1. The existing commercial face recognition systems and fair recognition datasets are biased towards Caucasian faces and are generally more accurate for lighter-skinned people and males, while other groups are underrepresented and can have lower accuracy rates. This leads to inconsistency across different groups and raises ethical concerns.

2. The evaluation metric used in this paper for tables is accuracy. The metric for figure 1 is the individual typology angle to measure the association of skin color and race. The metric for figure 5 is cumulative distribution function (CDF), and it's used to analyze the pairwise distances between faces and measure the diversity of the dataset.

3. Firstly, for the bias in dataset, the author checked the racial composition, drew random samples from dataset, performed t-SNE visualization, and examined the pairwise distance between faces. They found white faces are most common in existing dataset.
   To check the performance bias, the author measured the accuracy and the consistency of accuracy across different datasets. The author measured the cross-dataset classification accuracy on white races or non-white races in race, gender, and age. The model consistency is shown by the standard deviation and maximum accuracy disparity.