

Recovering belief structures using a language model on a naturalistic dataset of attitude change

Trisevgeni Papakonstantinou
University College London, London, United Kingdom

Antonina Zhiteneva
University of Oxford, Oxford, United Kingdom

Ana Ma
Arizona State University , Phoenix, Arizona, United States

Derek Powell
Arizona State University, Phoenix, Arizona, United States

Zachary Horne
University of Edinburgh, Edinburgh, United Kingdom

Abstract

On the Reddit forum ChangeMyView, users post beliefs and invite others to challenge them. In this study, we aimed to determine whether a GPT-4-based analytical pipeline could accurately recover belief structures from a subset of posts on predefined topics, identified through covariation statistics from a lab sample. This approach would enable us, in a second stage, to extract novel insights from naturalistic data on belief structures that have not been directly elicited in lab studies, providing a bottom-up examination at scale. Our findings suggest that the pipeline captures meaningful belief patterns, aligning moderately with human responses in structured surveys. Analyzing 3082 posts from 346 users revealed distinct ideological clusters and belief patterns that mirrored well-established political divisions. This method offers a scalable way to study belief networks, shedding light on their role in shaping societal attitudes.