Data Bias

There has been a lot of discussion in the literature and in the media about the use of biased algorithms. Facial recognition technologies in particular have been in the spotlight regarding (1) their misuse and (2) the fact that most available datasets are biased toward Caucasians and males. If we put privacy and ethical concerns aside for just a moment the problem is very similar to the imbalanced classes problem. There are underrepresented parts of the training data that need more data to become a better system. This is an important topic and if you want to learn more see the *Diversity in Faces* blogpost by IBM Research or the Diversity in Faces* research paper