



ETHIXAI

ALGORITHMIC AUDIT

Try out recommendations in this checklist to help guide your journey to de-biasing your Tech product. not every recommendation will apply to your product, but here are some suggestions to get you started.

PREPROCESSING RECOMMENDATIONS

CONSIDERATIONS TO DECREASE LABELLING BIAS

- Alter class labels of selected instances that fall on the decision boundary to increase representation of protected groups with "favorable" label. See page 16 on full guide.
- Assign different weights to instances based on their group membership so that for example protected groups with "favorable" label have greater weight. See page 16 on full guide.
- Ensure multiple people label each instance and that people labelling have diverse demographics to limit implicit biases. Is your labeling criteria explainable?
- Think critically about the source of your dataset. Are certain groups overrepresented in your dataset?

CONSIDERATIONS TO DECREASE SAMPLING BIAS

- Check that no group is over or underrepresented based on the population you are attempting to represent.
- Increase representation of protected groups in dataset by oversampling and undersampling certain data instances



ETHIXAI

ALGORITHMIC AUDIT

IN-PROCESSING RECOMMENDATIONS

In-Processing techniques are algorithm-specific, so research in-processing techniques to mitigate bias specific to your chosen learning algorithm.



Research in-processing techniques to mitigate bias specific to the learning algorithm you are using.



Look into whether using different fairness constraints is a useful technique for the problem you are trying to optimize. See table on page 23 on full guide for some constraint ideas for different algorithms.



Use multiple evaluation metrics that reveal whether predictions vary when looking only at protected vs unprotected groups. See page 22 on full guide.



ETHIXAI

ALGORITHMIC AUDIT

POST-PROCESSING RECOMMENDATIONS

Post-processing techniques must be an iterative process. As novel data is consumed by your learning algorithm, it may develop new biases, which makes constant monitoring imperative.

IMPLEMENTATION OF PROCESSES FOR ONGOING MONITORING

- Identify and implement a process for ongoing monitoring of your algorithm for potential bias.
- Consider hiring a third-party to check for and monitor potential biases in your learning algorithms over time.

CONSIDERATIONS TO DECREASE BIAS OF PREDICTIONS

- Tune thresholds to promote or demote predictions that fall close to the decision boundary lines.
- Alter probabilities of class membership so that the probabilities are closer to the true likelihood.

CONSIDERATIONS RELATED TO EXPLAINABILITY & TRANSPARENCY

- Be transparent about the assumptions used by your learning model.
- Be able to explain model's outputs so that a human can understand the decision-making process.