DetoxiPy Model Pipeline Reduce unintended bias Overall perf **Predictive** perf Bias **Text Processing** OUTt = yt - Homogeneizing selected proper nouns and contractions - Remove selected punctuation Replace identity words for ... neutral noun/pronoun [NN Models] [Non-NN Models] **Word embeddings** Naive Bayes **LSTM** Logistic **GRU** Random Regression **Forest** Train NN model Pre-trained to produce word embedding with dataset corpus Word2Vec GloVe Xt (Validation: tune hyperparams) Initial word embedding matrix [v[wi]] i=1,...,|V|