
- Choose a role for yourself in the company that developed the technology you're developing (e.g. data scientist, software engineer, business analyst, finance analyst, CEO, COO, etc.). Reflect on what influence your role has over the development and dissemination of that technology. Finally, develop a strategy that details (1) what your responsibility would be in terms of impact that the technology has (2) how you would use your role to avoid adverse outcomes and (3) your motivations for and against preventing the development and dissemenation of that technology.
- You will be graded on the quality of your proposed strategy specifically the detail in which you go into and your ability to think through how you could mitigate possible harms as a member of a corporation.
- *This deliverable should be around 500 words* – 50 to explain what your role is within the company and how it relates to the technology at hand, 150 to detail what your responsibility would be in terms of impact of the technology, 150 to layout how you could use your role to prevent adverse affects of the technology, and 150 to explain how you might deploy your strategy and what possible barriers/incentives might be

Deliverable 4 - Put it all together!
- Compile a write up, using the prior three deliverables to argue whether or not it would be ethical to develop and deploy the technology and how you would leverage your position in the company to guide the development of the technology in a way that mitigates possible ethical dilemmas.
- You will be graded on the quality of your writing, the clarity with which you describe the ethical dilemma at hand, your ability to utilize the moral frameworks to analyze your situation, and how well executed your mitigation strategy is.
- Your final report should include a bibliography **and** correctly formatted citations. You can use whatever citation style you like but it should be consistent throughout.

- Your final report should be at a minimum, 1500 words. It should not exceed 2000 words, please :)

In-class Final Presentation
You will present your analysis and strategy to the class in a 4 minute presentation. You may use slides for your presentation, though you don't have to. You may not exceed 5 minutes for your presentation, but you should have at least 3 minutes worth of content to present. Your presentation should include the following:

- An explanation of what your technology is
- Your definition of what it means to be ethical
- How you evaluated how ethical it would be to develop and disseminate the technology
- How you plan to mitigate possible ethical dilemmas in the development of the technology

If you know you will be absent for any reason on August 3rd, please let me know by July 29th!!!! Your final presentation makes up a portion of your final grade so its important that you are present in class that day!