

Explaining Sexism or Racism with GPT-3 (temp) *

Ke-Li Chiu *University of Toronto*

Rohan Alexander *University of Toronto and Schwartz Reisman Institute*

abstract...

Keywords: GPT-3; natural language processing; quantitative analysis; hate speech.

Introduction

Background

Methods

Result

Discussion

*Code and data are available at: <https://github.com/kelichiu/GPT3-sexist-racist-content-explanation>. We gratefully acknowledge the support of Gillian Hadfield and the Schwartz Reisman Institute for Technology and Society. We thank Amy Farrow, Haoluan Chen, Mauricio Vargas Sepúlveda, and Tom Davidson for helpful suggestions. Comments on the 05 July 2021 version of this paper are welcome at: rohan.alexander@utoronto.ca.