



INTRODUCTION TO HATE SPEECH ON SOCIAL MEDIA

What is hate speech?

This is a tricky question, since there's not a single internationally accepted definition of hate speech. But in broad terms, hate speech is a communication that denigrates people on the basis of their membership to a particular group. This can include any form of expression, such as images, plays and songs as well as speech. Some definitions even extend the concept of hate speech to include communications that foster a climate of prejudice and intolerance - the thinking here is that these kinds of communications may fuel discrimination, hostility and violent attacks later on. The International Covenant on Civil and Political Rights (ICCPR), UN treaty, calls on governments to prevent hate speech. Article 20(2) says: "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law". Working towards a consensus on hate speech, as of end of May 2016, Facebook, Twitter, Youtube and Microsoft, signed a code of Conduct on online hate speech with the UN.

Is hate speech illegal?

In many countries, yes, but it depends on how the country defines hate speech, if it does at all. Hate speech laws are a relatively modern phenomenon that appeared in Europe in the wake of World War II. The idea behind such laws was to curb the kinds of anti-Semitic and racist propaganda that gave rise to the Holocaust. Germany, Poland, Hungary and Austria passed hate legislation decades ago. Many other countries have since followed suit. For example, under Kenyan law, a person commits an offense if they stir up "ethnic hatred".

What about online hate speech?

Hate speech often shows up online,

"Hate Speech attacks a person or a group on the basis of race, religion, gender or sexual orientation. It is not freedom of speech. The International Covenant on Civil and Political Rights (ICCPR) requires states to prohibit Hate Speech"

EXERCISE 1.

List examples of different forms of hate speech that you have seen in the news & social media, and ask yourself the following:

What is the difference between offensive speech and hate speech?

Because there is no one definition of hate speech, it is sometimes difficult to judge what is just an offensive comment and what is hate speech. But basically a nasty comment about an individual isn't hate speech, unless it targets that person as a member of a particular group.

What are the effects of hate speech?

There are real-world examples of hate speech having disastrous, deadly results. Recent examples include:

- 1994 genocide in Rwanda, where it is widely believed that hate speech played a significant role in the massacre of 800,000 Tutsis and Hutus.
- In the aftermath of the December 2007 presidential elections in Kenya, violence erupted, mainly between Kenya's three largest ethnic groups. More than 1,100 people were killed. A popular radio broadcaster, Joshua Arap Sang, was accused of using his position to encourage ethnic attacks. Text messages were widely circulated calling on one group or another to "exterminate" ethnic rivals. Since then, Kenya has passed new laws prohibiting hate speech.

especially on social media. Facebook, Twitter and Google each has its own specific definition of hate speech and their approaches to dealing with it are evolving. Facebook's rules forbid bullying, harassment and threatening language (although critics say it does not always enforce these rules properly). Twitter: In 2015, the social media platform banned speech that could incite terrorism, or violence against people "on the basis of race, ethnicity, national origin, religion, sexual orientation, gender, gender identity, age, or disability."

New EU guidelines on hate speech and the framework legal decision on combating expressions of racism and xenophobia were signed by the major social media companies in Brussels on 31st May 2016.

Links between social media usage and offline activity reflected in-country: e.g. from Facebook to mobile phones to radio, graffiti and word of mouth.

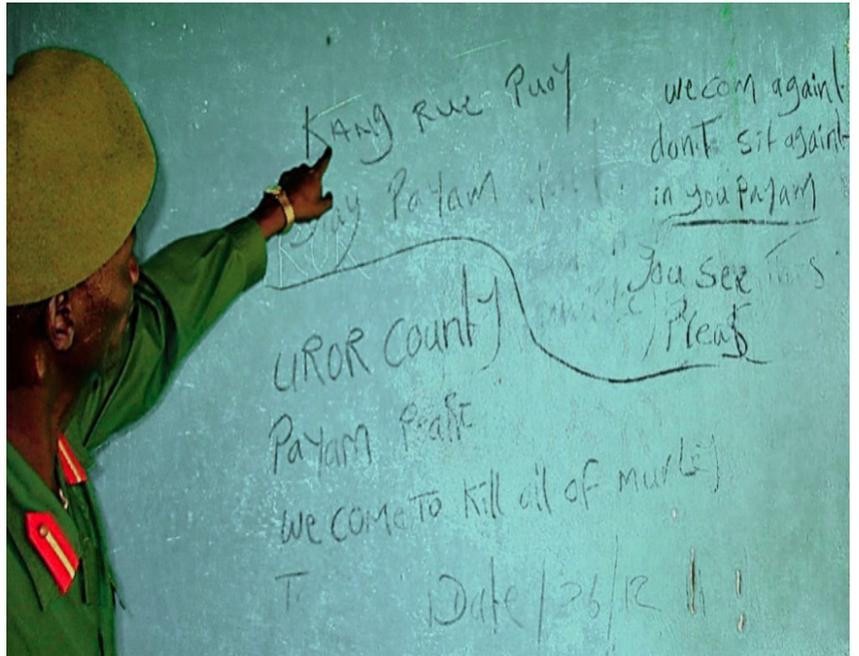


PHOTO CREDIT:Ferdinand von Habsburg-Lothringen

Dangerous Speech : A Proposal to Prevent

There are at least three ways in which inflammatory speech presents an opportunity for violence prevention, and therefore three distinct applications for the guidelines. First, such speech can serve as a key indicator for early warning, since it is often a precursor - if not also a prerequisite - for mass violence. Second, it may be possible to limit violence by finding ways to limit such speech or its dangerousness. Third, speakers may be held accountable for speech that constitutes crime.

The guidelines are designed to identify a subset of hate speech, which I have termed Dangerous Speech. Hate speech is variously defined in law and in common parlance, but is generally understood to mean speech that denigrates people on the basis of their membership in a group, such as an ethnic or religious group. This category of speech is too broad for successful early warning of mass atrocities, for two related reasons.

First hate speech is common in many societies, unfortunately, including those at minimal risk of genocide. Second, some hate speech does not appreciably increase the risk of mass violence, although it may cause serious emotional and psychological damage. In other words, speech can harm directly or indirectly, or both. It may directly offend, denigrate, humiliate or frighten the people it purports to describe – such as when a racist shouts at a person of color. Speech can also bring about harm indirectly - and with equal or even greater brutality – by motivating others to think and act against members of the group in question.

When an act of speech has a reasonable chance of catalyzing or amplifying violence by one group against another, given the circumstances in which it was made or disseminated, it is Dangerous Speech.

Benesch's model includes five variables for analyzing the dangerousness of hate speech:

1. The degree of the speaker's influence over an audience
2. The grievances or fears of the audience that can be cultivated by the speaker
3. Whether or not the speech act is understood as a call to violence
4. The social and historical context (such as previous episodes of violence)
5. Whether the means distributing the speech is also influential (such as when a media outlet is the sole broadcaster of information in that area)

EXERCISE

Determine the level of influence, context and intention of the hate speech listed earlier.

3 Primary Categories of Online Incitement :

1. Emotional

Buzz - lack of social media ethics - unclear of public effects of online activity.

2. Personal or group virulence

De-humanising the other, propaganda, image/fact manipulation.

3. Organized, aimed, directed Incitement to violence

Hoaxing / rumour to spark violence or armed action, potential genocide.

Drivers of Social Media Hate Speech :

- Children affected by or involved in violence at increasingly earlier ages - proximity to conflict, learning to hate, breakdown of positive norms/values.
- Youth feel frustration, social media offers an open platform, which can further entrench hate.
- Lack of accountability - no reconciliation, power of anonymity / distance.
- Political / tribal alignments - misconception on cultural diversity
- Lack of policies to guide on online freedom of expression, ethics and privacy
- See : <http://www.unesco.org/new/en/internetstudy>

- **Online reactions amplify tensions between citizens/groups increasingly splintering communities further, even within diaspora.**
- **Vicious circle of increasing brutality, dangerous and aggressive speech and directed incitement moving between inaccurate reporting of the conflict to de-contextualisation in social media.**
- **Increasing cases of networking between remote diaspora users. (e.g. in USA and Australia) using social media platforms to organize campaigns of directed incitement on the ground in South Sudan.**

#DefyHateNow Initiative

How to intervene without censorship?

- Remove abstraction of online hate: integrate monitoring, engaging, verifying, illustrating and exposing effects of hate coming from outside into peacebuilding activities.
- Urgency to act : target UNMISS / UNCT youth Peacebuilding activities e.g. UNESCO YouthMobile, UNICEF Youth Innovation, UNESCO Youth Peacemaker Network (YPN).
- Say NO, engage and offer positive messaging to counter intolerance whenever possible
- Verify the social media activities of high profile & spokespeople, ambassadors' and 'celebrities'!!

LINKS

FAQ: Hate Speech <http://www.dw.com/en/hate-speech-a-faq/a-19103744>

Guidelines on reporting hate speech – practical tips for journalists / DW Akademie #MediaDev
http://www.dw.com/en/reporting-hate-speech-practical-tips-for-journalists/a-19152896?maca=en-rss_en_akademie_mediadev-15969-xml-mrss

EU Guidelines: Hate Speech Code of Conduct
http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf

Framework legal decision on combating certain forms and expressions of racism and xenophobia by means of criminal law EU Brussels 31st May 2016
<http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3Al33178>

New EU Guidelines on Hate Speech
http://europa.eu/rapid/press-release_IP-16-1937_en.htm
<http://www.telegraph.co.uk/technology/2016/05/31/facebook-and-twitter-promise-to-crack-down-on-internet-hate-speech/>

Where can I learn more about hate speech?

Countering Online Hate Speech: Focus on the Internet, UNESCO report has good background information on hate speech.
<http://unesdoc.unesco.org/images/0023/002332/233231e.pdf>

Factsheet – Hate Speech: This document by the European Court of Human Rights looks at a variety of real-world cases regarding cases concerning incitement to hatred and freedom of expression.
http://www.echr.coe.int/Documents/FS_Hate_speech_ENG.pdf

Hate Speech: Article by press freedom group Article 19 exploring hate speech in international and domestic law.
[tps://www.article19.org/pages/en/hate-speech-more.html](https://www.article19.org/pages/en/hate-speech-more.html)

#defyhatenow is an urgent community peacebuilding, training and conflict reconciliation project aiming to strengthen the voices and support the actions of peace & youth oriented civil society organisations in South Sudan.

#DefyHateNow aims to identify and develop mitigation responses in 3 key areas and target groups

- Awareness - offline grassroots level / peace mobilizers / students
- Social Media Literacy - youth / IT training / teachers / NGOs / lawyers
- Counter-Messaging Skills - correspondents / monitors / experts.



#DefyHateNow is an initiative to combat social media hate speech by the r0g_agency for open culture and critical transformation, Berlin, and the Community Empowerment for Progress Organisation (CEPO), Juba, funded by the ifa – Institut für Auslandsbeziehungen (zivik) with means of the German Federal Foreign Office.



Images, videos, sound recordings, reports and other media produced by the project shall bear the name of the author (and institutional affiliation where applicable) and are released under a Creative Commons Attribution - Share Alike 4.0 International License (CC BY 4.0)