**Lesson 9 Discussion**

This week, you explored ethical issues in Machine Learning. Specifically, you were asked to think about how, in any given machine learning system, the data or the model may be biased towards certain groups.

After this lesson's reading, share some takeaways you have on the issues related to bias in machine learning. What do you do, or plan to do, in your own career to mitigate?

What other issues come to mind when exploring your work in machine learning? Have you thought about the impacts of this work? Explain.

As we know the machine learning models are as good as the data that we input. For this reason, is key to understand the data we are working with and stop for a moment to ask important questions about the assumptions that the person/group/organization had when asking the questions either in the structure of the questions or in the answers provided. This part is complicated specially when companies apply machine learning without a better understanding on the limitations of these types of algorithms.

In the last years as the topic of Machine Learning is maturing in the industry and academia, people have started asking questions about the possibility of unintended bias against specific groups given by the type of algorithm used, the chosen optimization or more commonly the data that was input to the model. As these models get more and more complex the interpretability part gets more and more difficult, which in return makes more difficult to understand how to identify this in the model and corrected (if/as needed). This has brought a better understanding of the responsibility that we have as machine learning practitioners (or future ones) as the impact of an unattended bias could be catastrophic on the long run as what we have lately seen in our society.

I recently took a short class on Introduction to Machine Learning on Azure; the last lesson was on Responsible AI and it was a great introduction to the governing principles on this topic: inclusiveness, transparency, accountability, fairness, reliability & safety and privacy & security. At the end of the training there is a small demo on one of the tools used to verify if there is bias in the output given by the model (which probably comes from the data) and the correction options as well as the impact that might have in accuracy (The python package is called “fairlearn”) .