I'm working on a list of the different forms and definitions of online hate and harassment, as well as the various forms of intervention.

Vectors:

- In the study *Counterspeech on Twitter: A Field Study* the authors identify a distinction in counterspeech conversations that is useful more generally in looking at other forms of contentious speech online. These vectors are based on the number of participants in each stage or side of an exchange. By classifying the different exchanges, it is easier to assess the different nature and consequences of the speech as well as the effective forms of intervention. The four types of exchanges are:
    - One-to-one: one person deploying counterspeech against one person's hate speech
    - One-to-many: one person deploying counterspeech against many people's hate speech
    - Many-to-one: many people deploying counterspeech against one person's hate speech
    - Many-to-many: many people deploying counterspeech against many people's hate speech
        - This is especially prevalent with temporal clustering around a publicized incident or a hashtag
- While the vectors' definitions outline the exchanges as they relate to counterspeech, they could also be useful in assessing who is being targeted by hateful speech.

Types of responses to hate speech:

- Inoculation is a long-term method for fighting against hate speech that takes some time. It involves instilling values in a society that oppose hate speech, and deals especially with building the social-psychological tools necessary so that groups of people don't fall victim to the pressures of engaging in hate speech or being incited by it.
    - An example of a group that deals with Inoculation is Radio la Benevolencija (RLB) a dutch nonprofit that produces entertainment for countries in central africa that deals with the psychology underlying incitement to hate and violence.
- Citron and Norton suggest that internet intermediaries and society at large play a stronger role in fostering digital citizenship http://web.a.ebscohost.com.ezproxy.cul.columbia.edu/ehost/pdfviewer/pdfviewer?vid=1&sid=1c840371-b7ff-4ba7-a74e-8848b5bae30a%40sessionmgr4008
- Counterspeech : https://dangerousspeech.org/counterspeech/
    - While inoculation and fostering digital citizenship are longer processes, Counter Speech is a more immediate-term solution which seeks to effectively refute hate speech and forestall its potential consequences.
    - There are two types of counterspeech: organized counter-messaging campaigns and spontaneous, organic responses. For more information see my summaries *Counterspeech on Twitter: A Field Study* and *Considerations for Successful Counterspeech*.

    - Counterspeech can be very effective when done by someone who the speaker affiliates with
- Calling out or Doxxing

Hate Speech:

- “An expression that denigrates or stigmatizes a person or people based on their membership of a group that is usually but not always immutable, such as an ethnic or religious group. Sometimes other groups, defined by disability or sexual orientation, for example, are included.” Source: Benesch, Susan. *Defining and diminishing hate speech*. 2014.

Hateful Speech:

- The authors of *Counterspeech on Twitter: A Field Study* define hateful speech as “speech which contains an expression of hatred on the part of the speaker/author, against a person or people, based on their group identity” (13). This is distinguished from **hate speech** by focussing on the expression of hatred rather than the intent of hatred.

Dangerous Speech

- The authors of *Counterspeech on Twitter: A Field Study* define dangerous speech as “speech that can inspire or catalyze intergroup violence” (13).
- Benesch also puts forward the following conditions that make the likelihood of speech resulting in group violence :
    - there is a “powerful speaker with a high degree of influence;”
    - there is a receptive audience with “grievances and fear that the speaker can cultivate;”
    - a speech act “that is clearly under-stood as a call to violence;”
    - a social or historical context that is “propitious for violence, for any of a variety of reasons;”and
    - An “influential means of dissemination.”” (17)
-

Dogpiling

Deadnaming-referring to someone by their name given at birth instead of their preferred name. Usually done against trans people intentionally to inflict harm

Flesh search engine

Geoblocking https://en.wikipedia.org/wiki/Geo-blocking

Blackmail

Cyber bullying

Cyber stalking

Sexual Predation

Revenge Porn/Non consensual pornography (NCP)

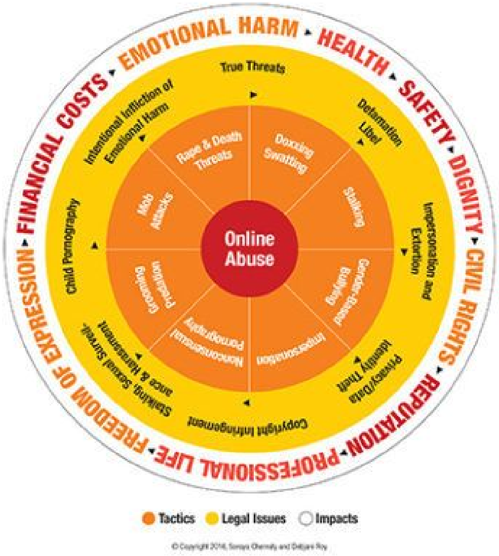
Threats

Identity based discrimination

Roasting

Women's Media Center has a super useful primer on Online Abuse where they list and define a variety of forms of online abuse:

www.womensmediacenter.com/speech-project/online-abuse-101/



****
These are Cyber-Harassment Typologies from Digital Rights Foundation that they used to classify the types of calls that their cyber harassment helpline received over its first six months.

Source:
https://digitalrightsfoundation.pk/wp-content/uploads/2017/07/Cyber-Harassment-Helpline-Six-Month-Report.pdf

Fake Profile Fake profile or impersonation is when someone's identity is appropriated without their permission. This manifests itself in profiles purporting to or belonging to someone on social media websites, and contacting people through texts or calls pretending to be someone else

Blackmailing This often involves using personal information or psychological manipulation to make threats and demands from the victim.

Unsolicited Messages are unwanted and repeated contact with someone which may include spam, repeated requests for contact personalised threats, blackmail, or any unwanted messages that make the receiver feel uncomfortable.

Hacking Gaining unauthorized access to someone's electronic system, data, account and devices

Non-Consensual Usage of Information: This involves using, sharing, disseminating, and manipulating data such as photographs, phone numbers, contacts, and other personal information without consent and in violation of the privacy of a person

Online Stalking Online stalking is keeping track of someone's online activity in a way that it makes the subject of the stalking uncomfortable. For the purpose of this report, online stalking also refers to (repeatedly) contacting a person's friends and/or family.

Doxxing Doxxing is the practice of leaking and publishing information of an individual that includes personally identifiable information. This information is meant to target, locate and contact and individual usually through social media, discussion boards, chat rooms and the like; and more often than not, is accompanied by cyberbullying and cyberstalking.

Gender-based Bullying Any actions, statements, and implications that targets a person based on their gender identity or sexual orientation. Evaluation for gender-based bullying takes into account the overall connotations attached to actions and verbal communications within the larger system of gendered oppression and patterns of behaviour that signify abuse.

Non-Consensual Pornographic Pictures This is obtaining, using, distributing or retaining pictures, videos or graphic representations without a person's consent that violate their personal dignity

Financial Fraud Intentional actions of deception perpetrated by a person for the purpose of financial gain and profit; this includes using someone's financial data to gain access to accounts and make purchases.

Non-Consensual Photoshopped Pictures The manipulation, distortion or doctoring of images without the permission of the persons to whom they belong. This is often accompanied by distribution and sharing of such pictures as well.

Non-Cooperation from Social Media Platforms: These complaints refer to a situation when a person has reported a case of cyber harassment to the relevant social media team, but has not received a decision in their favor

Threats These are all threats directed at the victim of online harassment that do not fall under the category of gender-based threats or sexual/physical violence

Defamation Any intentional, false communication purporting to be a fact that harms or causes injury to the reputation of a natural person

Hate Speech Any communication that targets or attacks an individual on the basis of their race, religion, ethnic origin, gender, nationality, disability, or sexula orientation. Hate speech becomes a matter of urgent action when it puts its target in physical danger or the reasonable apprehension of physical danger. However hate speech is not restricted to just incitement to violence, it is hate speech if it leads to the exclusion of or creation of a hostile online environment for its target.