**Rajat Sethi – CPSC 8580 – Paper Summary 3 – A Survey on Bias and Fairness in Machine Learning:**

This article is a conglomeration of several years of research in AI-centered fairness and the problems that could occur with such biases. Some ML algorithms may be misleading for several reasons. The data might not be inclusive enough, the data could be missing a hidden variable, user data might be inherently discriminatory or provided by different demographics, and the list goes on. With all these potential biases, this paper attempts to compile a list of possible solutions for the different error inducers.

The best solution to prevent bias is to improve the true positive rate. Using a comprehensive testing data set with an even amount of each demographic, an AI engineer can determine how much bias is in their system and what their next actions should be. One of the paper’s proposed solutions is to remove any discriminatory variables. For example, in an algorithm that determines criminal bail, race and gender should not be factors.

Most of this paper provides a set of similar problems and solutions. The key novelty of this paper is that it’s a centralized document of a lot of research. It does not provide “new” information, but rather collects information from several sources and condenses them into this one paper. Developers and engineers no longer have to search through 30-40 papers to solve their bias issue, they can look at this one long paper instead. I believe that researchers in other fields should follow this example and make their field more accessible and centralized.

Since this paper does not provide much “original” research, there are not too many weaknesses or limitations. In a future version of this paper, I would hope to see a little more pseudocode or math to visualize the solutions. Having a formula or code that fixes the algorithm, plus statistics to back up these changes, would be quite helpful in similar, future scenarios.