**Problem 4** (Bonus Ethics Assignment, 0pts)

*Estimated total time for completion*: 45 minutes.

In our lecture from class, we discussed philosophical and legal frameworks to examine algorithmic discrimination. But these aren't the only frameworks! A growing body of work in the humanities and social sciences, particularly in feminist studies, critical race studies, sociology, anthropology, and the history of science has emerged to study the social consequences of machine learning.

In this bonus problem, you will be introduced to one framework inspired by scholarship in science and technology studies. To complete the below questions, first watch the 28-minute 2019 NeurIPS talk "The Values of Machine Learning" by Ria Kalluri. Please write no more than 1 paragraph in response to each of the below reflection questions:

1. In their talk, Ria discussed opportunities for shifting power to each of four possible stakeholders: an input source, data source, expert, and decision-maker. Choose one of these stakeholders, and discuss one possible way we as machine learning practitioners and model-builders could shift power to them.
2. What do you think will it take to achieve a world where AI can shift power towards historically marginalized communities? What obstacles or barriers stand in between our current world and your own AI dreams?

## Solution

1. As machine learning practitioners and model-builders, we are the expert stakeholders in the machine learning process. Given the power that we have in the machine learning process, we have a responsibility to do what we can to shift power to marginalized communities, who predominantly represent the input source and data source stakeholders. One way we can shift more power to the data source stakeholders is through enhancing their agency and privacy in the machine learning process. One main source of data for machine learning models is census data, since it is a vast data source that is well-suited for machine learning. However, rarely are the individuals actually asked for their consent of the use of their census data. Given that reporting census data is compulsory, we as machine learning practitioners can give these individuals more power by always asking for the consent of these individuals before their census data is used in our model. This gives the data source stakeholders more agency in deciding if their data is used or not, as well as more privacy as they can not give their consent to use their data on projects that they feel would infringe on their right to privacy. Overall, this is akin to the idea of giving individuals a greater right to their own data.
2. My AI dream would be the achievement of equitable outcomes in the actions dictated by our machine learning models for marginalized and non-marginalized communities. This would shift a vast amount of power to historically marginalized communities, as in the ideal world our machine learning models should not ever lead to different outcomes for non-marginalized and marginalized community. Shifting from our current status quo of unequal outcomes to equal outcomes would vastly increase the power of historically marginalized communities, as model-builder and decision-maker stakeholders will need to take into greater considerations the effects of their machine learning models on marginalized communities. This is a lofty and difficult goal to achieve, as many of the barriers to achieving this AI dream are societal and structural. For instance, marginalized communities tend to be less technologically literate, and therefore are both less capable of exercising agency when they are part of the data source or input source stakeholders. For the same reason, marginalized communities are less likely to be a part of the model-builder or decision-maker stakeholders because of structural barriers in education and career advancement for marginalized communities. Moreover, considering how the research pipeline for machine learning is funded, most money is spent on creating models that don't benefit marginalized communities. Therefore, another obstacle would be to increase the amount of

funding available for machine learning projects that are designed to enhance the power and equality of marginalized communities.