

Professional Ethics in Computing – Seminar 9

Trustworthy AI

In today's seminar you will work on a dataset which records the output of an AI system. You will analyze it to determine which fairness criteria have been met. You may work individually or in your team.

Step 1. Open the file `model_outcomes.csv` in Excel, Python, Matlab or whichever data analysis package you prefer to use.

The file has the following variables:

A1 : sensitive attribute for sex (0=female; 1=male).

A2: age (1=elderly; 0=not elderly).

A3: ethnicity (1=minority; 0=majority).

Y: outcome (0/1).

R: model response (decision).

Step 2. Analyze the data to determine which of the sensitive attributes A1, A2 or A3 are fair by the Independence or Sufficiency definitions.

Allow for +/- 5% error between probabilities.