

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Studying the Effect of Moderator Biases on the Diversity of Online Discussions: A Computational Cross-linguistic Study

Permalink

<https://escholarship.org/uc/item/34k4m5qg>

Journal

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

Authors

Hassan, Sabit

Atwell, Katherine J

Alikhani, Malihe

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

Studying the Effect of Moderator Biases on the Diversity of Online Discussions: A Computational Cross-linguistic Study

Sabit Hassan

University of Pittsburgh, Pittsburgh, Pennsylvania, United States

Katherine Atwell

University of Pittsburgh, Pittsburgh, Pennsylvania, United States

Malihe Alikhani

UPitt , Pittsburgh, Pennsylvania, United States

Abstract

The methods by which people harm others evolve with changes in, and in access to, technology. Several cognitive, linguistic, and behavioral theories have suggested that biased language use is correlated with dominance and can reduce the diversity and inclusivity of a community (e.g. Poteat et al, 2010). We present a cross-cultural and cross-linguistic study of moderators on Reddit in English, Arabic, and French. We collect and analyze a large Reddit moderation dataset and use machine learning models to study cognitive and behavioral differences of moderation across cultures. We then work with expert linguists who analyze and evaluate our results. Finally, we explore the implications of our models for studying how we might shut down voices from different communities by not moderating online content properly. Our preliminary results reveal biases towards women and minority groups, and more broadly affirm our hypothesis that culture and topic of discussions bias moderation decisions.