Biography
Solon Barocas is an assistant professor of information science at Cornell University. Barocas focuses on the ethics of machine learning, particularly applications that affect people’s life chances and their everyday experiences on online platforms. He is currently exploring issues of fairness in machine learning, methods for bringing accountability to automated decision-making, the privacy implications of inference, and the role that privacy plays in mitigating economic inequality.

Abstract
Normative debates about bias in machine learning have focused overwhelmingly on problems of unfair allocation: systematically denying desirable opportunities to deserving members of legally protected classes (e.g., women, racial minorities, etc.). Understanding fairness in these terms has limited the field's ability to recognize and address harms other than those involving economic opportunities withheld. We survey a number of examples where the harms of bias in machine learning are representational in nature—where applications of machine learning propagate derogatory messages about specific social groups, reinforce cultural stereotypes, fail to recognize members of marginalized and minority communities, or under- or over-represent them in depictions of valorized or censured activities. Using a series of real-world examples of machine vision and natural language processing, we assess how bias in machine learning can have the effect of privileging certain identities and denigrating others—and we contrast these with the more common problems of allocation addressed by the existing literature. While allocative harms spring from discrete moments of decision-making (e.g., whether or not someone gets a job), representational harms tend to be much more diffuse (e.g., cultivating general beliefs about the people most suited to certain careers). And while the long term effect of these kinds of representational harms might be uneven access to societal resources, we argue that representational harms need to be addressed in and of themselves.