




The Virtual Public Square

DEFINING HATE SPEECH ON SOCIAL MEDIA PLATFORMS

Hamza Khan, Master of Arts Candidate
Human Dimensions of Organizations, The University of Texas at Austin

Introduction

- 
- Over 2 billion people communicate daily using social media.
- 
- Hate speech is extremely prominent today and its prevalence is magnified through the use of social media.
- 
- This capstone will research the question: “how is hate speech defined to be policed on social media platforms?”.

Hate Speech


For the purpose of this research we will refer to the definition of hate speech, as stated by Grace Udoh-Oshin:

“Speech that offends, threatens, or insults groups, based on race, color, religion, national origin, sexual orientation, disability, or other traits. Examples of hate speech can include racist cartoons, anti-Semitic symbols, ethnic slurs, or other derogatory labels for a group, burning a cross, politically incorrect jokes, sexist statements, anti-gay protests...etc. Although we see a lot of hate speech online, hate speech is not new and neither is the goal of hate speech.” (Udoh-Oshin, 2017)


Impact of Social Media

- Why People Share on Social Media
- 84%


To support a cause or issues they feel strongly about


- 68%


Use social sharing to build image and demonstrate who they are and what they stand for


- 69%


To participate and feel involved in things happening in the world


- 78%

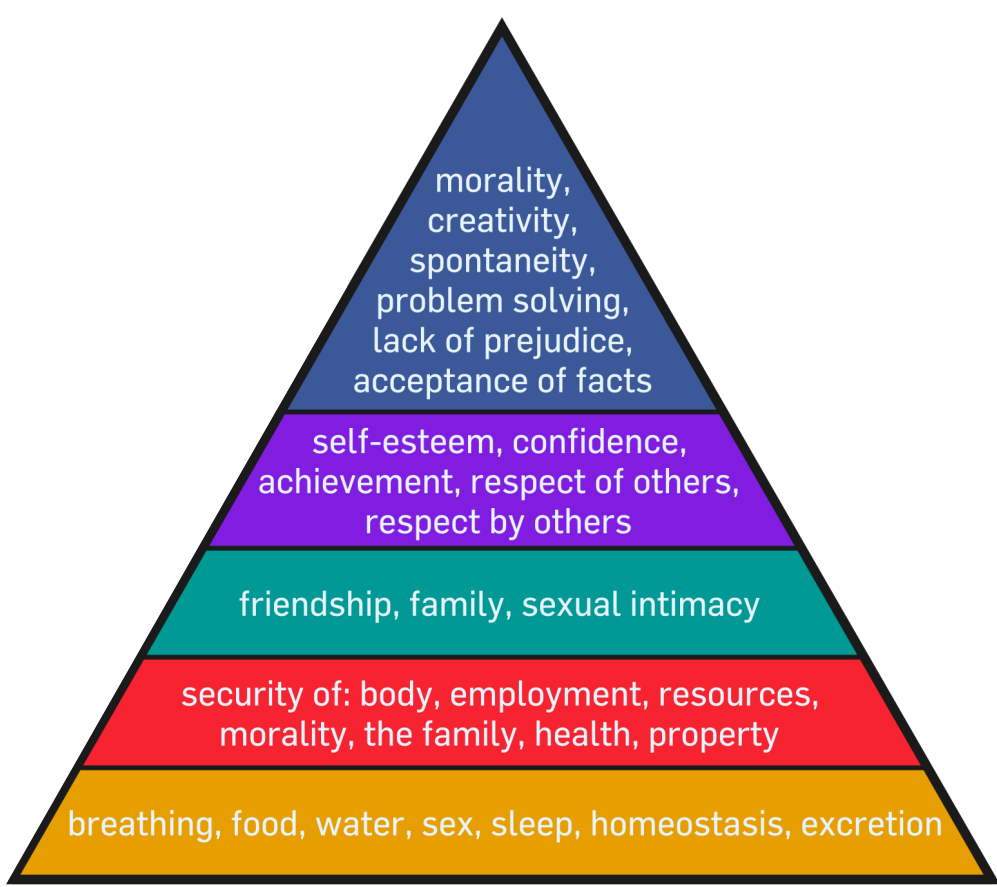
To interact, grow, get a sense of fulfillment, nur relationships and stay in touch with others


- 94%



Share to pass valuable information



Maslow’s Hierarchy of Needs



Methods

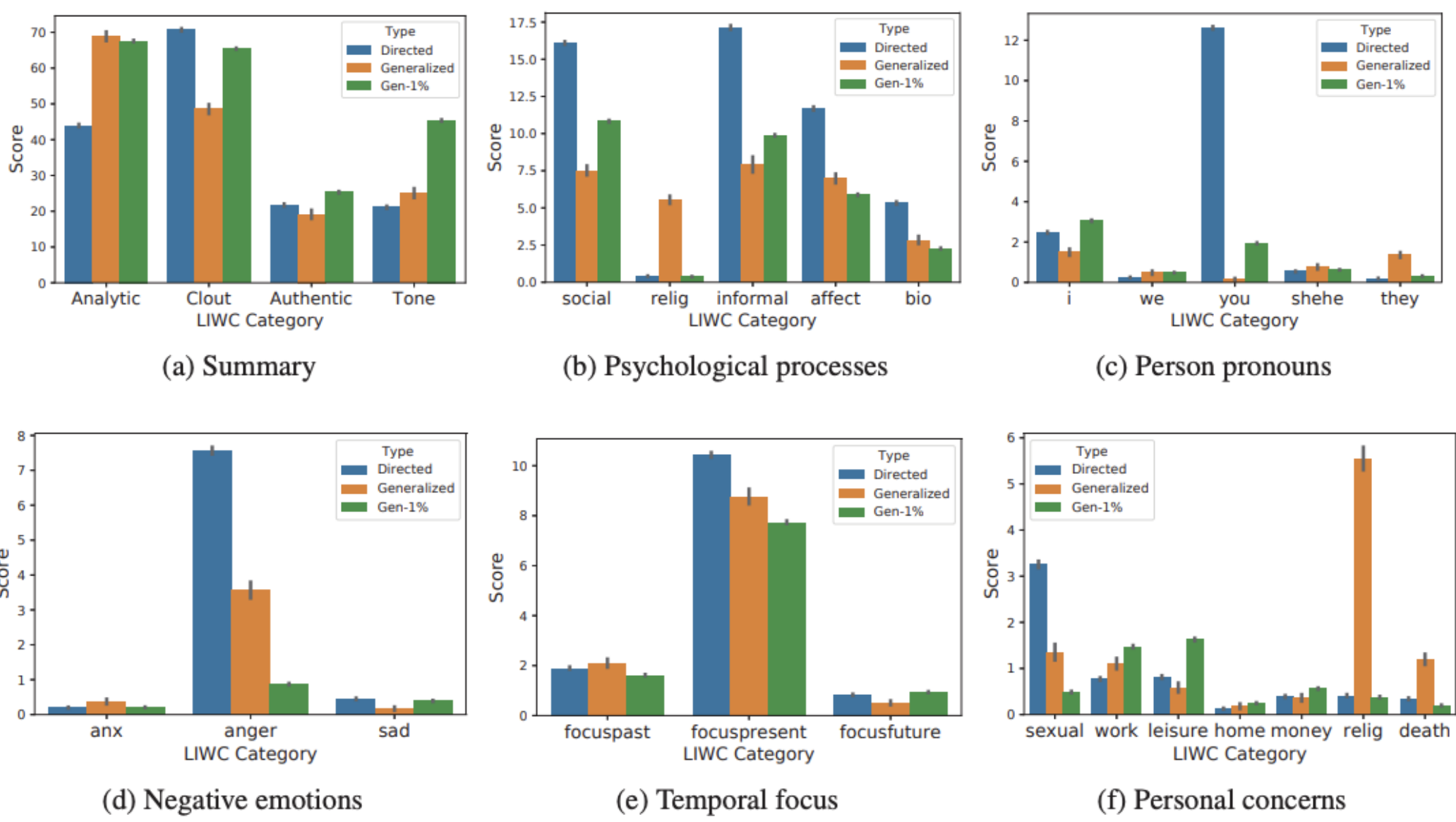
- To research this topic, this study will identify:
- 
- What is the impact of social media in society today?
- 
- Which traits do scholars and humanities professionals utilize to define hate speech?
- 
- How are two prominent social media platforms (Twitter and Reddit) defining their hate speech policies?

Literature Review

- The Need of Human Security and Hate Speech
- The tension between free speech and hate speech is the concept of two competing positions of free speech versus equality. (Onbasi, 2015)
- While there is no universal definition of hate speech, the universal value of human security should be used as an analytical tool to determine content moderation for hate speech online.
- Exceptions in the Classification of Protected Hate Speech by the First Amendments
- Obscenity
 - Child pornography may be banned
 - Defamatory statements may be prohibited
 - Commercial Speech may be banned

Directed vs. Generalized Hate Speech

At the University of California Santa Barbara, a team of researchers addressed this integral topic of hate speech, the target. Using the definitions of hate speech gathered from literature review, their team set out to establish a target based linguistic experiment in which they analyzed hate speech on Twitter for directed hate and generalized hate.



Literature Cited

Twitter: number of active users 2010-2019 | Statista. n.d. <<https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>>.

U.S. Reddit usage reasons 2019 | Statista. n.d. <<https://www.statista.com/statistics/262331/digital-communication-platforms-of-us-internet-users/>>.

Content Policy - Reddit. n.d. <<https://www.redditinc.com/policies/content-policy>>.

Elshierief, Mai, et al. "Hate Lingo: A Target-Based Linguistic Analysis of Hate Speech in Social Media." 2018.

Gençoğlu Onbasi, Funda. "Social Media and the Kurdish Issue in Turkey: Hate Speech, Free Speech and Human Security." Turkish Studies 16.1 (2015): 115-130.

Udoh-Oshin, Grace. "Hate Speech on the Internet: Crime or Free Speech?" n.d.

What is the Major Impact of Social Media? n.d. <<https://www.simplilearn.com/real-impact-social-media-article>>.

Acknowledgments

This wouldn't have been possible without my loving wife Abeer, my encouraging family, my supportive friends, and the staff of HDO program. Every single person played a very integral part in development of this Capstone.

Case Studies

Twitter Hateful Conduct Policy



Social media activities on select social networks by social media users in the United States in February 2019

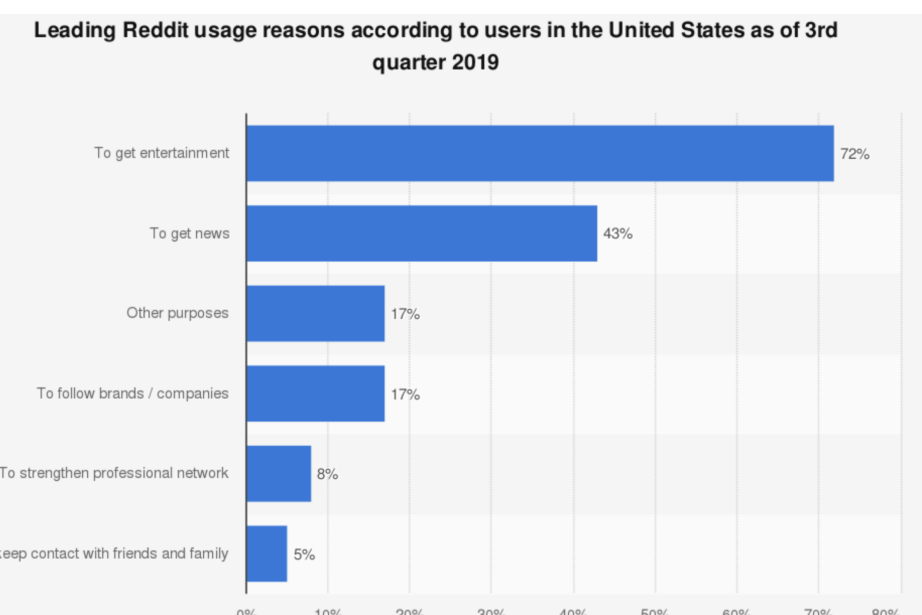
	Instagram	Facebook	Snapchat	Pinterest	Twitter
Viewing photos	77%	65%	64%	59%	42%
Watching videos	51%	46%	50%	21%	32%
Sharing content with everyone	45%	57%	46%	21%	32%
Sharing content one-to-one	31%	43%	45%	12%	20%
Networking	23%	33%	21%	10%	26%
News	18%	38%	17%	9%	56%
Finding/shopping for products	11%	15%	5%	47%	7%
Promoting my business	9%	7%	6%	5%	7%

You may not promote violence against or directly attack or threaten other people on the basis of:

- Race
- Ethnicity
- National origin
- Sexual orientation
- Gender
- Gender identity
- Religious affiliation
- Age
- Disability
- Serious disease.

We also do not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories.

Reddit Content Policy



Content is prohibited if it:

- Is illegal
- Is involuntary pornography
- Is sexual or suggestive content involving minors
- Encourages or incites violence
- Threatens, harasses, or bullies or encourages others to do so
- Is personal and confidential information
- Impersonates someone in a misleading or deceptive manner
- Uses Reddit to solicit or facilitate any transaction or gift involving certain goods and services
- Is spam

Conclusion

While companies such as Twitter and Reddit are struggling to define the line between limiting free speech and allowing hate speech, they should involve scholars in the exercise to make these definitions more complete.