

## **Fair Decision-Making via Counterfactual Sensitivity Analysis**

Fei Huang

UNSW Sydney  
School of Risk and Actuarial Studies  
Australia

Joint work with: Silvana Pesenti

This paper proposes an individual fairness criterion via counterfactual sensitivity analysis. It achieves fairness by eliminating the sensitivity of decision-making to protected attributes. Different from traditional fairness criteria, which only focus on fair machine learning tasks, this fairness criterion is applied to a two-step decision-making process, which can be potentially applied to various applications, especially in insurance and finance. The first step is a machine learning task estimating a loss random variable; the second step is for decision-making by computing a distortion risk measure of the loss random variable. The counterfactual sensitivity fairness criterion is directly applied to the final decision-making stage and ensures fair outcomes. In this paper, we use fair insurance pricing as an example for empirical analysis.