Varnika Sinha Section B

Are We Making AIs Racist and Sexist?

As technology advances with our culture, computers are becoming more like ourselves through the use of algorithms. An algorithm is a step-by-step method a computer runs ~~through~~ based on decisions that evaluate to true or false. Computers today use algorithms to learn and model specific human behaviors. Artificial Intelligence (AI) computers take large amounts of data and extract patterns to predict human behaviors and mannerisms. Unfortunately, these data sets ~~more~~ often ~~than not~~ contain social bias that computers learn.

Every day, we use our phones at least once. Whether it is to find a good restaurant or directions to a friend's house, we are giving raw data to the computer. This raw data helps the computer analyze what we are like as individuals. By analyzing the given data, the computer is able to improve its optimization search for things we like for the next time the search is used. A computer is like a newborn baby, we give it knowledge based on our actions and diction. This ability may be beneficial, however, computers learn us as different individuals and pick up on the slightest bias on topics that even human analysts wouldn't pick up.

Although having a machine learning system starting equivalent to a newborn is beneficial, it often falls prey to blatant stereotypes. When a computer receives audio or text, it does not only read each word, but also observes the relationships between words. This analysis is based on many factors including how often certain words are used together. Scientists decided to test how well word relationships are identified by using word analogy puzzles. The skeleton of the analogy used was "he is to X as she is to Y." When scientists tested "he is to brother as she is to Y," the computer responded sister, which is the correct answer, but when using "he is to doctor as she is to Y," the computer responded nurse, reflecting clear gender stereotypes.

A computer algorithm makes its decisions based on the frequency at which certain words appear next to each other. Unfortunately, the source documents that doctor is more closely associated to males and nurses to females, forcing the algorithm to learn those biases too. ~~This~~ search becomes difficult when search engines such as Google adapt AIs to conform to our appeals. The results shown are bias, because the collected human bias skewed the computer data. These results can be potentially harmful, because computers, like their human counterparts, are being unfair, and the results shown will not be completely accurate, further damaging these stereotypes.

We can build algorithms that collect data without bias. We can collect the data first, then remove the bias from it before the computer analyzes it. By compiling a list of the most common misused gender stereotypes, it will remove all bias before the computer can go through all the data. This solution is not optimal, but it is the best solution the scientists currently have.