

# ML FAIRNESS

Human biases { how do they creep in?  
how to mitigate?

- humans remember/store things based on what is typical to them
- these can affect ML... in many steps

## Collection and annotation of data

- reporting bias: what is "interesting" (e.g. bananas vs. green bananas)
- selection bias
  - overgeneralization bias
  - out-group homogeneity bias
  - confirmation bias
  - automation bias ("ML is more reliable than humans!")

## Designing for fairness

1. Consider the problem
2. Ask experts
3. Train model to account for bias
4. Interpret outcomes (i.e. how are the outputs in the real world)
5. Publish with context