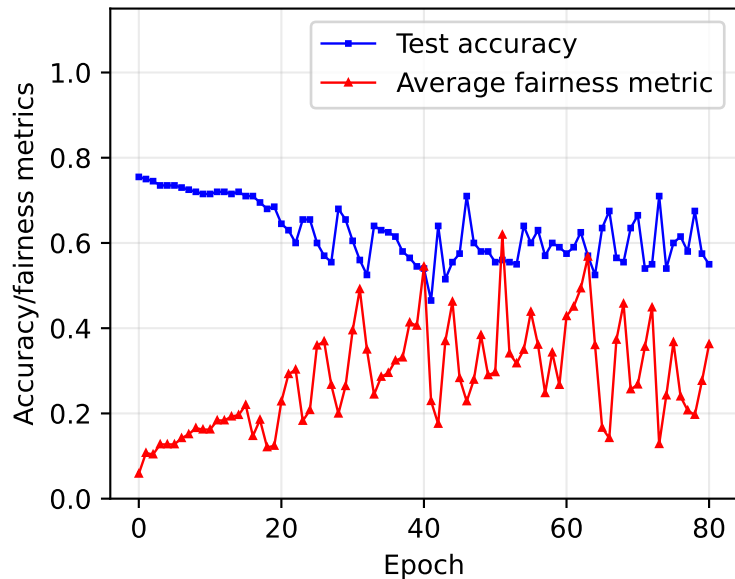**Difference between the two stopping metrics**
($\varepsilon = 0.5, \lambda = 1$)

German dataset (Solans attack)

Drug dataset (IAF attack)

Test accuracy
Average fairness metric

Accuracy/fairness metrics

Epoch