Call for Abstracts
Workshop on Algorithmic Fairness

Venue: University of Copenhagen (with an option to participate online)

Date: November 12-13, 2020

Keynote speakers: Benjamin Eidelson (Harvard Law School) and Deborah Hellman (University of Virginia School of Law)

Organizers: Sune Holm (University of Copenhagen) and Kasper Lippert-Rasmussen (University of Aarhus)

Description

The topic of algorithmic fairness was kick-started by an article (Angwin et al. 2016), which claimed that an algorithm called COMPAS, widely used in American courts to decide whether defendants awaiting trial are too dangerous to be released on bail, was biased against blacks. The basis for this claim was that COMPAS' error rates for black and white defendants are very different. Blacks are almost twice as likely as whites to be classified as too dangerous for release, when they are not. Whites are almost twice as likely as blacks to be classified as low risk, when they are not. In the wake of these criticisms it became clear that similar concerns arise in other domains including education, credit, hiring, social services, and medicine. Furthermore, researchers have shown that in ordinary circumstances it is mathematically impossible to equalize false positive and false negative rates and simultaneously ensure that the algorithm's predictions are equally accurate for members of the two groups (Chouldechova 2017, Kleinberg et al. 2017). More generally, the scrutiny of the COMPAS algorithm's outcomes and alleged discrimination of black defendants has led to a realization that widely held conceptions of algorithmic fairness cannot be satisfied simultaneously by an algorithmic decision maker.

This workshop is prompted by a call for philosophical reflection on how to assess the fairness of algorithmic decision making. In order to realize the potential benefits of applying computer systems to make consequential decisions about people in both the public and private sector, a wide range of stakeholders, including policy makers, public administrators, businesses, NGOs, and academics, have expressed a concern for the development of fair algorithms. How to define and model algorithmic fairness is currently a topic of intense debate in the field of AI known as machine learning. However, as several researchers in that field have pointed out, what it *means* for an algorithm to be fair is not a question that can be solved mathematically. It is a question of ethics. Hence there have been calls from the machine learning community for philosophers to engage in and apply their expertise to the topic of algorithmic fairness. The aim of this workshop is to do just that. Questions that may be discussed by the contributions include but are not limited to:

- How does the algorithmic fairness problem relate to definitions of discrimination, e.g., standard definitions of statistical discrimination and the cut between direct and indirect discrimination?
- Should variables such as race and gender be used as the basis for algorithmic classifications of individuals if it can e.g. increase accuracy and equality of error rates?
- Why are instances of disparity between (socially salient) groups resulting from the use of algorithmic instruments (such as COMPAS) objectionable, when they are?
- Are existing moral frameworks, such as the ideal of equality of opportunity, adequate for analysing different definitions of algorithmic fairness, such as equality of error rates, or does algorithmic fairness give rise to novel ethical problems and principles?
- How should trade-offs be made between context-specific decision-making goals, such as maximizing public safety, and wider societal goals such as reducing racial disparities?

We invite scholars interested in giving a presentation to submit an abstract of approximately 400 words to Sune Holm (suneh@ifro.ku.dk).

Deadline for submissions: September 8, 2020.
Notification: September 15 2020.

The workshop will be held online with a possibility of participating in person, if conditions allow it.

The workshop is organized by the project Bias and Fairness in Medicine funded by Independent Research Council Denmark, and the Centre of Excellence for the Experimental-Philosophical Study of Discrimination (CEPDISC) funded by the Danish National Research Foundation.

References

- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big Data, 5(2):153–163.
- Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. and it's biased against blacks. ProPublica.
- Kleinberg, J., Mullainathan, S., and Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. In Proceedings of Innovations in Theoretical Computer Science (ITCS).