

CONTENT MODERATION AND HUMAN RIGHTS: WHAT WORDS DO (AND WHAT PLATFORMS SHOULD DO WITH THEM)

BROWN BAG SERIES

Hoffman Center for Business Ethics

Thursday, September 19th, 2024

12:30 p.m. to 1:45 p.m.

[Register Here](#)

This paper presents a new theoretical account of the human rights obligations of online platforms with regard to the moderation of hateful and discriminatory content. We build upon speech act theory to characterize the type of human rights harm that is caused when hateful and discriminatory speech is allowed to remain online. Based on feminist and critical race approaches to speech act theory, we characterize hateful and discriminatory speech acts as injurious in themselves, and we show that this type of injury consists of a violation of the principles of equality and non-discrimination, and potentially of freedom of expression, of every individual affected by the relevant social structure of oppression. We characterize the different types of content moderation decisions that platforms can adopt as a response, and we determine the platforms' level of involvement in human rights violations in each case. Our main contention in this paper is that content moderation is not just an act that balances interests of equal value, but rather a fundamental choice as to whether structures of oppression are reinforced or opposed in society. We conclude that platforms may play a very direct role in human rights impacts and so incur specific responsibilities that, in some cases, will include remediation, and we consider a proposal for appropriate remedy.



Rita Mota

Assistant Professor
ESADE Business School
Barcelona, Spain

Presented by the Hoffman Center for Business Ethics at Bentley University.

FOR MORE INFORMATION

gsands@bentley.edu | 781-891-2981

bentley.edu/cbe

Bentley University, 175 Forest Street, Waltham, MA 02452 USA



BENTLEY UNIVERSITY

W. Michael Hoffman
Center for Business Ethics