UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

The Quantified Moral Self

Permalink

https://escholarship.org/uc/item/1dk2f1cc

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors

Purcell, Zoe A

Rahwan, Iyad

Bonnefon, JF

Publication Date

2022

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

The Quantified Moral Self

Zoe Purcell
University of Toulouse, Toulouse, Occitanie, France

Iyad Rahwan
Max Planck Institute for Human Development, Berlin, Berlin, Germany

JF Bonnefon
Toulouse School of Economics, Toulouse, France

Abstract

Artificial Intelligence (AI) can be harnessed to create sophisticated social and moral scoring systems – enabling people and organisations to form judgements of others at scale. While this capability has many useful applications – e.g., matching romantic partners who are aligned in their moral principles, it also raises many ethical questions. For example, there is widespread concern about the use of social credit systems in the political domain. In this project, we approach this topic from a psychological perspective. With experimental evidence, we show that the acceptability of moral scoring by AI depends on its perceived accuracy, and that perceived accuracy is compromised by people's tendency to see themselves as morally peculiar, and thus less characterizable by AI. That is, we suggest that people overestimate the peculiarity of their moral profile, believe that AI will neglect this peculiarity, and resist for this reason the introduction of moral scoring by AI.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*. ©2022 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).