5.

Article Title: Study finds gender and skin-type bias in commercial artificial-intelligence systems

(a) Summary:

This article details a study inspired when a research group was using a commercial facial recognition program. When they were trying to demonstrate their project, they found they had to, "rely on one of the lighter-skinned team members to demonstrate it". To further investigate this, a researcher, Joy Buolamwini, began submitting photos of herself to other commercial facial-recognition programs, and they consistently classified her incorrectly, or even failed to detect her as human. This led her to systematically evaluate this problem. First she compiled a set of 1,200 images and coded them based on skin tone. She then applied three different commercial facial-analysis system. She found that error rates were higher for females than for males, and the error rate increased as someone's skin got darker. For people with the darkest skin, some algorithms had an error of 46.5 and 46.8 percent—almost random guessing.

(b) This issue relates heavily to sampling bias, generalization and overfitting. The task we want the algorithm to learn is to classify faces of all colors as male or female—binary classification. However, Joy found that the algorithms were having trouble generalizing what they had learned to certain types of test data, namely test data of people with dark skin. This implies that the algorithms were trained on a dataset that was sampled in a biased way, one mainly with lighter skinned people, and they over-fit the data towards people with lighter skin. They are good at that type of data they over-fit, but when they need to generalize to other people, they fall short.

(c) To address this, we need to include more diverse data in our training set. This will help the models not over-fit the learning problem to just the data they see. I remember us talking about a "representative" test set or randomly drawn set. A set with mostly lighter skinned people may be convenient but it does not represent the world population nor is randomly drawn. Another issue to fix this problem could be to apply a bit of regularization. If we do this, it's possible that models will learn more general facial structures (eyes, mouth), instead of more specific patterns associated with lighter/darker skin. This may bring the testing error up for lighter skinned people, but could overall bring the generalization down for the while dataset. One would need to test his to see if it works.