Algorithmic Discrimination

Ethics, Data, and Technology

David Gray Grant

**Algorithmic discrimination**

Wrongful discrimination resulting from the use of a predictive algorithm to decide how to treat individuals.

**Wrongful discrimination**

A decision procedure *wrongfully discriminates* against social group *x* if and only if:

1. there is a social group *y* such that the procedure treats the members of *x* less favorably than the members of *y*;
2. part of the explanation for the difference in treatment is their membership in *x* and *y*, respectively; and
3. the difference in treatment is not morally justified.

**Two kinds of wrongful discrimination**

1. *Direct discrimination* = wrongful discrimination resulting from a negative attitude toward the social group.
2. *Indirect discrimination* = wrongful discrimination that does not result from such an attitude.

**What makes wrongful discrimination wrong?**

Two possibilities:

1. Failing to treat similar people similarly.
2. Reinforcing structural discrimination without adequate justification.

*The Aristotelian conception of fairness:* treating people fairly requires treating similar people similarly

* i.e., treating people who are *similar in all morally relevant respects* in *similarly favorable ways*

*Structural discrimination* = “When the rules of a society’s major institutions reliably produce disproportionately disadvantageous outcomes for the members of certain salient social groups and the production of such outcomes is unjust” (Altman, “Discrimination,” *Stanford Encyclopedia of Philosophy*).

**Three causes of algorithmic discrimination**

1. Biased training data.
2. Biased proxy.
3. Accurate predictions, but decisions that reproduce preexisting inequalities.