Discrimination by People and by Algorithms

Tamar Kricheli-Katz & Noya Matalon

The article describes three types of human discrimination: taste-based discrimination, founded on personal preferences; stereotypical discrimination, primarily based on prejudices; and statistical discrimination, based on statistically accurate but normatively undesirable generalizations.

While artificial intelligence systems may reduce discrimination stemming from human preferences, such as taste-based and stereotypical discrimination, they cannot avoid statistical discrimination, which may discriminate against some product users or third parties. Artificial intelligence operates on the basis of large databases and inherently performs statistical generalizations. Therefore, when gaps are statistically accurate, they will be manifested in algorithmic output. Furthermore, even when actively attempting to remove discriminatory variables such as gender or race, algorithmic systems can infer these variables from other variables, and tracing the discriminatory variable in the decision is nearly impossible. Thus, in certain cases, the results obtained by algorithmic systems may actually reproduce existing societal discrimination and inequality.

The article argues that the algorithmic project poses a significant challenge to existing antidiscrimination laws. We propose splitting the legal treatment: continuing to apply traditional antidiscrimination laws to human decisions, while developing a separate legal framework for algorithmic decisions that focuses on examining their outcomes and distributive effects.

This approach will require complex normative determinations regarding the scope of permissible gaps between groups and how to distribute the costs involved in reducing them.