Questions

"Training"

1. How can an unrepresentative training data pose fairness issues?
2. If we removed sensitive features of a model, will our model achieve have *equalized odds*?
3. What is one problem with using *demographic parity* as a measure of fairness?
4. What is one problem with using *equalized odds* as a measure of fairness?
5. What evidence is there that word2vec embeddings can be problematic?

"Generalization"

1. How can the gesture recognition model we trained in assignment 3 pose a fairness issue?
2. How can the spam detection model we trained in assignment 5 pose a fairness issue?
3. Show, using example (hypothetical) data, of how *equalized odds* and *demographic parity* can be contradictory goals.