

On the Fairness of Machine-Assisted Human Decisions

Talia Gillis
gillis@law.columbia.edu
Columbia Law School
New York, NY, USA

Bryce McLaughlin
brycem@stanford.edu
Stanford Graduate School of Business
Stanford, CA, USA

Jann Spiess
jspiess@stanford.edu
Stanford Graduate School of Business
Stanford, CA, USA

ABSTRACT

When machine-learning algorithms are deployed in high-stakes decisions, we want to ensure that their deployment leads to fair and equitable outcomes. This concern has motivated a fast-growing literature that focuses on diagnosing and addressing disparities in machine predictions. However, many machine predictions are deployed to assist in decisions where a human decision-maker retains the ultimate decision authority. In this article, we therefore consider how properties of machine predictions affect the resulting human decisions. We show in a formal model that the inclusion of a biased human decision-maker can revert common relationships between the structure of the algorithm and the qualities of resulting decisions. Specifically, we document that excluding information about protected groups from the prediction may fail to reduce, and may even increase, ultimate disparities. While our concrete results rely on specific assumptions about the data, algorithm, and decision-maker, they show more broadly that any study of critical properties of complex decision systems, such as the fairness of machine-assisted human decisions, should go beyond focusing on the underlying algorithmic predictions in isolation.

CCS CONCEPTS

• **Human-centered computing** → HCI theory, concepts and models; • **Applied computing** → Law; Economics.

KEYWORDS

Protected Classes, Human Computer Interaction, Decision Support Systems, Fairness, Machine Learning

ACM Reference Format:

Talia Gillis, Bryce McLaughlin, and Jann Spiess. 2022. On the Fairness of Machine-Assisted Human Decisions. In *Proceedings of FAccT '22: ACM Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/3531146.3533152>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
FAccT '22, June 21–24, 2022, Seoul, South Korea
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9352-2/22/06.
<https://doi.org/10.1145/3531146.3533152>