

Iowa Initiative for Artificial Intelligence

Final Report

Project title:	Algorithmic Personalization and Online Radicalization	
Principal Investigator:	Rishab Nithyanand	
Prepared by (IIAI):		
Other investigators:	Brian Ekdale, Timothy Havens	
Date:	Nov 3, 2020	
Were specific aims fulfilled:	Y	
Readiness for extramural proposal?	Y	
If yes ... Planned submission date	Submission made and project awarded in July 2020	
Funding agency	Department of Defense	
Grant mechanism	Minerva Initiative	
If no ... Why not? What went wrong?		

Brief summary of accomplished results:

Proposal was submitted and eventually funded.

Related publication in review (accepted with revisions at ACM CSCW)

Research report:

Aims (provided by PI):

Our overall objectives for this project are, first, to identify the technological, personal, relational, and cultural factors that contribute to radicalization of vulnerable populations on social media and, second, to develop tools that can identify and predict radicalization on social media.

- Aim 1: Use a longitudinal survey to identify personal, technological, relational, and cultural factors that predict change in the radicalization of vulnerable populations on social media. We will use a longitudinal survey of politically-engaged U.S. adults to examine the psychological factors that make an individual more or less vulnerable to radicalization. The two-wave survey, distributed one year apart, will measure participants' views on a range of political, cultural, and social topics, which will allow the research team to identify individuals whose views have become more or less extreme over time. The survey also will measure participants' sense of belonging and social capital, their motivations for using social media, and their susceptibility to conspiracy theories (Brotherton, French, & Pickering, 2013). In addition, the survey will adapt a methodology recently utilized by the Pew Research Center (Hitlin & Rainie, 2019) to measure users' knowledge, attitudes, and behaviors about algorithmic personalization. We will analyze the relationships among these variables to determine the persona, technological, relational, and cultural factors that predict how people use social media. This use of social media is likely to

lead to the experience of personalized content, which might be an explanatory factor in subsequent radicalization.

- Aim 2: Use in-depth interviews to identify strategies individuals use to respond to and make sense of algorithmic personalization and radical content on social media. We will use in-depth interviews with politically-engaged U.S. adults to examine users' mental constructs of what personalization algorithms are and what they do (Bucher, 2018). Interview questions will address participants' knowledge, experiences, attitudes, and responses to algorithmic personalization in their everyday lives as well as what participants do when they encounter radical content on social media. The interviews will incorporate three tasks involving participants' own social media accounts. During each task, participants will be asked how they make sense of and respond to personalized content as well as radical content they encounter on social media. We will use the interview data to understand the different interpretive strategies users employ in their daily experiences with algorithmic personalization.
- Aim 3: Use automated methods to approximate user behavior for continued monitoring of algorithmic personalization experienced by various online communities. We will design and test a sock-puppet methodology for auditing the algorithmic personalization experienced by different sets of users online. We will seed online personas (i.e., build automated profiles that approximate actual user behavior) based on two different data sources: the first acquired through algorithms which only incorporate public interactions of online communities on social media platforms and the second acquired through collecting browsing data from actual members of these communities. We will then compare the algorithmic personalization experienced by these two sets of online personas to determine if publicly available data is sufficient for realistic testing of algorithmic personalization.
- Aim 4: Use the results from previous aims to develop techniques to identify communities vulnerable to future exposure to extremist ideologies due to algorithmic personalization. We will apply findings from Aims 1, 2, and 3 to predict future exposure to and adoption of extremist ideologies. We hypothesize that models of community evolution which capture inter- and intra-community interactions in conjunction with our measures of attitudes and behaviors related to algorithmic personalization can be used to identify communities vulnerable to future exposure to extremist ideologies due to algorithmic personalization. Ultimately, we aim to design a model that is able to predict communities likely to adopt extremist ideologies as well as understand the impact of specific community attributes on the process of radicalization.

Data: Obtained from surveys and interview questionnaires, Reddit.

AI/ML Approach: Explainable ML and feature importance

Experimental methods, validation approach: NA

Results: NA

Publications resulting from project:

Hussam Habib, Maaz Bin Musa, Fareed Zaffar, Rishab Nithyanand, "To Act or React: Investigating Proactive Strategies for Online Community Moderation". Currently in review at ACM CSCW 2020