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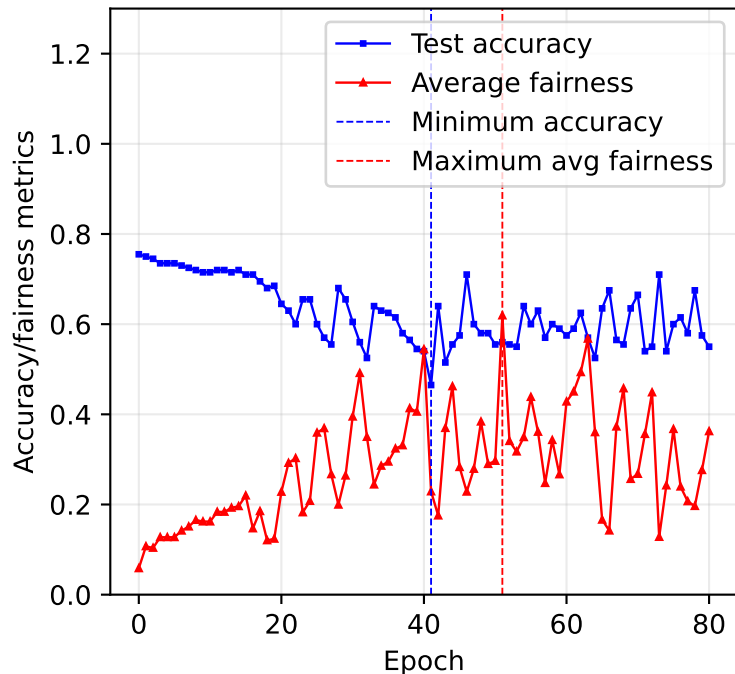**($\varepsilon = 0.5$, $\lambda = 1$)**

German dataset (Solans attack)

- Test accuracy
- Average fairness
- Minimum accuracy
- Maximum avg fairness

Drug dataset (IAF attack)

- Test accuracy
- Average fairness
- Minimum accuracy
- Maximum avg fairness