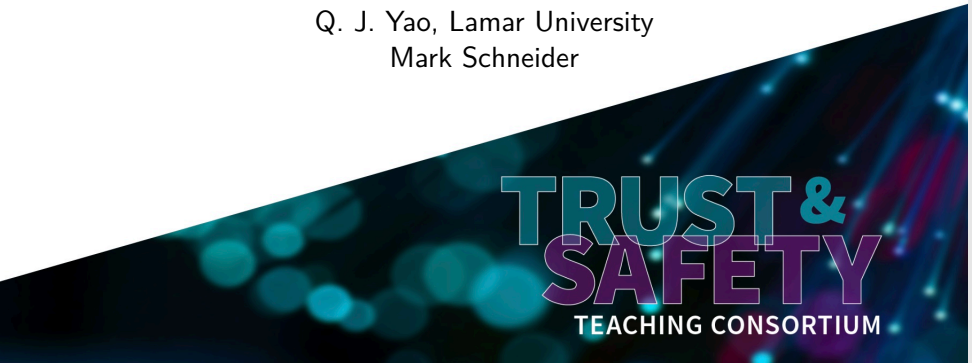
Harassment and Hate Speech

Peter Carragher, adapted from Ina Kamenova, University of Massachusetts Lowell
Q. J. Yao, Lamar University
Mark Schneider

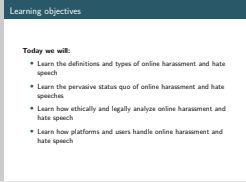


Harassment and Hate Speech

Peter Carragher, adapted from Ina Kamenova, University of Massachusetts Lowell
Q. J. Yao, Lamar University
Mark Schneider

Learning objectives

- Learn the definitions and types of online harassment and hate speech
- Learn the pervasive status quo of online harassment and hate speeches
- Learn how ethically and legally analyze online harassment and hate speech
- Learn how platforms and users handle online harassment and hate speech



A Definition of Online Harassment

- Definition: "interpersonal aggression or offensive behavior(s) that is communicated over the internet or through other electronic media."(Slaughter & Newsman 2022)

A Definition of Online Harassment
• Definition: "interpersonal aggression or offensive behavior(s) that is communicated over the internet or through other electronic media."(Slaughter & Newsman 2022)

The Online Harassment Experience Questionnaire (OHEQ; for users to assess online harassment; Slaughter & Newsman 2022)

- Non-traumatic questions: “I was impersonated by someone.” “I was excluded from an online group.” “Offensive or hurtful comments were directed at me or posted about me or I was insulted.called names.” “Someone spread untrue rumors about me.”
- Potentially traumatic questions: “Someone threatened to harm me.” “I experienced unwanted sexual attention.” “My personal information was posted online where other could access it.” “Someone hacked, stole, or otherwise gained access to my online account(s) without my permission.”
- Notes: those eight items are measured on a 6-point scale: 0 = never; 1 = less than once a month; 2 = 2 to 4 times a month; 3 = 2-4 times a week; 4 = daily; 5 = multiple times a day.

The Online Harassment Experience Questionnaire (OHEQ; for users to assess online harassment; Slaughter & Newsman 2022)
- Non-traumatic questions: “I was impersonated by someone.” “I was excluded from an online group.” “Offensive or hurtful comments were directed at me or posted about me or I was insulted.called names.” “Someone spread untrue rumors about me.”
- Potentially traumatic questions: “Someone threatened to harm me.” “I experienced unwanted sexual attention.” “My personal information was posted online where other could access it.” “Someone hacked, stole, or otherwise gained access to my online account(s) without my permission.”
- Notes: those eight items are measured on a 6-point scale: 0 = never; 1 = less than once a month; 2 = 2 to 4 times a month; 3 = 2-4 times a week; 4 = daily; 5 = multiple times a day.

The State of Online Harassment (Pew Center 2021)

- Surveyed 10,093 American adults in September, 2020
- Roughly four-in-ten have experienced online harassment, half due to political reasons and half experiencing more severe behaviors
- Specifically, harassment includes:
  - More severe forms: physical threats (14%); stalking (11%); sustained harassment (11%); sexual harassment (11%); all percentages raising from the past years
  - Less severe forms: offensive name-calling (31%); purposeful embarrassment (26%); all percentages raising from the past years

https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/

The State of Online Harassment (Pew Center 2021)
- Surveyed 10,093 American adults in September, 2020
- Roughly four-in-ten have experienced online harassment, half due to political reasons and half experiencing more severe behaviors
- Specifically, harassment includes:
  - More severe forms: physical threats (14%); stalking (11%); sustained harassment (11%); sexual harassment (11%); all percentages raising from the past years
  - Less severe forms: offensive name-calling (31%); purposeful embarrassment (26%); all percentages raising from the past years

Link to the course:
https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/

The State of Online Harassment (Pew Center 2021, Cont.)

- Reasons for being a victim: Political reasons (50%); gender (33%); race/ethnicity (29%); religion (19%); sexual orientation (16%); all percentages raising from the past years
- Venues of online harassment: Social media site (75%); forum/discussion site (25%); texting/messaging app (24%); online gaming (16%); personal email account (11%); online dating site/app (10%); multiple locations (41%).
- 32% believe social media addressing harassment poorly and 47% think of it only fair. They proposed methods to reduce harassment: 1). Permanent suspension (51%); real identity disclosure (48%); criminal charges (43%); proactive deletion (40%); temporary suspension (32%).

The State of Online Harassment (Pew Center 2021, Cont.)
- Reasons for being a victim: Political reasons (50%); gender (33%); race/ethnicity (29%); religion (19%); sexual orientation (16%); all percentages raising from the past years
- Venues of online harassment: Social media site (75%); forum/discussion site (25%); texting/messaging app (24%); online gaming (16%); personal email account (11%); online dating site/app (10%); multiple locations (41%).
- 32% believe social media addressing harassment poorly and 47% think of it only fair. They proposed methods to reduce harassment: 1). Permanent suspension (51%); real identity disclosure (48%); criminal charges (43%); proactive deletion (40%); temporary suspension (32%).

A Framework of Online-Harassment Assessment for Platform Regulation (Practices of Patio)

- When to moderate:
  1. The following correspondence between the violator and the platform moderators is rough
  2. Platform users or the victim of the violation has no other way to address the violation

A Framework of Online-Harassment Assessment for Platform Regulation (Practices of Patio)
- When to moderate:
  1. The following correspondence between the violator and the platform moderators is rough
  2. Platform users or the victim of the violation has no other way to address the violation

A Framework of Online-Harassment Assessment for Platform Regulation (Practices of Patio, Cont.)

- Three criteria to decide if moderation (social sanction) is needed (Foley & Gurakar, 2022):
  1. Intensity: severity of the violation
  2. Specificity: the size of the target(s) of the violation; the more specific the more danger the violation is
  3. Persistence: frequency of the violation

A Framework of Online-Harassment Assessment for Platform Regulation (Practices of Patio, Cont.)
- Three criteria to decide if moderation (social sanction) is needed (Foley & Gurakar, 2022):
  1. Intensity: severity of the violation
  2. Specificity: the size of the target(s) of the violation; the more specific the more danger the violation is
  3. Persistence: frequency of the violation

Online Hate Speech

Ina Kamenova, ABD, LCMHC

Online Hate Speech

What is hate speech?

What is hate speech?

"a kind of speech act that contextually elicits certain detrimental social effects that are typically focused upon subordinated groups in a status hierarchy"

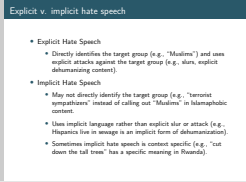
(Demuyakor, 2020, p. 1014; United Nations, 2019, p. 2)

Definitions are variable and highly contested as legal concepts differ across EU and US and the landscape is shifting.

Hietanen, M., & Eddebo, J. (2022). Towards a Definition of Hate Speech—With a Focus on Online Contexts. Journal of communication Inquiry, 01968599221124309.

Explicit v. implicit hate speech

- Explicit Hate Speech
  - Directly identifies the target group (e.g., “Muslims”) and uses explicit attacks against the target group (e.g., slurs, explicit dehumanizing content).
- Implicit Hate Speech
  - May not directly identify the target group (e.g., “terrorist sympathizers” instead of calling out “Muslims” in Islamaphobic content.
  - Uses implicit language rather than explicit slur or attack (e.g., Hispanics live in sewage is an implicit form of dehumanization).
  - Sometimes implicit hate speech is context specific (e.g., “cut down the tall trees” has a specific meaning in Rwanda).



How is hate speech different from harassment?

- The message applies to a characteristic shared by a group of people
- The impact is broader, described as concentric circles of hate impact.
- The threat is potentially more unpredictable and unavoidable
- There can be intersections between harassment and hate speech, such as when targeted harassment includes hate speech, or when online hate speech also leads to harassment campaigns

How is hate speech different from harassment?
- The message applies to a characteristic shared by a group of people
- The impact is broader, described as concentric circles of hate impact.
- The threat is potentially more unpredictable and unavoidable
- There can be intersections between harassment and hate speech, such as when targeted harassment includes hate speech, or when online hate speech also leads to harassment campaigns

Castaño-Pulgarín, S. A., Suárez-Betancur, N., Vega, L. M. T. & López, H. M. H. (2021). Internet, social media and online hate speech. Systematic review. *Aggression and Violent Behavior, 58*(101608), 101608. https://doi.org/10.1016/j.avb.2021.101608

Examples of hate speech categories

- Religious Hate Speech
- Racist Speech
- Gender and Sexuality Hate Speech

Examples of hate speech categories
• Religious Hate Speech
"inflammatory and sectarian language to promote hatred and violence against people on the basis of religious affiliation through the cyberspace" (Castaño-Pulgarín et al., 2021)
• Racist Speech
As with all hate speech, racist messages can be direct (using slurs and describing characteristics associated with race unfavorably) or indirect: alluding to purity, nationality and other coded concepts.
• Gender and Sexuality Hate Speech

Comparative Legal Frameworks

Comparative Legal Frameworks

US

In the US racist, sexist, and other hateful language online is not a crime. It can matter in criminal settings to categorize another crime as a hate crime and the offender may receive a hate crime enhancement to their sentence.

UK & Europe

The UK and Europe have much more stringent laws that make certain kinds of speech a crime in themselves.

Freedom of Speech Exceptions in the US

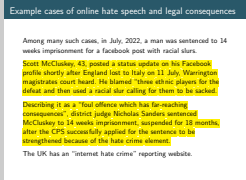
Freedom of Speech Exceptions in the US

The exceptions are very narrow and refer to criminal sanctions. Private institutions can regulate speech as they choose and have routinely done so.

Example cases of online hate speech and legal consequences

Scott McCluskey, 43, posted a status update on his Facebook profile shortly after England lost to Italy on 11 July, Warrington magistrates court heard. He blamed "three ethnic players for the defeat and then used a racial slur calling for them to be sacked.

Describing it as a "foul offence which has far-reaching consequences", district judge Nicholas Sanders sentenced McCluskey to 14 weeks imprisonment, suspended for 18 months, after the CPS successfully applied for the sentence to be strengthened because of the hate crime element.



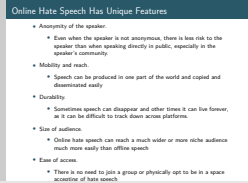
https://www.report-it.org.uk/reporting_internet_hate_crime

https://www.theguardian.com/world/2021/sep/08/man-sentenced-over-racist-post-after-euro-2020-final

https://www.cps.gov.uk/mersey-cheshire/news/cheshire-man-sentenced-racist-abuse-england-players

Online Hate Speech Has Unique Features

- Anonymity of the speaker.
  - Even when the speaker is not anonymous, there is less risk to the speaker than when speaking directly in public, especially in the speaker's community.
- Mobility and reach.
  - Speech can be produced in one part of the world and copied and disseminated easily
- Durability.
  - Sometimes speech can disappear and other times it can live forever, as it can be difficult to track down across platforms.
- Size of audience.
  - Online hate speech can reach a much wider or more niche audience much more easily than offline speech
- Ease of access.
  - There is no need to join a group or physically opt to be in a space accepting of hate speech



Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities*, 18(3), 297–326. https://doi.org/10.1177/1468796817709846

Impact of online hate speech

- Direct harm to targeted people and groups of people
  - E.g. psychological distress, defamation, social repercussions (e.g. in outing someone)
- Link with violent attacks
  - E.g. linked with increasing hate crime and hateful ideology radicalization, which is thought to be related to increased single-actor attacks precipitated by participation in online, channels and other spaces elaborating on hateful ideologies or directly attacking people with targeted characteristics.

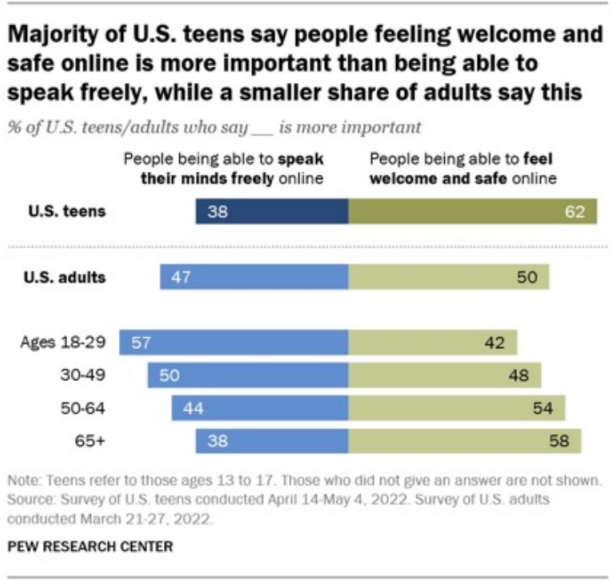
Impact of online hate speech
- Direct harm to targeted people and groups of people
  - E.g. psychological distress, defamation, social repercussions (e.g. in outing someone)
- Link with violent attacks
  - E.g. linked with increasing hate crime and hateful ideology radicalization, which is thought to be related to increased single-actor attacks precipitated by participation in online, channels and other spaces elaborating on hateful ideologies or directly attacking people with targeted characteristics.

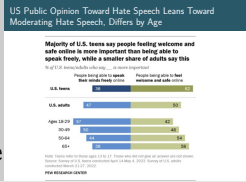
Impact of online hate speech

- Influences Political Discourse
  - E.g. Hate speech online normalizes hateful ideologies, influences electoral process, in turn legislation and even judicial decisions.
- Disincentivizes others from participating in the discourse
  - E.g. people with targeted characteristics will be less likely to engage in public discourse knowing they will become targets for hate speech and harassment. Other people may also refrain because they do not want to read hateful content.

Impact of online hate speech
- Influences Political Discourse
  - E.g. Hate speech online normalizes hateful ideologies, influences electoral process, in turn legislation and even judicial decisions.
- Disincentivizes others from participating in the discourse
  - E.g. people with targeted characteristics will be less likely to engage in public discourse knowing they will become targets for hate speech and harassment. Other people may also refrain because they do not want to read hateful content.

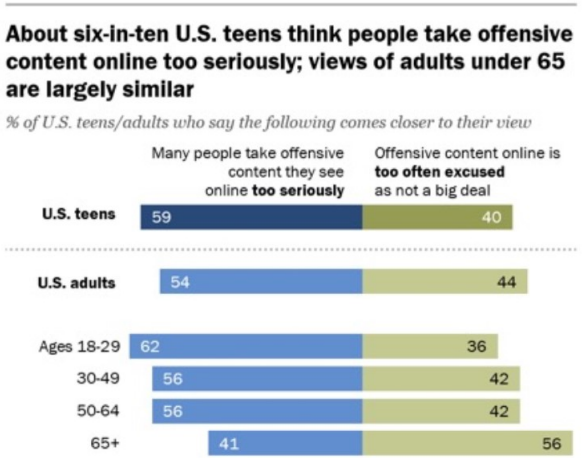
US Public Opinion Toward Hate Speech Leans Toward Moderating Hate Speech, Differs by Age

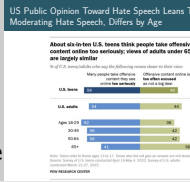




https://www.pewresearch.org/fact-tank/2022/08/30/more-so-than-adults-u-s-teens-value-people-feeling-safe-online-over-being-able-t

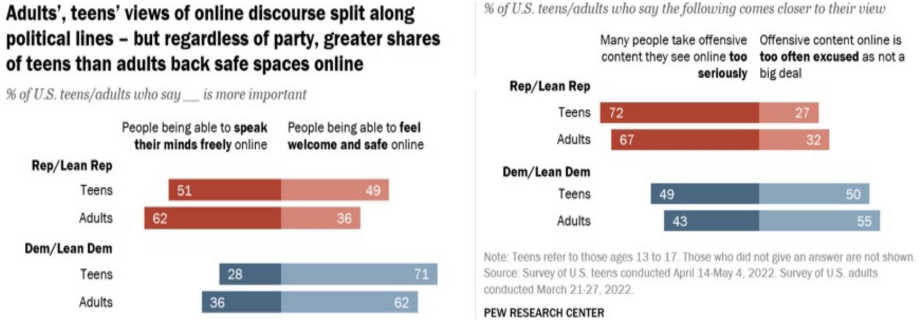
US Public Opinion Toward Hate Speech Leans Toward Moderating Hate Speech, Differs by Age

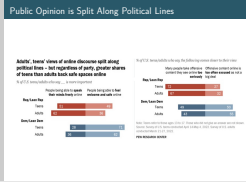




https://www.pewresearch.org/fact-tank/2022/08/30/more-so-than-adults-u-s-teens-value-people-feeling-safe-online-over-being-able-t

Public Opinion is Split Along Political Lines





Platform Response

- Hate speech, or hateful speech based on identity characteristics is mentioned in various ways in the content policy of platforms.
- Check current platform policies for most up-to-date information.
- These policies and practices are also changing as new legislation take effect in the EU and US.

Platform Response
- Hate speech, or hateful speech based on identity characteristics is mentioned in various ways in the content policy of platforms.
- Check current platform policies for most up-to-date information.
- These policies and practices are also changing as new legislation take effect in the EU and US.

Main Questions Investigated Differ by Discipline

- In the United States, legal scholars primary ask questions about free speech and the limits of free speech.
- Another currently contested area for legal scholars and legislators are regulatory expectations of platforms and to what extent offline speech principles apply such as the “public square” doctrine.

- Considers social and individual impacts of hate speech. Research shows that there are similar effects to other traumatic events and effects vary by the type and amount of exposure.


Main Questions Investigated Differ by Discipline

Law

- In the United States, legal scholars primary ask questions about free speech and the limits of free speech.
- Another currently contested area for legal scholars and legislators are regulatory expectations of platforms and to what extent offline speech principles apply such as the “public square” doctrine.

Psychology & Sociology

- Considers social and individual impacts of hate speech. Research shows that there are similar effects to other traumatic events and effects vary by the type and amount of exposure.

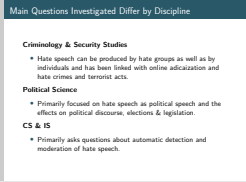
Often cited opinion pieces by legal scholars include: Balkin, Jack M., To Reform Social Media, Reform Informational Capitalism (September 6, 2021). Social Media, Freedom of Speech and the Future of Our Democracy; Lee Bollinger and Geoffrey R. Stone, eds., Forthcoming, Available at SSRN: https://ssrn.com/abstract=3925143 or http://dx.doi.org/10.2139/ssrn.3925143

https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/the-ongoing-challenge-to-define-free-speech/in-the-age-of-socia-media-first-amendment/

https://scholarlycommons.law.case.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1076&context=jolti

Main Questions Investigated Differ by Discipline

- Hate speech can be produced by hate groups as well as by individuals and has been linked with online adicaization and hate crimes and terrorist acts.
- Primarily focused on hate speech as political speech and the effects on political discourse, elections & legislation.
- Primarily asks questions about automatic detection and moderation of hate speech.



Often cited opinion pieces by legal scholars include: Balkin, Jack M., To Reform Social Media, Reform Informational Capitalism (September 6, 2021). Social Media, Freedom of Speech and the Future of Our Democracy; Lee Bollinger and Geoffrey R. Stone, eds., Forthcoming, Available at SSRN: https://ssrn.com/abstract=3925143 or http://dx.doi.org/10.2139/ssrn.3925143

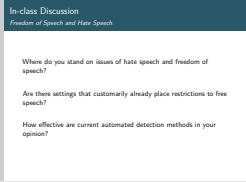
https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/the-ongoing-challenge-to-define-free-speech/in-the-age-of-socia-media-first-amendment/

https://scholarlycommons.law.case.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1076&context=jolti

In-class Discussion

Freedom of Speech and Hate Speech.

In-class Discussion
Freedom of Speech and Hate Speech.

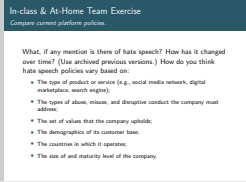


Customize and flesh out these answers as needed and pair students or assign small groups with specific readings and handouts based on the readings.

In-class & At-Home Team Exercise

Compare current platform policies.

- The type of product or service (e.g., social media network, digital marketplace, search engine);
- The types of abuse, misuse, and disruptive conduct the company must address;
- The set of values that the company upholds;
- The demographics of its customer base;
- The countries in which it operates;
- The size of and maturity level of the company.



Links to the current policies of several platforms: https://www.tiktok.com/community-guidelines?lang=en#39

Sources of Information

- Journal of Online Trust and Safety
- Arbiters of Truth Series of the Lawfare Podcast

Sources of Information

Since this a rapidly developing field, the most recent information on new scholarship and information can be very field-specific. Below is a list of sources that are have specifically dedicated to Online Trust & Safety issues, including free speech.

- Journal of Online Trust and Safety
- Arbiters of Truth Series of the Lawfare Podcast

Areas for Further Study & Recent Issues

- Repeated calls for interdisciplinary collaboration to make progress on these issue.
- The magnitude of the issue is difficult to pin down and scholars are calling for more research.
- Multi-platform cooperation and academic-industry collaborations.
- Moral and legal principles are continuously under debate alongside technological capabilities to execute desired policies for intervention and prevention

- Legislation debates
- Tech industry layoffs of teams responsible for hate speech detection and moderation

Areas for Further Study & Recent Issues
Further Study
- Repeated calls for interdisciplinary collaboration to make progress on these issue.
- The magnitude of the issue is difficult to pin down and scholars are calling for more research.
- Multi-platform cooperation and academic-industry collaborations.
- Moral and legal principles are continuously under debate alongside technological capabilities to execute desired policies for intervention and prevention

Recent issues
- Legislation debates
- Tech industry layoffs of teams responsible for hate speech detection and moderation