Film: Coded Bias (Discussion Reflection)

Reflection Question – *In 300-500 words, describe your perspectives on the responsibilities machine learning practitioners have to give thoughtful consideration to potential societal impacts of the algorithms they develop, and how can they exercise their responsibilities.*

Libo Zhang (NetID: lz200)

Answer – In class discussion, our group indeed talked a lot about what societal impacts there could be once a new algorithm is published for commercial use. I think machine learning engineers should grasp and understand two key points when they are developing algorithms. First, they should realize that the results generated by their algorithms could have significant influence on other people's life. According to the film, I realize that algorithms are everywhere. Algorithms can evaluate your credit rating, which could influence your house renting. Algorithms can predict your probability of committing crimes, which could cause you to be stopped and searched by police officers when walking on the street. Algorithms can automatically filter applicants' resumes, which could make you unable to find a job simply because of your gender. Given the above three film examples our group discussed in class, I hope that machine learning practitioners should deeply understand the fact that their algorithms could have significant, or even decisive effects on everyone's daily life. Second, I think machine learning engineers should realize that predictions/evaluations made by their algorithms are not always accurate. In the film, for example, a hard-working, highly respectable teacher was poorly rated and suggested to be fired by an algorithm introduced by school to help evaluate faculty. The teacher works very hard to help his students succeed, but the algorithm mistakenly judges him as an incompetent employee. Before implementing algorithms in the society, machine learning practitioners should hold the idea that their models are not panaceas to everything in the world, and they should take the responsibility of acknowledging the algorithms' flaws and mistakes in the first time.

How can machine learning practitioners exercise their responsibilities described above? Firstly, since we are focusing on societal impacts, I think the most important work is still model debiasing. In class, our group actively discussed current efforts made by researchers and engineers to eliminate or reduce bias among gender, race, age, income level and so on. After discussion, I think debiasing could be conducted starting from data preprocessing, to even building another machine learning model to detect and reduce bias in existing algorithms. Clearly a huge amount work remains to be done in the area of model debiasing. Secondly, our group also discussed what emphasis machine learning practitioners should have. I think when engineers are developing and deploying their algorithms, they should concentrate more on aiding/helping people to make accurate decisions, instead of replacing people to make important decisions simply according to model evaluation results, which could be completely absurd given the above discussed examples. In conclusion, I think we still have a very long way to go, and much work to do, to eventually eliminate the "Coded Bias".