Mireille Hildebrandt

Vrije Universiteit Brussel

The issue of bias: trade-offs and balance in ML

There is a lot of talk about bias when it comes to machine learning. From a computer science perspective, this may refer to the productive bias that enables ML, both at the level of picking the training set and at the level of training the algorithms. It reminds one of David Wolpert's 'no free lunch theorem', if not of Humean scepticism or Gadamer's acknowledgment of constitutive presumptions. This relates to e.g. the trade-off between the size of a training set, its relevance, the types of algorithms used, and the accuracy and/or speed of the results. From a societal perspective, bias may refer to unwarranted or even unlawful discrimination. It is crucial to distinguish these two conversations about bias, while teasing out where they meet, interfere and cause Babylonian confusions. In this talk I will propose that lawful employment of behavioural data requires asking very precise questions about trade-offs, in order to answer legal questions of balance. This incorporates and accounts for ethical questions about proportionality and trustworthiness.