

Philosophy & Ethics of Artificial Intelligence

Master in Philosophy of Science

Philippe van Basshuysen

09 December 2019

Worksheet 7 – Algorithmic Bias and Fairness

1. **Algorithmic Bias**

Define different kinds of sources of algorithmic bias and give one simple example of each. (You can use Danks & London’s 2017 proposed taxonomy for sources of biases but you should come up with your own examples.)

Are the biases in your examples ethically problematic? Why, or why not? How should we (as philosophers, policy makers, society as a whole) react to biases?

1. **Fairness**

Why are the mental state and the generalisation account of discrimination problematic to account for algorithmic discrimination?

Why or why not should we regard the COMPAS system as unfair?

1. **Next Week: Ethical Decision-Making with Algorithms**

Readings:

Alexandra Chouldecheva (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. arXiv:1703.00056v1 [stat.AP]. NOTE: This is a more technical paper. Try to read pp. 1-7.

Anders Persson and Iordanis Kavathatzopoulos (2017). How to Make Decisions with Algorithms: Ethical Decision-Making Using Algorithms within Predictive Analytics. *ACM Computers & Society* 47(4)

Optional readings:

Kirsten Martin (2018). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics* 1-16.