PEACETECH LAB

COUNTERING HATE SPEECH STRATEGIES & BEST PRACTICES

A TRAINING MANUAL

Table of Contents

Foreword

When our team at the PeaceTech Lab began implementing the Countering Hate Action Network (CHAN) program in 2021, we did so with the aspiration to form new partnerships and to continue fostering the expansion of technology for social good — both fundamental to the mission of the PeaceTech Lab. The CHAN program is central to this mission. CHAN participants span 13 countries in Africa and the Middle East and work on a diverse set of issues in their communities. Through the use of low-cost, high-impact technology, participants have gone back to their communities better equipped to detect, deactivate, and defend against hate speech.

The purpose of this workbook is to reflect upon the lessons learned from the first two sessions of the CHAN program and the lexicon research used to inform its content. In this workbook, we review the four modules and numerous case studies that were used in the CHAN program to build participants' ability to identify hate speech, use low-cost technologies to counter it, and create long-lasting community resiliency. With the materials and exercises found in this workbook, we hope program coordinators and practitioners who did not participate in the CHAN program may be similarly equipped to address the complexities associated with countering hate speech online and offline.

I am incredibly proud of the participants, facilitators, and staff that made the CHAN program successful. Without their thoughtful contributions and diligence, this program would not have been successful. Their excellent work has been recognized in local media, beneficiary communities, and in the Cooper Hewitt Smithsonian Design Museum's recent exhibit on *Designing Peace*, in which the *Hate Speech Lexicons* research is highlighted. This workbook is a testament to the efforts of these individuals, a celebration of the program's accomplishments, and a powerful tool for future peacebuilding practitioners.

Sheldon Himelfarb
Founder & CEO, PeaceTech Lab

Glossary

Actor mapping: a situational analysis of a particular context or conflict that shows who the actors are in the situation including their roles and how various actors are interrelated. Should precede narrative analysis in any specific context.

Advocacy: an umbrella term that describes collective and strategic engagement with certain stakeholders to get their attention, support, or move them to certain actions.

- **Community advocacy:** involving community leaders and public figures in advocacy can give campaigns a powerful boost. The participation of community leaders and decision makers can increase the likelihood that the public will participate in campaigns.

- **Social Media advocacy:** working with social media companies to 1) verify content, 2) flag false information (e.g., COVID, election rumors), and 3) redirect the user to more trustworthy sources or psycho-social services.

- **Legal advocacy:** using legal or regulatory leverage to advance the cause against hate speech. It may consist of engaging with (or pressuring) the national government to adopt a hate speech law (if there isn't one), improving existing law, improving the implementation of that law, improving regulations, and strengthening a company's own community guidelines. Legal advocacy may also use the judicial system to advance particular goals for change.

Amplification (social media): the technological capacity to increase the spread and reach of speech or content. This is increasingly a problem with hate speech, dangerous speech, and mis/disinformation.

Anti-discrimination laws and regulations: protection of human dignity including protection for religious belief systems or political entities.

Counterspeech: according to the Dangerous Speech Project, "is any direct response to hateful or harmful speech which seeks to undermine it." Counterspeech can be conducted individually or collectively.

Critical episodes ("catalyst events"): are events such as elections, violent incidents, natural disasters, assassinations, and other events that can polarize opinions at the regional or national level. They may also exacerbate tensions and conflict dynamics.

Dangerous speech: *"Dangerous Speech is any form of expression (speech, text, or images) that can increase the risk that its audience will condone or participate in violence against members of another group." (Susan Benesch)*

Data gathering: the process of collecting information about a topic or phenomenon. This can be achieved through various types of collection processes both human-led and automated.

Difficult content: online hate and hate speech as well as other forms of content that may cause trauma or difficulty for monitors and researchers.

Disinformation: false information that is spread by a person who knows that what they are sharing is ***not*** true. They spread false information intentionally, for the purpose of manipulating their audience for self-enrichment (e.g., gaining political advantage). Often, actors spreading disinformation use false 'news' to instigate their followers against a particular community.

Doxxing (doxing): when a person (or network of people) publishes another person's private, personal information in online forums with the intent of inciting others users to take the information and harass the target individual (doxx is an online shorthand for "documents").

EWER: Early Warning and Early Response (EWER) programs or guides are tools used in contemporary conflict prevention and peace building that show how states and state partners can advance commitments to work for peace and cohesion.

First responders: counterspeakers who are dispatched as a first line response to online hate speech. First responders also often act as data collectors and monitors.

Freedom of speech: the rights of individuals to express their ideas and beliefs, particularly ideas and beliefs that are political within the individual's context. These rights are often codified (in varying ways) in national laws or regulations. As a democratic ideal, particularly in relation to social and online media, freedom of speech adds levels of complexity to combatting hate speech.

Genocide: is recognized as a form of mass atrocity crime that is often preceded by increased hate speech and conflict. According to Sheri P. Rosenberg, *"Genocide is a process, not an event. It did not start with the gas chambers, it started with hate speech."*

Hate speech: is *any form of expression that seeks to promote hate and targets individuals or groups based on shared and innate characteristics.*

- **illegal hate speech** – *speech that has been categorized by legislation as unlawful speech or banned speech; legality will vary by national legal context.*

- **legal hate speech** – *speech that has not been categorized by legislation as "illegal" or "banned" speech, but which is still recognized as harmful*

Hate speech literacy: a type of situational awareness that enables individuals to better identify hate, dangerous, and inflammatory speech, particularly online. Hate speech literacy instruction is an important aspect of civic and peace education and should be an integral part of the discourse for combating hate speech and as a preventive measure.

Inflammatory speech (content): speech (content) that is intended to excite anger, disorder, conflict, or violence.

Information campaigns: programs to raise concerns about hate speech issues in public spaces to spread awareness, foster discussion about what hate speech is, and to inform people about a specific topic or issue.

In-person or digital dialogue: The aim of dialogue in countering hate speech is to engage diverse and divided people in a constructive conversation in order to break down stereotypes as well as rebuild trust and empathy.

Media literacy: a type of situational and informational awareness that enables individuals to better identify false information (both mis- and disinformation) and unreliable sources with an eye to increasing an individual's critical assessment capacities.

Misinformation: a form of false information where the person circulating the content believes that the information is true. For instance, many people who believe that climate change is a myth may spread false information including conspiracy theories, bad data, or other incorrect or misleading communication, because they think it is the truth.

Narrative analysis: approaches that provide a way to understand how stories function in communities and how to intervene when destructive stories circulate by examining what is being said in the hateful content to understand the stories and discourses that underpin it.

Nonviolent action: nonviolent action is the practice of achieving goals (such as political or social gains) through collective and nonviolent actions like mass protests, boycotts, strikes, sit-ins and various forms of non-cooperation.

- **Strategic nonviolent action:** When nonviolent action tactics are used in strategic ways with the goal of shifting the power dynamics at play on both sides of a given cause (or struggle).

Online footprint: all the sites where a person or organization maintains a presence online, has posted online materials, and posted material about themselves, particularly identifying information that can be utilized to harass or bully the person or organization.

Peace promotion: a practice that involves strengthening peace and non-violence through education, advocacy and media, amongst other activities. Digital campaigns and activities can be incorporated into already existing works to promote peace which includes on and offline activities.

Quantitative analysis: a form of data analysis that relies on numbers and statistics. Particularly useful for analyzing large amounts of data. Includes computational analysis of digital data.

Qualitative analysis: a form of data analysis that relies on language-based data such as stories, narratives, and interviews. Particularly useful for providing rich and nuanced analysis, typically used on small volume data.

Resilience building (youth / peer to peer): programmatic approaches to countering hate focused on youth engagement to build media literacy, resilience against hate and radicalization, as well as generating the capacity for direct interventions.

Social cohesion: a framework that allows us to understand the interactions between populations. It creates a space for shared advocacy by people within a common process of development. It also enables a better framework for responses to the issues involved in intercultural and interreligious dialogues. Social cohesion is based on individuals' abilities to interact with others to the benefit of society as a whole.

Stakeholders: persons who have an interest (or "stake") in resolving hate speech in any context; this may include community members, survivors / victims, practitioners, NGOs, scholars, elected officials, law enforcement, or teachers, among others.
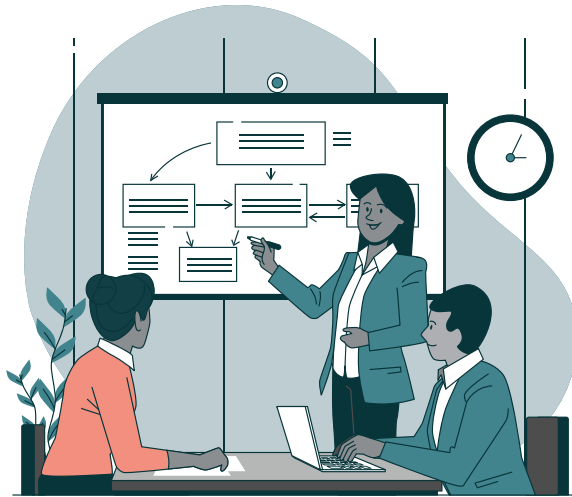
Module 1
Frameworks for Defining Hate Speech

You have all worked in the space of peacebuilding and countering hate speech and are familiar with the main concepts of hate speech. However, let's take a moment to review some of the key concepts and frameworks relevant to this work.

What are the differences between hate speech, inflammatory speech, and dangerous speech? How are they each defined? And, how do different actors approach these concepts in the context of their field? Along with traditional understandings of hate, inflammatory, and dangerous speech, what roles can disinformation and misinformation play?

You will find resources to deepen your study, should you wish to do so, at the end of each submodule. Finally, before you dive into the content, please take a moment to respond to the discussion questions!

Learning Objectives

By the end of this module, you will be able to...

- Understand the landscape of hate speech definitions and frameworks and know how to situate your work in it;
- Define and discuss the dangerous speech framework and contrast it with the hate speech definitions and frameworks;
- Discuss issues that come up around hate speech laws, social media company's policy, and free speech;
- Understand what mis- and disinformation are, and how they relate to hate speech.

Featured Tech Tools

[Jamboard](#) is a whiteboard integrated into Google suite, accessible during google meet sessions. Attendants can collaborate with writing and stickies. Great for group brainstorming and project planning. We will be using this tool throughout the module!

[Here](#) is the Jamboard trailer video

[Here](#) is the Jamboard training video

Discussion Questions

- Describe hate speech in three words!
- Do you and your organization have a clear working definition of hate speech?
- How do we best communicate about hate speech to promote a common understanding of related issues and concerns, and minimize confusion and/or backlash/resistance?
- How do we educate ourselves and others about hate speech?

Activity

- Discussion Groups

  Exercise Instructions: Participate in the interactive discussion using Jamboard

  – Discuss: Does your country have a hate speech law? Discuss similarities and differences with other national contexts.

- Research and Report

  – Exercise Instructions: Find an example of a post on social media (i.e. a tweet or Facebook post) that includes both disinformation and hate speech. Assess: How is this disinformation? What narratives about the conflict or division does it allude to? How is hate speech used in this context?

Submodule 1: Defining Hate Speech

Hate speech is a widely recognized issue and phenomenon, but it can have varying definitions depending on the field of study and practice you approach it from. For example, legal or regulatory approaches seek to define it in order to sanction it. While the field of conflict prevention and resolution looks at hate speech through the lens of the harm and the impact it has on communities or countries in conflict, still other approaches seek to map and monitor hate speech in order to understand its impact on society and conflict. Thus, the hate speech landscape includes a wide array of definitions determined by national laws and regulations that vary from country to country, definitions used by the 'community standards' enforced by social media companies on their platforms, as well as working definitions proposed by international organizations and instruments.

Conceptually, most agree that at its core, hate speech is *any form of expression that seeks to promote hate and targets individuals or groups based on shared and innate characteristics.*

Hate speech is also widely recognized as a stepping-stone and precursor to the perpetration of mass atrocity crimes, such as genocide: "*Genocide is a process, not an event. It did not start with the gas chambers, it started with hate speech.*" (Sheri P. Rosenberg).

Because of its connection with violence and crime, legality is another way of categorizing hate speech. There is hate speech that is illegal (at the international or national level) and there is hate speech that is not illegal, but still recognized as harmful – especially in fragile and volatile social and political environments.

PeaceTech Lab does not adopt a strict working definition of hate speech and instead expands its lens to incorporate "insulting and inflammatory terms and phrases." This framework straddles the ideas of hate speech and dangerous speech (see Submodule 2), but offers a little more flexibility. Focusing on impact, rather than a strict definition or the intent of the speech, allows a nuanced approach to mapping and monitoring the phenomenon in its specific contexts. This is an important asset when analyzing the impact of hateful speech in regards to conflict dynamics.

The wide array of definitions, field-specific missions, regulatory drivers, and legal variations has led to some conflation of concepts and definitions – as different actors use these different reference points for their work – which in turn leads to a dilution of the concepts and makes implementation difficult.

It is therefore important that your organization positions itself and its work in this field and adopts a clear definition and frame of reference for hate speech. As noted above, there are multiple approaches you can draw from. Beyond the definitions of hate speech offered by major international CSOs or institutions (see Featured Resources), the Dangerous Speech Project proposes a concept and detailed framework, and the UN recently clarified their working definition of hate speech along with its practical implications for their programs and member states' obligations.

Featured Resources

- Council of Europe Definition of Hate Speech
- Rights for Peace: What is Hate Speech?
- The Ten Stages of Genocide
- The Pyramid of Hate
- Listen: Free Speech vs. Hate Speech
- Hatebase

Discussion Questions

- Which working definition(s) of hate speech best applies in your context for your work?
- Would merging two or more definitions create a better fit in your context for your work?
- What challenges do you face with defining hate speech in your country?

Submodule 2: Dangerous Speech

To provide some clarity on the concept of hate speech, scholar Susan Benesch proposes an alternative framework of "dangerous speech". In this framework, the focus on "dangerous" expressions highlights the effects and impacts of inflammatory as well as hateful communication, specifically how such communication promotes harm: *"Dangerous Speech is any form of expression (speech, text, or images) that can increase the risk that its audience will condone or participate in violence against members of another group."*

Based on this definition, the Dangerous Speech Project offers a detailed framework to identify and understand the impact of dangerous speech. Explore further here.

Case Studies

President Rodrigo Duterte's Dangerous Speech in the Philippines

As President of the Philippines, Rodrigo Duterte has used dangerous speech, inciting violence against drug users as a part of his nationwide campaign against drugs. When running for president in 2016, he explicitly encouraged the murder of drug users and traffickers. He dehumanized these individuals, claiming their "brains do not function," "they are beyond redemption," and they are merely violent murders. His speeches have included explicit orders to kill, provided impunity to law enforcement officers responsible for killings, and were largely delivered before law enforcement officers who were responsible for enacting his war on "drugs" (i.e., drug users). His speeches have had disastrous consequences. From July 1, 2016-October 25, 2017, Duterte's war on drugs has left 3,967 "drug personalities" dead. These deaths alarmed the International Criminal Court which has promised to conduct an examination into whether or not Duterte committed a crime against humanity. The real life, violent implications of Duterte's speech provide an interesting example of dangerous speech.

Featured Resources

- Video: What is Dangerous Speech?
- What is Dangerous Speech?
- Video: The origins of Mass Violence and Genocide in 100 seconds
- Guide: The Dangerous Speech Framework

Discussion Questions

- Have you used the dangerous speech framework in your work? How so?
- How does dangerous speech differ from the 'classic' understanding of hate speech?
- How do you differentiate between hate speech and dangerous speech in your context? Why is it important to engage with these differences?

Submodule 3: Laws and Regulations

National hate speech laws are often based in principles of anti-discrimination or protection of human dignity. The content and application of these laws and regulations vary from country to country, especially when it comes to including protection for religious belief systems or political entities.

Social media companies – as private, for-profit ventures – set their own conditions and standards of use for their platforms. Facebook, Twitter, and Youtube have community standards and codes of conduct that include their own definitions of what constitutes unacceptable content and conduct within their ecosystems. Based on these definitions, social media companies have developed policies and strategies about how to handle violations of their rules. But, even though these rules are regularly updated, social media companies still struggle to effectively mitigate the issue of inflammatory content. In large part, this is because highly-charged and emotion-driven content is a primary driver of interaction on social platforms and is therefore deeply intertwined with their business model and foundational algorithms.

There is increasing awareness among international actors and states of the grave threat posed by social media's supercharging effects in amplifying and spreading hate speech, dangerous speech, and mis/disinformation. As a result, many states are adopting laws and regulations designed specifically to target online content or to force social media companies to take more aggressive measures to contain hateful and toxic content.

However, combating hate speech brings up important issues around freedom of speech. In both democratic and non-democratic contexts – to varying degrees – laws and regulations on hate speech can be misused to silence dissent, target political opponents, and crack down on minorities. It is a complex and sensitive issue,

but one way to navigate that tension can be to contrast national laws and their application with the dangerous speech framework (see Submodule 2) which can be used as both reference and standard.

Case Studies

Responding to 'Hate Speech': Comparative Overview of Six EU countries

This report provides a comparative overview of six EU countries' laws regarding hate speech. Such laws are often framed through the legal issue of defamation which is a criminal offense in Austria, Germany, Italy, and Poland. Germany specifically allows for the prosecution of "group" defamation, the concept most clearly applicable to 'hate speech' cases, and strictly outlaws Holocaust denial. The UK, Germany, Italy, and Poland outlaw the improper use of electronic communication to send such messages (inter alia, "offensive" or "menacing" messages); and "threats" with bias motivation. These laws aim to decrease hate speech and dangerous speech, both online and in person.

Turkish Government Wants Silicon Valley to Do Its Dirty Work

Turkey has a history of restricting online speech and has extensive systems in place to criminalize and remove such speech. Social media platforms have historically provided an outlet to uplift marginalized voices and social activism in Turkey. In October of 2020, the Turkish Government issued new internet regulations that heavily censor free speech. The amendment requires major social

media companies to appoint Turkish legal representatives to address the government's demand for censorship. The representatives are expected to remove content at the government's request within just 48 hours or else face stiff fines and prosecution. Should companies refuse to appoint a representative, Turkey will essentially block access to the site by slowing down the internet bandwidth. Thus, Turkey's increased online censorship will deprive Turkish people of their only remaining source of reliable nonpartisan news.

Report of the Special Rapporteur on the Situation of Human Rights in Myanmar (09/2020)

In September 2020, the United Nations Special Rapporteur on the situation of human rights in Myanmar delivered his first report to the UNGA since the beginning of his mandate in May 2020. Under current legislation in Myanmar, prohibited speech includes language that "can harm dignity and morality", questions or "disrespects" existing legislation, criticizes the Tatmadaw, "defames" the nation or "tarnishes" the image of the country, tarnishes the Union's sovereignty and territorial integrity, or incites civil servants to oppose the Government. Meanwhile, nationalist groups thrive on Facebook, posting hate speech that targets State Counsellor Aung San Suu Kyi and members of the Government, as well as Muslims, Rohingya, and political parties deemed supportive of freedom of religion. These posts contain dangerous speech, hate speech, and disinformation, posing significant challenges for upcoming elections and religious tolerance.

Featured Resources

- Facebook Accused of Violating its Hate Speech Policy in India
- South Africa: Prevention and Combating of Hate Speech Bill

- International Comparison of Approaches to Hate Speech (Germany, Israel, EU and USA)
- German Hate Speech Laws Also Cover Misogynist Abuse, Court Rules
- European Commission Code of Conduct: Illegal Online Hate Speech Q&A
- But Facebook is Not a Country: How to Interpret Human Rights Law for Social Media Companies
- Facebook Discussion: Who Should Decide What is Hate Speech in an Online Global Community?
- Report. Hogan Lovell: The Global Regulation of Hate Speech (2020)
- Wiki: Hate Speech Laws by Country
- Video: Facebook's Role in Brexit
- Freedom of speech vs. hate – Skokie: The Legacy of the Would-be Nazi March in a Town of Holocaust survivors
- European Commission Code of Conduct

Discussion Questions

- How do hate speech and freedom of speech relate to each other?
- Describe a time when there was a tension between hate speech and freedom of speech in your context? If so, how did it manifest? How was it addressed?

Submodule 4: Misinformation and Disinformation

What is the difference between mis and disinformation? How are they related to hate speech? Much of the debate around "fake news" uses the terms "misinformation" and "disinformation interchangeably. However these terms are not synonymous. The difference lies in the intentions of the person circulating the information. In the case of misinformation, the person circulating the content believes that it is true. For instance, many people who believe that climate change is a myth may spread false information including conspiracy theories, bad data, or other incorrect or misleading communication, because they think it is the truth. Disinformation, on the other hand, is spread when the person sharing false information knows that what they are sharing is **not** true. They spread false information intentionally, for the purpose of manipulating their audience for self-enrichment (e.g., gaining political advantage). Often, actors spreading disinformation use false 'news' to instigate their followers against a particular community.

Even though the intention behind the spread of mis-/disinformation may be different, both types of false information can do a lot of harm. Just like hate speech, mis-/disinformation can be, and is increasingly, inflammatory. In these cases, similar elements as those used in analyzing hate speech can be used to evaluate their potential scope and impact. Primary evaluative components include the context, who is speaking, and the speaker's reach (potential audience size). Hate speech and mis-/disinformation can be closely intertwined, using both direct and indirect messages of hate, promoting narratives that stoke existential fear, intensifying the sense of division between communities, and ultimately enabling calls for mass violence.

hashtag #CoronaJihad prevented accurate information about the coronavirus from reaching the public, which in turn exacerbated the spread of the virus. This report examines how disinformation is used to spread hate against a community.

Featured Resources

- What are mis- and disinformation? Read up on them here
- A Finder's Guide to Facts
- Video: The Role of Misinformation, Disinformation, and Hate Speech in Conflict
- Video: Why We Are Susceptible to Misinformation

Case Studies

Coronajihad: Islamophobic Hate Speech and Disinformation During the First Wave of the COVID-19 Outbreak in India

During the initial phase of COVID-19 spread in India, many political actors used disinformation to instigate Islamophobia. Muslims were falsely accused of spreading COVID-19 among local communities as a way of performing Jihad. The

Discussion Questions

- How do you think mis-/disinformation interacts with hate speech in your context? What impact do they have?
- Can you think of recent examples of how mis-/disinformation and dangerous speech were used together to increase hate and conflict in your context?

Technology Tools Corner

Jamboard

Introduction

Jamboard is a whiteboard integrated into Google Suite, accessible during Google Meet sessions. Participants can collaborate with writing and stickies. The tool is great for group brainstorming and project planning. We will be using Jamboard throughout the module!

We chose Jamboard because it is easy to use and widely available for anyone with a Google account. Specifically in this module, you will be using Jamboard to help you visualize all you've learned about the definitions of hate speech and dangerous speech, to interact with other participants, and to share any other ideas you may have.

Here are the major things you can do with Jamboard:

- Apply a variety of board backgrounds
- Pens for drawing
- Eraser for erasing
- Selector for selecting board elements
- Image uploading
- Circle tool for adding circles
- Text boxes
- Sticky notes
- Laser for pointing to elements on your board

Getting Started

To get started with Jamboard, ensure you have logged onto your browser with your Google account (Gmail). Any browser should work fine. Once logged in, navigate to https:// jamboard.google.com, this should open the Jamboard homepage where you can see any recent boards you may have created or have been shared with you. Click the plus (+) icon on the bottom right of this page to create a new board.
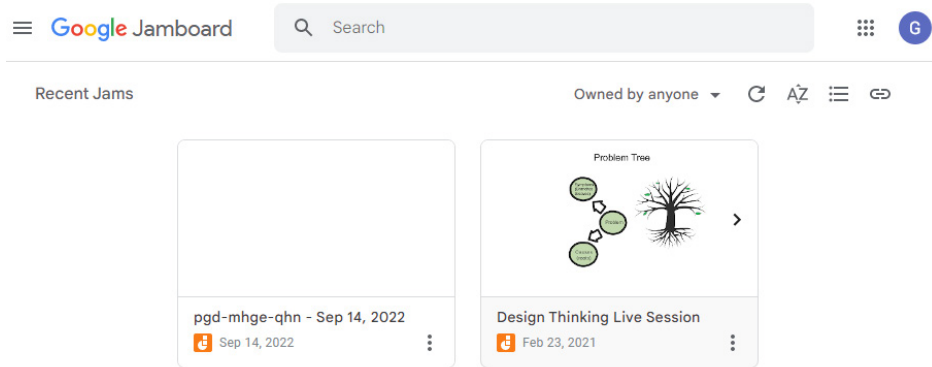
Similarly, if you are in a Google Meets call, you can access Jamboard from the Activities button on the bottom right of your screen. Once this is open, select Whiteboarding as the activity of choice.
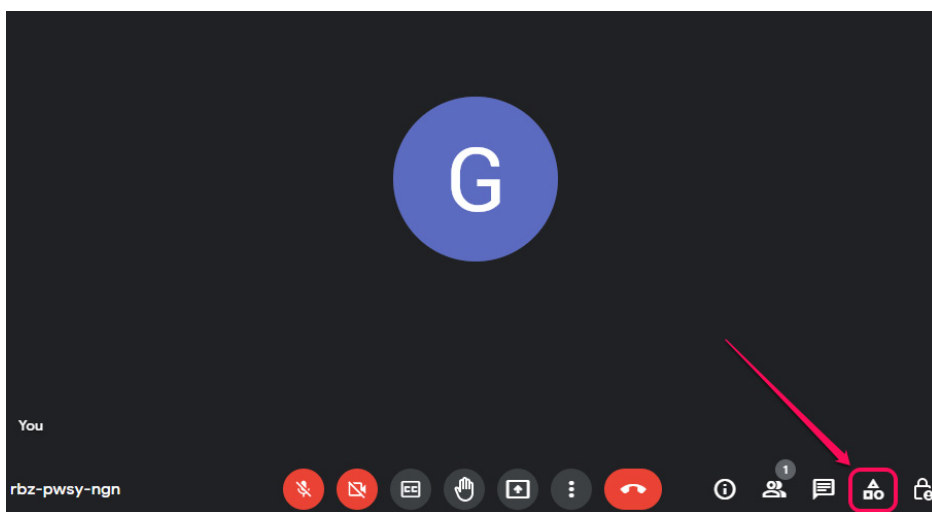
Once this is open, select Whiteboarding as the activity of choice.
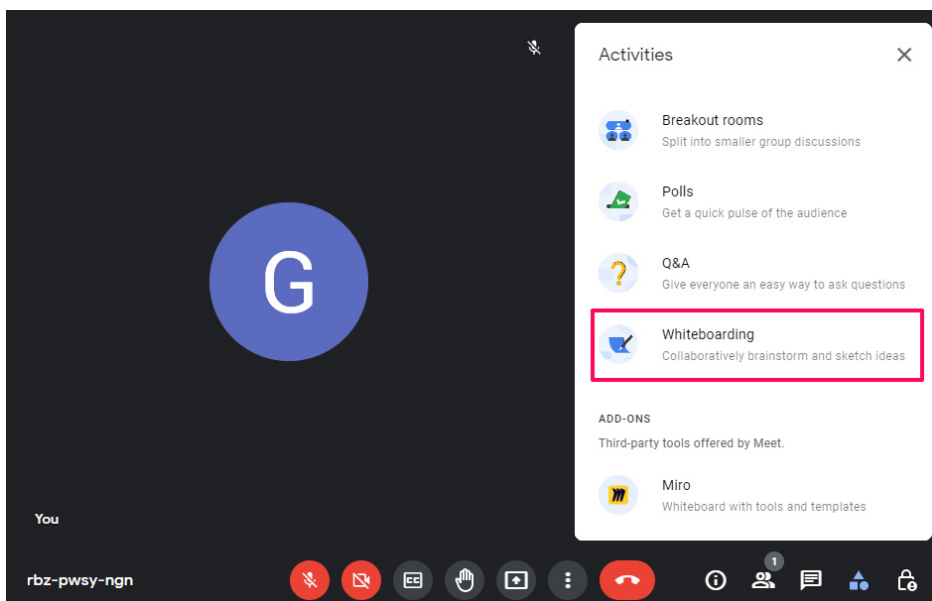
You are now ready to start sharing your ideas. Feel free to experiment with the different editing tools available until you can create a board that is able to communicate your ideas and creativity as you imagined it. Here is a sample board created for Module 1.

Tech Tool Practice

Use Jamboard to explain/collaborate with a team on how hate speech has been affecting the social lives and wellbeing of people in your area.

1. Open a new Jamboard. You can share it with your team members for collaboration.

2. Set a nice background for your Jamboard by choosing one from the 'Set background' option.
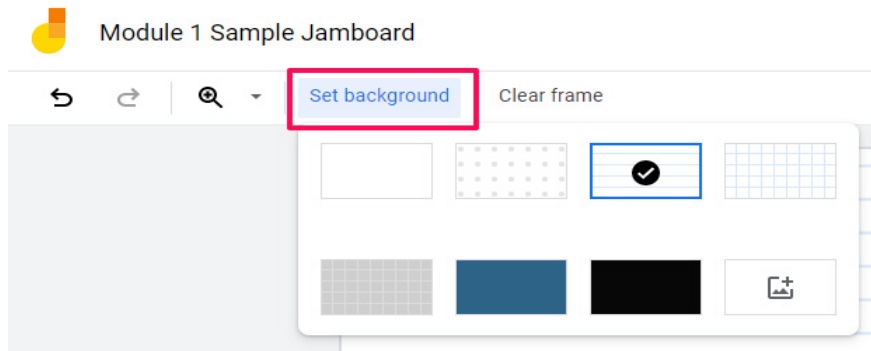
3. Create a title for your board with the text box option. You can customize it as you'd like.
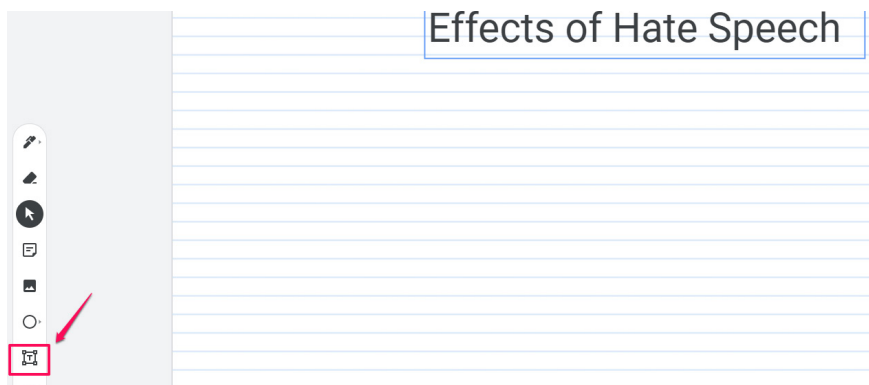
4. Use sticky notes on one side of your board to jot down any ideas that you think of, or any contributions made by your team members.

5. Use the 'Add Image' option to add some images that will drive your point home. You can also use shapes and drawings.
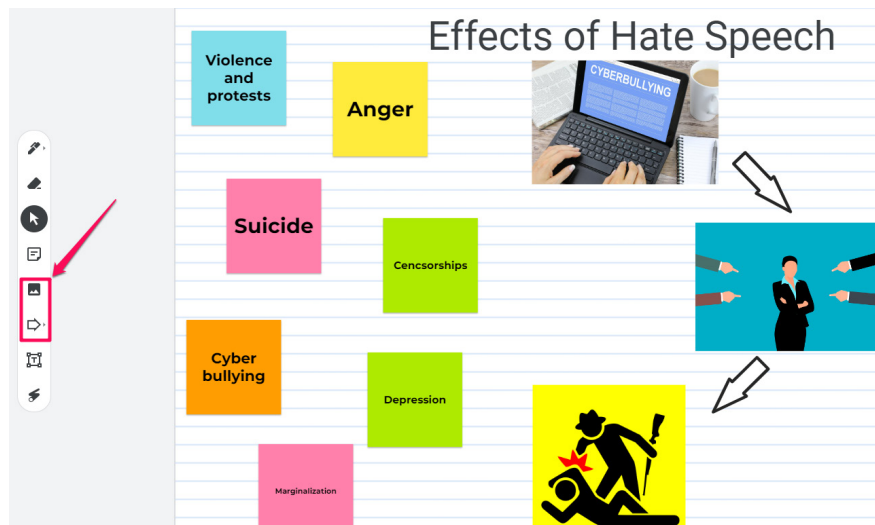
6. Lastly, you can add links to tools or articles you think can be helpful to your topic using the text box.
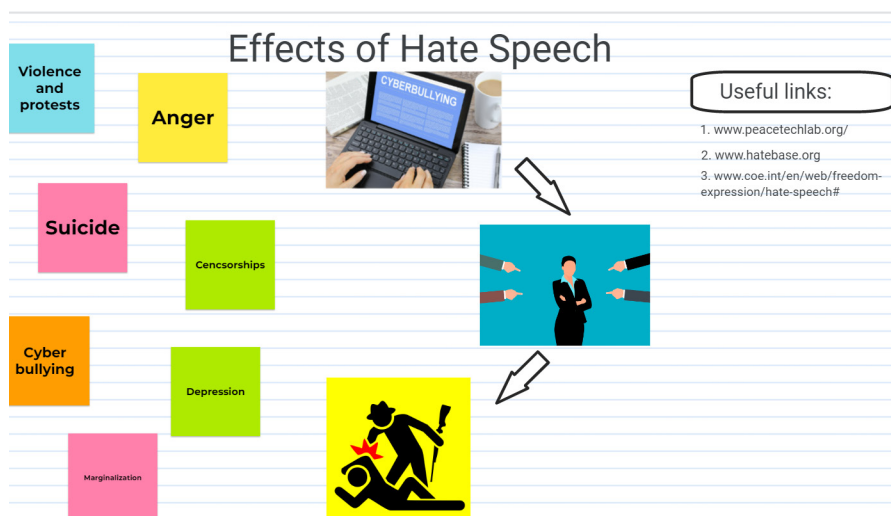
7. Add as many frames (pages) as you'd like using the Frame bar at the top of the page.

8. Lastly, be sure to share it with your team members or use it within a presentation. It works similar to Powerpoint.

Module 2
Hate Speech Monitoring: Social Media Monitoring, Data, Analysis, and Narrative Analysis

Before elaborating a strategy and deciding on what actions to take against hate speech in your context, it is important to understand two essential aspects of taking action: online security and the nature of hate speech. Taking action includes gathering data, monitoring media, analyzing the data, and formulating a response. Much of this work today incorporates online activity, but even when the activity is offline, bad actors may seek to find online data and information about team members from the internet and social media. So, it is essential that you and your organization take the proper steps to ensure the safety, particularly online safety and security, of team members working to take action. Second, you must understand the general makeup of the problem in your context: where does it occur, who are the main players, and what are the main narratives that serve to proliferate hate, dis- and misinformation that feed into conflict, undermine social cohesion and stoke fear? The insights and patterns that emerge can inform decisions as to what tools and types of interventions would be most effective in combating hate speech in your context.

Mapping and monitoring hate speech can occur at the start of an initiative and on an ongoing basis to provide updated information on what can be a very dynamic situation. Setting up a monitoring and analysis process can also provide a foundation for an early warning system.

We propose a four-step process to get started:

1. Preparing an online safety and security plan;
2. Gathering information and setting up a monitoring system;
3. Analyzing the data that is gathered; and then
4. Diving deeper into understanding the narratives that underpin the hate speech you seek to address.

Featured Tech Tools

KoBo Toolbox collects data using powerful form creation engines that cater for internet unavailability and instability. Get your data all in one place and as clean as possible.

Here is the KoBo Toolbox trailer video

Here is the KoBo Toolbox 30-minute training video

Discussion Questions

- How familiar are you/your organization with online safety and security?
- How familiar are you/your organization with data collection?
- How familiar are you/your organization with media monitoring?
- How familiar are you/your organization with data analysis?
- How familiar are you/your organization with narrative analysis?

Activity

For this module's activity, we will focus on the aspects of creating an action plan using the four-step approach outlined above. We will use Jamboard again to share our thoughts and record our ideas.

- **Step 1**: First quickly identify a hate-based problem in your context that you might be interested in working against.

- **Step 2**: List its features (Does it involve hate speech or dangerous speech? Is it primarily online or offline or both? Who are the main players you can think of? etc.)
- **Step 3**: Now focus on your assigned step from the four-step process (Online Safety, Data Collection, Media Monitoring, or Data Analysis). How would you plan to implement that step of the process for your identified problem and context?
- **Step 4**: Post the outline of your plan to a Jamboard page.
- **Step 5**: Discuss!

It is okay if your outline has questions or gaps that need to be filled, as you go through the submodules below, you will get more information and directed questions to assist you with a more complete development of your action planning.

Submodule 1: Online Security and Safety

When responding to hate speech in online contexts, counter speakers may experience mental and emotional responses to hateful content, and they may become targets of hate proponents they are interacting with or who see the interactions and utilize the dialogue to interject additional hate speech. It is important to plan effectively prior to working with difficult content (online hate and hate speech) as well as for protecting online identity prior to participating in counter speech dialogues or campaigns.

Protecting your identity online

Before participating in counter speech dialogues or campaigns online it is important to assess your online "footprint" – all the sites where you maintain a presence online, have posted online materials, and posted material about you – particularly identifying information that can be utilized to harass or bully you. This will help reduce the capacity of bad actors and hate proponents to engage in harassment campaigns known as "doxxing" (doxx is short for documents) or publishing your personal information in public or private online forums with the intent of having others use the information. In particular, email addresses, phone numbers, home addresses, and other personal information should be removed from online contexts wherever possible. Additionally, former posts and shares of information should be assessed for their potential use in such campaigns. While it is impossible to prevent all potential harm, the likelihood and materials available can be substantially reduced.

Being prepared for online harassment once you have taken steps to protect your identity online includes having a plan for what steps you should take if harassment occurs. This includes documenting the harassment, reporting the harassment, as well as approaches to protecting your mental and emotional well-being during and after harassment. Options for documenting harassment will vary according to the platform or technology features available where the harassment is occuring (and in some cases on multiple platforms/

technologies such as social media and email). It is possible to prevent messaging by **making your accounts private** which means you will not experience as many messages, but does not prevent bad actors from doxxing your information or harassing through other forums. You can also **mute responses** on some social media platforms so that they are still occurring but you do not have to experience them in real time. Muting allows you to retain the messages for documentation purposes without having to be bombarded during a harassment campaign.

Dealing with Hateful Content – Working With Traumatic Imagery

Along with the potential for experiencing online harassment and the need to protect your identity, many people addressing hate speech online may experience mental and emotional responses when working with, reading, and responding to hateful content over time. It is important to find ways to decompress from online engagement in order to sustain your mental and emotional well-being to ensure your capacity to continue working against hate.

Case Studies

Doxxing in Tunisia

According to a report published by SMEX, a Lebanese digital rights NGO serving the MENA region, cases of doxxing in Tunisia last year impacted "...human rights defenders, women, and LGBTQI+ activists who took part in anti-government protests were subjected to a

growing number of social media campaigns of bullying, forced outings and doxxing of their personal information." The report references case studies including the case of one woman whose personal information was doxxed by pages affiliated with police unions and another victim who was targeted on social media following her participation in protests. The report's findings indicate that "pro-police unions pages and accounts on Facebook shared her home address and telephone number along with her personal photos and comments threatening her. She also received hundreds of messages on Facebook including bullying comments and death and rape threats forcing her to deactivate her social media accounts."

'It Gets to You.': Extremism Researchers Confront the Unseen Toll of Their Work

Recently scholars and other researchers of extremism, hate, and political violence have been speaking out about their struggles with "difficult data," particularly analyzing violent images and hateful content. Many have begun speaking out about the need to have outlets both formal and informal to deal with the mental and emotional toll of doing their work. These researchers have come up with means of dealing with their "vicarious trauma" while continuing to conduct research, but are calling for better information, training, pre-planning, and acknowledgement of these potential harms specifically training new analysts before they begin working with these difficult subjects.

Numerous resources exist to introduce activists to the basics of digital security, such as how to use VPNs, end-to-end encryption, and air-gapped computers. In addition, multiple organizations have arisen to help assist activists and researchers with protecting their safety and security as well as helping to manage mental and emotional well-being.

Featured Resources

- Surveillance Self-Defense has software recommendations such as how to use Tor browser and how to enable two factor authorization and how to train others in online security with lesson plans and handouts in English.
- Security-in-a-box provides simple step by step instructions on topics such as phone settings with helpful screen shots.
- SMEX Digital Safety Help Desk provides users in Lebanon and other Arabic-speaking countries with digital safety tips, rapid response to digital attacks, and threat mitigation.
- Speak Up & Stay Safe(r): A Guide to Protecting Yourself From Online Harassment Details current best practices for security and safety for most online environments.
- PEN America Online Harassment Field Manual covers preparation, response, self-care, legal issues, support, and online abuse.
- Frontline Defenders Workbook on Security helps users to assess their security and produce a plan.
- DART Center for Journalism and Trauma has created a practical guide for dealing with online trauma.

Discussion Questions

Online Safety and Security

Online safety and security are always important, but become especially so when you are taking action against hateful people and organizations. In our digitally mediated world, silencing and dominance practices foreground abuse, and data-driven tactics. Moreover, regular exposure to hateful content including speech, imagery, videos, and music can take

a toll on those studying and responding to hate. These questions are intended to facilitate thinking about how online safety and security can become a regular aspect of your work.

- Do you know what your online "footprint" is?

- Are there affordable services in your area to help you manage your data and online presence?

- What are some ways you can disconnect and decompress from exposure to hateful messaging and media?

Submodule 2: Gathering Information and Monitoring

Technology and data tools can substantially enhance your information gathering and monitoring capability, especially when you are looking to monitor online hate speech. However, these tools also have technical limitations, and hate speech does not occur in isolated environments. Social media exponentially enhances this problem. It can therefore be helpful to take a broader view and seek to understand the 'landscape' of hate speech before seeking to address the phenomenon itself. This can include identifying and monitoring which online and offline communication channels and media are being used, how hateful messages and mis/dis information travel before they reach social media, and who the main actors are that drive the spread of hateful content.

A robust information gathering and monitoring system should be tailored to your context. It can associate low tech and high tech, as well as offline and online methods, especially when it comes to gathering information. All these different components can be combined to adapt your approach to the challenges and capabilities of your given environment. In addition to your context, another aspect to consider when choosing your method for gathering data are the audience you are gathering the information from, the platforms, tools, and methods where they get and share information, the information you are looking for, and the tools to be used to collect this information.

Humans gathering information online and offline

Frameworks for gathering information this way include online and in-person surveys, text messaging apps, and interviews to name a few. Each of these methods presents advantages and limitations. It is also important to be strategic by having a clear idea <u>what</u> information needs to be gathered.

Here are two example questionnaires used to gather information about hate speech terms and phrases in <u>Ethiopia</u> and <u>Sudan</u>.

Human monitoring and "human-led automatic monitoring"

A number of tech tools can support the speed and scale at which you can investigate hate speech activity online, especially on social media. For example, some tools can help you go through large amounts of data gathered from several social media platforms simultaneously, others can help you organize or even visualize the information you find for easier analysis. Each tool has different levels of effort involved and range from simple to very complex to use. Browse the Technology Tools Corner in each module and look into what tools are best adapted to your and your teams' capabilities and needs!

🎯 Case Studies

Umati: Monitoring Online Dangerous Speech During the 2015 Elections

Nigeria has a history of election violence. The Centre for Information Technology and Development (CITAD) and iHub Research worked together to launch a hate speech monitoring system which was active during the entire Nigerian 2015 election cycle. For this project, they contextualized dangerous speech in Nigeria as "any speech act that denigrates people on the basis of their membership in a group, such as an ethnic or religious group,

and that has a reasonable chance of catalyzing or amplifying violence by one group against another, given the circumstances in which it is made or disseminated". They developed a structured online form to collect hate speech content about the perpetrators, presentation, and content of the hate speech. The project adopted the Umati project's categorization spectrum, which ranges from offensive speech, to moderately dangerous, and extremely dangerous speech based on the speaker's influence and the content's call to violence. The project aimed to generate analyses of dangerous speech that could be used for early warning systems and to respond to the likelihood of conflict. This project is a great example of how to effectively conduct data collection and analysis about hate speech.

PeaceInsight: Monitoring Online Hate Speech to Inform Programming

In Kyrgyzstan, there is a growing presence of hate speech in the media, in public discourse, and on the internet targeting religious minorities and the LGBTQ+ community. To combat the spread of hate speech, the School of Peacemaking and Media Technology has developed programs to reduce the risk of ethic conflict in Kyrgyzstan. The organization regularly monitors the internet, studying public opinion and using this data to develop innovative training materials for journalists, editors, and activists to train them on how to overcome hate speech. They monitored the internet by analyzing the content of 58 media outlets issued and published in Kyrgyzstan including Facebook groups, online media outlets, newspapers and TV channels. These media outlets were selected on the basis of their popularity among the audience, availability and distribution in all regions of Kyrgyzstan. The overall study tracked the languages used, the phobias mentioned, the minority groups referenced, the leading perpetrators, and the tone of the content. This data was ultimately used to prepare workshops with partner NGOs about decreasing hate speech.

Mapping and Monitoring Hate Speech Directed at Jewish Lawmakers in The US

The Anti Defamation League (ADL), conducted a study analyzing antisemitism on Twitter directed towards Jewish US Congress members seeking reelection in 2020.The goals of the study was to make policy recommendations to technology companies and explore how social media is used to exploit stereotypes about particular identity groups based on religion, race or other characteristics. The study was conducted by analyzing 5,954 tweets out of a total dataset of 337,689 tweets (the total number of tweets directed at Jewish incumbents over the one-month period.) An Online Hate Index (OHI) classifier was applied on all tweets which output a score (ranging from 0% to 99%) of the OHI's confidence that a given tweet is antisemitic. The top 200 scored tweets for each of the 30 elected office holders were manually reviewed and analyzed. Manually reviewing tweets enables the ADL to better understand the contemporary manifestations of antisemitism.

Featured Resources

- Count me in! Collecting human rights-based data
- Basic Tools for Data Collection
- Mapping and analyzing networks: how to use Gephi
- Umati: Monitoring online dangerous speech during the 2015 Elections
- PeaceInsight: Monitoring online hate speech to inform programming
- Mapping and monitoring hate speech directed at Jewish lawmakers in the US

Discussion Questions

Understanding the Information Ecosystem

Understanding the information ecosystem is an important part of understanding what information you want to gather or monitor, and who or where to gather that information. These questions are intended to facilitate thinking about the information ecosystem in the context where you are working, the type of information you seek, and where to find it.

- What does the information ecosystem look like in your context? Where are people getting their trusted information? How does information travel?

- What considerations guide your selection of which information to gather, who to target, and what method and/or tool to use to gather that information?

- Are there other approaches to gathering information that you use regularly, or think would work well in your context?

- Are there other media platforms such as radio, television, print media, or other platforms that should be considered? What are potential challenges of using them (e.g. radio, taking notes, etc...)?

Submodule 3: Analyzing the Data

Data gathered through monitoring activities can provide a number of analytical insights. One way of looking at the data is **quantitatively**: For example, which terms are most prevalent online, on what platform, and used by what actors? Another way to analyze the data is **qualitatively**: How are hateful terms being used, what other terms are they associated with, does their meaning evolve?

Here are a few examples of the types of questions that can guide your analysis:

Quantitative content analysis:

- Which terms are most prevalent online?

  Among the terms and phrases you are monitoring, which ones are the most commonly used? Maybe some are used only rarely, while others are much more common? Sometimes those ratios can also change over time: a rare term can become widely used, and vice versa. These patterns in turn provide insights such as: how conflict dynamics may be evolving, whether tensions are rising or easing, etc.

- Are there (offline) events that drive spikes in hate speech volume, or the use of specific terms?
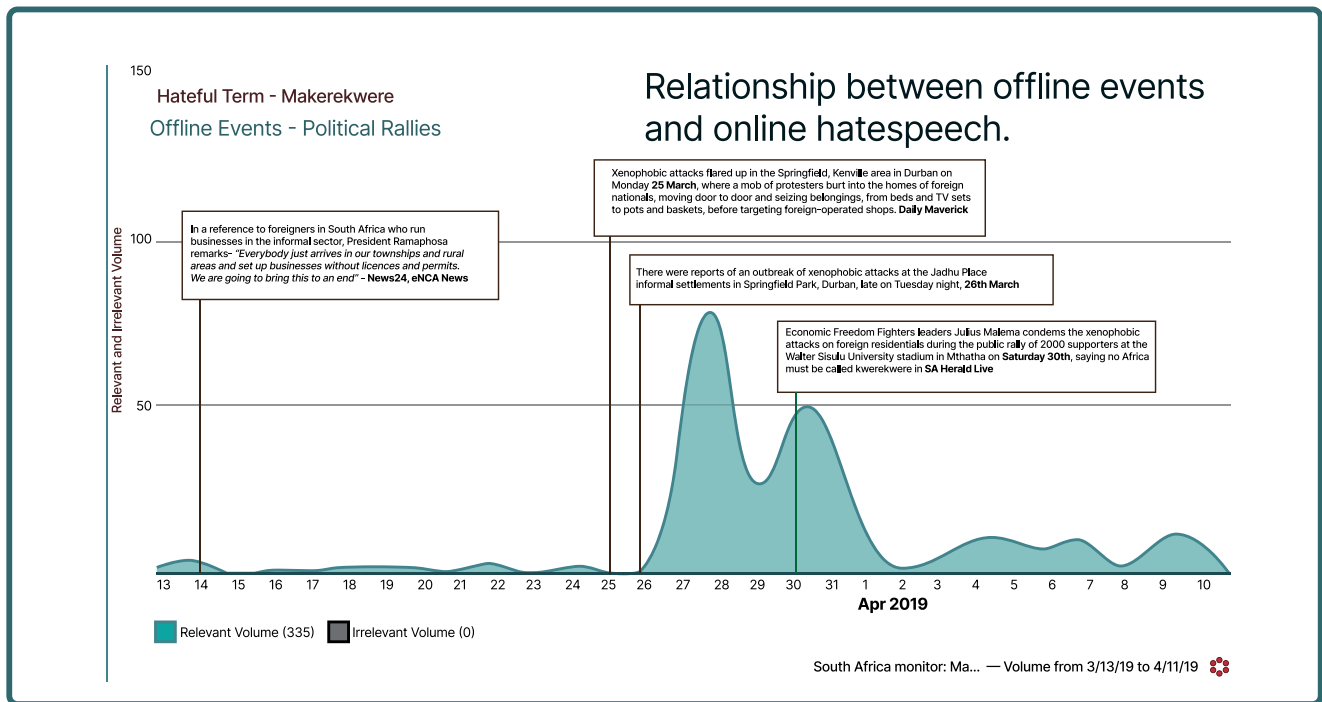
  "Critical episodes" or "catalyst events" are events such as elections, violent incidents, natural disasters, assassinations, and other events that can polarize opinions at the regional or national level. They may also exacerbate tensions and conflict dynamics. Some events are predictable and can be prepared for (elections), others cannot be predicted (for example, the assassination of Ethiopian popular singer Hundessa Hachaluu). All of these events impact online discussion, and online discourses in turn further exacerbate already intense feelings aroused by the 'catalyst event.' This creates a dynamic

that can have repercussions on the existing conflicts/divisions. Examining hate speech data related to these events can provide additional insights into the unfolding climate and conflict dynamics in the context. Some analysts have used this type of analysis to build indicators for early warning and violence prediction efforts.

- Are there instances of online hate speech that cause(d) violence offline?

  Incitement (hate speech that openly calls for violence against a group based on a common characteristic) has a high risk of leading to offline violence. Especially when the message is spread by a person (or group) with authority and influence. But there are also other configurations of hateful speech that can lead to offline violence. Tracking this kind of information can provide insights about potential correlations.

  For example, PeaceTech Lab, in partnership with Media Monitoring Africa, monitored and analyzed language trends throughout South Africa's election season to offer insights on the potential relationship between hateful language on social media and instances of violence seen in municipalities throughout the country. Here is one of the study reports and below is a graph that shows the relationship between online hate speech trends and offline violent events.

Relationship between offline events and online hatespeech.

Hateful Term - Makerekwere
Offline Events - Political Rallies

150
100
50

Relevant and Irrelevant Volume

In a reference to foreigners in South Africa who run businesses in the informal sector, President Ramaphosa remarks- *"Everybody just arrives in our townships and rural areas and set up businesses without licences and permits. We are going to bring this to an end"* – **News24, eNCA News**

Xenophobic attacks flared up in the Springfield, Kenville area in Durban on Monday **25 March**, where a mob of protesters burt into the homes of foreign nationals, moving door to door and seizing belongings, from beds and TV sets to pots and baskets, before targeting foreign-operated shops. **Daily Maverick**

There were reports of an outbreak of xenophobic attacks at the Jadhu Place informal settlements in Springfield Park, Durban, late on Tuesday night, **26th March**

Economic Freedom Fighters leaders Julius Malema condems the xenophobic attacks on foreign residentials during the public rally of 2000 supporters at the Walter Sisulu University stadium in Mthatha on **Saturday 30th**, saying no Africa must be called kwerekwere in **SA Herald Live**

13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 1 2 3 4 5 6 7 8 9 10
Apr 2019

■ Relevant Volume (335) ■ Irrelevant Volume (0)

South Africa monitor: Ma… — Volume from 3/13/19 to 4/11/19

- Who are the actors/accounts that are the most prolific hate speech?

  Identifying the actors and social media accounts that are the most prolific hate speakers and have the widest reach provides important information when considering targeted actions and strategies against hate speech. It will also lay the groundwork for actor mapping activities (see Submodule 4). There are also some tech tools available to assist in conducting quantitative analysis, such as:

  – Voyant (tools for text analysis)

  – Excel

Qualitative content analysis:

Multiple methods can be used to analyze data qualitatively. The goal is to understand the *meaning* of the data in its context. For example, hate terms may often be couched in supposedly humourous content (memes are a well-known online example of this), which can often be difficult to parse in simple quantitative terms. Moreover, meaning of terms changes over time and in different geographic and cultural contexts. Qualitative analysis may involve surveys, focus groups, or narrative analysis of texts and images.

All of these specific methods can provide a variety of insights as to how hate terms are being used, whether conflict and tension are increasing or decreasing, and if violence is being promoted. The analysis is typically not done at as large a scale as quantitative analysis because qualitative analysis takes more time. But, qualitative analyses can provide nuanced and rich findings that are especially useful for creating interventions and developing counter speech.

A few questions to consider:

- How are these terms used and are they associated with other terms?

  Collecting this information and understanding how terms relate to each other will help you divide even deeper and conduct narrative analysis (see Submodule 4).

- Is the meaning of the terms changing, even evolving over time?

  Sometimes, because of events or changes in the conflict dynamics, the meaning of terms can evolve. When monitoring terms online, looking for these shifts can provide valuable clues.

Case Studies

Data Collection, Analysis and Use in Protracted Humanitarian Crises (Bangladesh & Iraq)

This report was produced by Publish What You Fund to provide a view of how data reporting has been conducted in cases of humanitarian crises. The program, launched in 2008, is a global campaign for aid and development transparency. It seeks to ensure that aid and development information is accessible and useful for a variety of purposes including evidence-based advocacy, effective decision-making, accountability and creating sustainable change for better humanitarian outcomes. The report provides information on data gathering and analysis.

Featured Resources

- [6 Methods of Data Collection and Analysis](#)
- [Acquiring and Analyzing Data In Support of Evidence Based Decisions](#)
- [History of Media Monitoring: From Press Clipping to the Internet | What Is Media Monitoring?](#)
- [Introduction to Network Analysis](#)
- [Cartographie de la Haine en Ligne](#)

Discussion Questions

Analyzing the Data

Data Analysis is directly related to the information gathering and monitoring steps from [Submodule 2](#).

- What kinds of data analysis have you done in the past (quantitative? qualitative?) and where would you like to go with your data analysis related to address hate and division?
  - What kinds of insights are you seeking to glean from your analysis?
  - What questions are you seeking to answer through your data analysis?
- How does clarifying the insights and questions you seek to answer also inform the information you seek to collect and the places or people you collect it from?

Submodule 4: Narrative Analysis

Narrative analysis approaches provide a way to understand how stories function in communities and how to intervene when destructive stories circulate. We conduct narrative analysis by examining what is being said in the hateful content to understand the stories and discourses that underpin it. We might ask: What are the major themes and topics that emerge? What are the grievances and fears that are being stoked? What existing or latent conflict dynamics are at play? Resource allocation? Historical injustices? Political marginalization? Or something else?

Why is Narrative Analysis Important?

"Different and competing narratives are an intrinsic part of conflict: People that argue fiercely with each other and groups that fight with each other tend to construct their own narratives in ways that justify their own behavior and throw the blame on the opponent. 'Historical' truth is quickly a serious casualty in heated conflict. Accepting one narrative over others would make the analyst(s) appear clearly biased in the eyes of various actors in the conflict, and undermine trust. Peacebuilders often have to explicitly engage with the contesting narratives which requires that contesting narratives are 'accepted' as part of the reality." (Interpeace, Peacebuilding Guide)

Furthermore Sara Cobb explains: "[...] because history itself is narrative, a story about what happened to whom and why, the narrative/conflict intersection must include not only attention to time, but also the layered way that historical narratives accrue descriptions that become, like sedimentary rock, "truth", over time. And in this process, historical narratives, as well as truth itself, hide from us the messy details that could contradict, or challenge, our understanding. From this perspective, a narrative approach to conflict analysis and resolution calls for the interrogation of history and accumulated facts that all too often simplify, rather than complexify, our understandings of a given conflict."

What Narrative Analysis Can Involve

Sarah Federman proposes the three forms of analyses to provide a variety of starting points for those interested in understanding the role of narrative in conflict. These approaches can be used together or separately at different points within the research process. Read this summary of how you can conduct narrative analysis.

Actor Mapping and Narrative Analysis

Narrative analysis relies on one's understanding of the history of conflict and current conflict dynamics, as well as the actors and parties to the conflict(s). Therefore, it will be important to have already carried out a context/conflict analysis prior to conducting a narrative analysis, and it will be helpful to have that with you and your team as you map out the narratives.

Case Studies

An analytical framework for reconciliation processes – Two case studies in the context of post-war Bosnia and Herzegovina

This paper discusses two analytical frameworks for developing reconciliation processes in the post-war context and in light of the act of genocide. The two cases included are the work of The Karuna Center for Peacebuilding through Project DiaCom, and the Association of Citizens Truth and Reconciliation of Bosnia and Herzegovina. The author provides a discussion and definition of truth and reconciliation, and utilizes qualitative methods to analyze the implementation and impact of the two cases discussed.

💡 Featured Resources

- [Peacebuilding How? Good Practices in Conflict Analysis](#)
- [Preventing and Combating Hate Speech by Understanding and Managing It](#)
- [Engaging Narrative for Peace](#)
- [We CAN Guide: Narratives](#)
- [The Genesis of a Conspiracy Theory](#)
- [Countering Dangerous Narratives in Times of Uncertainty](#)

👥 Discussion Questions

Narrative Analysis

Because narratives and discourses are stories that we tell socially, we can often become so used to them we do not realize that there could be entirely different stories (narratives and discourses). This process whereby we become used to narratives is called naturalization. It is one reason why narrative analysis can be difficult to complete initially. One way to begin to see how narratives are working in our context is to imagine our stories differently. This process can also help with developing counter-narratives.

- What are some of the common hate or conflict related narratives in your context?
- How do those hateful or conflict narratives get adapted to different situations (political, religious, educational, etc.)?
- What are other narratives (remember these are stories) that could be utilized to counter those hateful narratives?

KoBo Toolbox

💡 Introduction

KoBo Toolbox is, at its very core, an open source data collection toolkit with a few features that makes it great for gathering information in areas that do not have great connectivity to the internet. For this reason, KoBo Toolbox is regularly used in humanitarian crisis situations to paint a picture of what is happening on the ground. As a peacekeeper, this tool will come in handy when collecting data in both regular use cases and in challenging conditions and remote areas.

✈ Getting Started

KoBo Toolbox is a free tool and easy to get started with. Go to www.kobotoolbox.org to sign up for a free account. Be sure to use the **Researchers, Aid Workers & Everyone Else** section.

GET STARTED

Get started with KoboToolbox now: it's free, and it takes under a minute. We have two publicly-available instances of KoboToolbox you can choose from:

Unlimited Use for Humanitarian Organizations

Provided by UN OCHA
Unlimited Submissions
Unlimited Data Storage
Unlimited Projects

CREATE AN ACCOUNT

or login

Researchers, Aid Workers & Everyone Else

Provided by Kobo, our own nonprofit
10,000 Submissions Per Month
5GB of Data Storage Per Month
Unlimited Projects

CREATE AN ACCOUNT

or login

For more information on the differences between the two servers and to help determine which one is right for you, please read our Which Server Should I Use? support article.

Advanced users can also install KoboToolbox on their own server (or on a local machine) using Docker. See our kobo-install repository on GitHub for details.

Once logged in, the interface is pretty basic with two tabs for your projects and a library. For each project, you can have it either deployed, as a draft or in the archive.
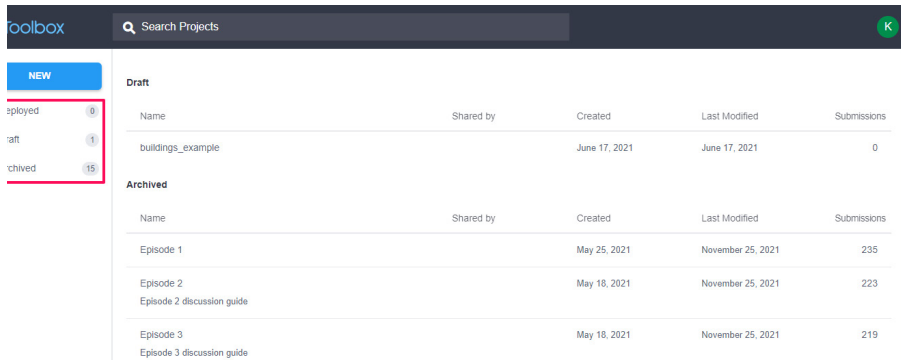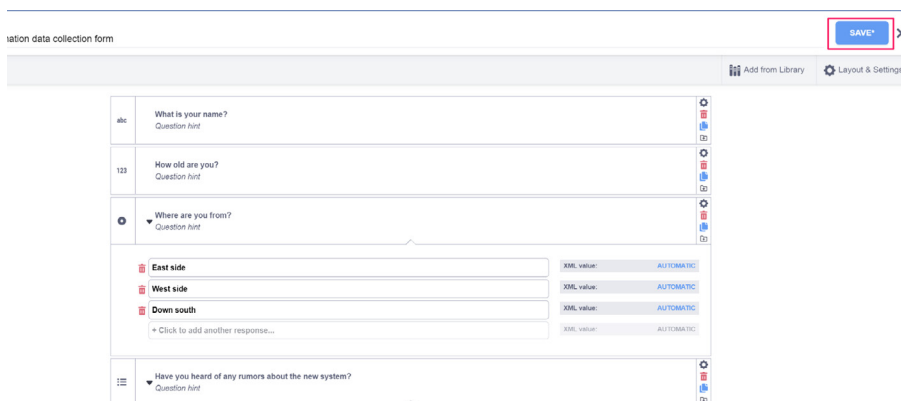

To create a new project, click on the **New** button and select the **Build from scratch** option.
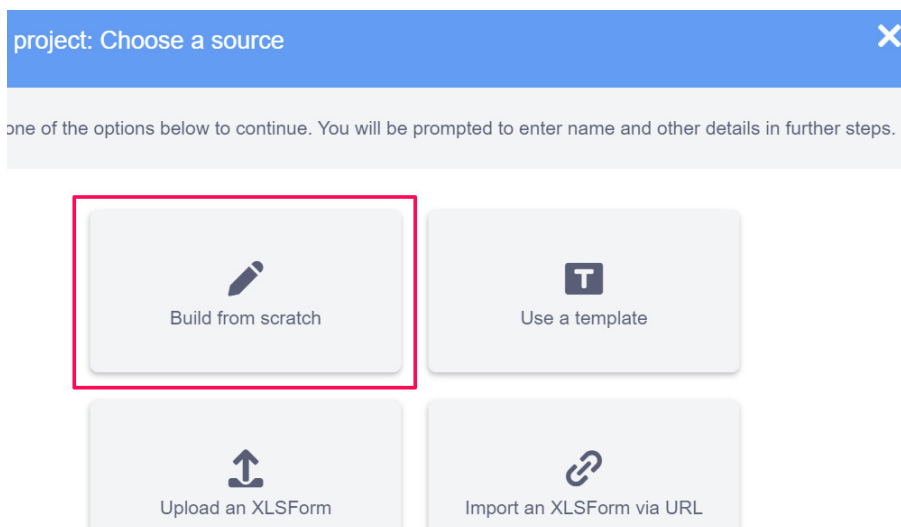

After filling in the form details, you will be directed to a formbuilder page where you can start designing the questionnaire.

Feel free to add as many survey questions as is required and click on the Save button when done.

Be sure to watch the [video tutorial](#) on Kobo Toolbox to learn all about the different ways of creating surveys. Once saved, the form will be stored under the Draft section by default. To make the form usable, it has to be deployed. You can do this by clicking on the deploy button.
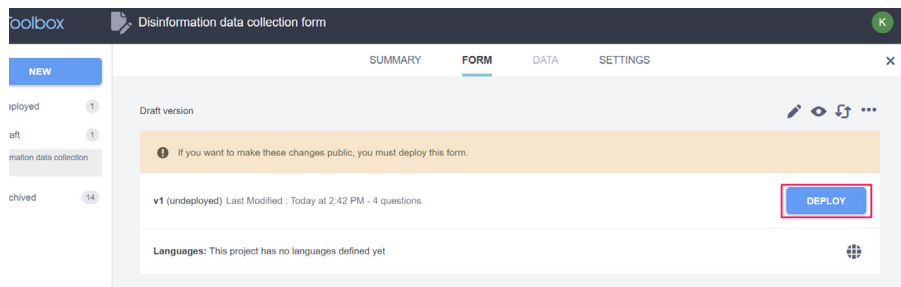
You are now ready to share the link to the form and start collecting data.

 Tech Tool Practice

Use KoboToolbox to create a simple data collection form that collects information on the following:

- Sex and age of respondents
- If they have experienced/spotted misinformation online
- Details on what was spotted
- Actions taken (If any)

1. Start by creating a form on KoboToolbox. We will name it: **Misinformation collection project**. We will begin by adding a note to inform the respondents what the survey is all about.


2. Next, we will add questions regarding the name and age of the respondents.
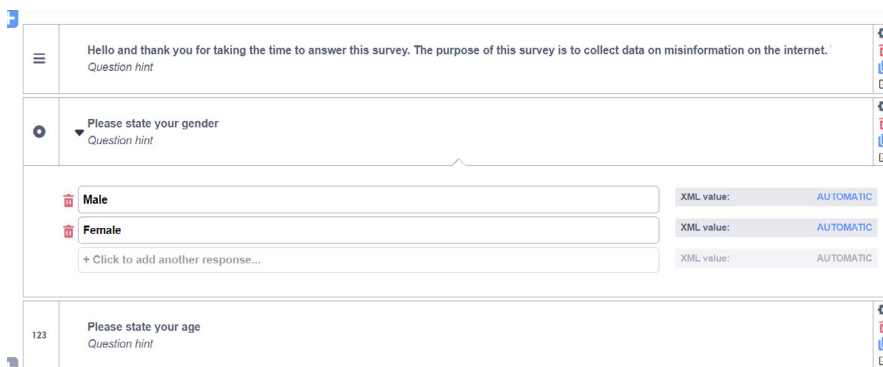
3. Be sure to limit the age question to accept reasonable values i.e. No values above 100. We will then add questions about misinformation spotted online.

4. For the last two questions, it is important to add a skip logic criteria based on the question that asks if the respondents have spotted any misinformation online. This ensures that the two questions only appear if the respondent responds with a **Yes** to the third question.

5. Finally, save your form and deploy it. You are now ready to share your form with respondents. Under the Form tab find the data collection URL and send it to your respondents.
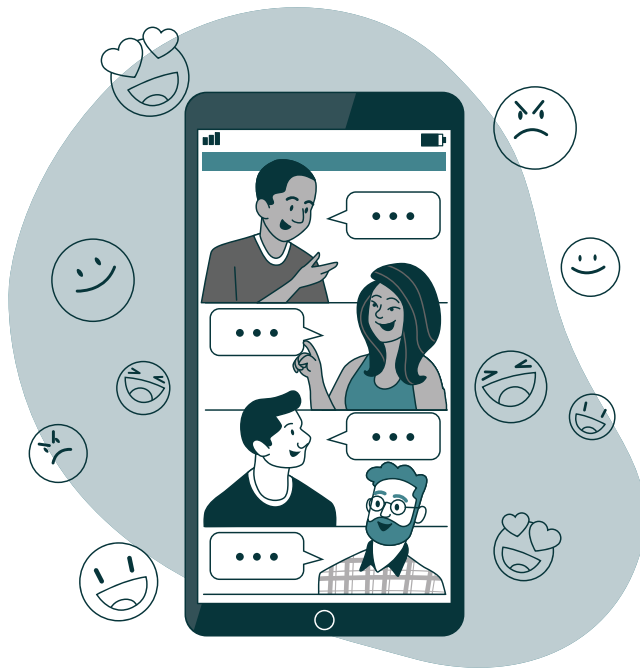
Module 3
Combating Hate Speech: Strategies and Practices – Part 1

Effectively combating hate speech is as complicated as the nature of the problem. Over the course of the next two modules (Module 3 and 4) we will explore the different strategies, tactics and practices organizations and stakeholders all over the world have developed and tried to address this phenomenon and mitigate or prevent its worst consequences.

To manage harmful content is to take action. There are various ways one can take action, and in this module, you will learn how the following strategies can reduce conflict and violence: Direct Response to Hateful Tweets or Messages, Dialogue and Other Engagement: In person and Digital Dialogues, Building Response Mechanism to Prevent or Mitigate Offline Violence, and Strategic Nonviolent Action.

Learning Objectives

By the end of this module, you will...

- Become familiar with practical methods to combat hate speech
- Identify methods that will address the problem you have identified in your context

Featured Tech Tools

Crowdtangle Link Checker is a Google Chrome extension that searches for the URL you have specified through Facebook, Twitter, Reddit and Instagram to see if any of the accounts, Pages, or profiles in the CrowdTangle database have shared that link. This can be useful to check if and how a certain link has been shared across these social media network sites.

> Here is the Crowdtangle Link Checker trailer video part 1
>
> Here is the Crowdtangle Link Checker trailer video part 2
>
> Here is the Crowdtangle Link Checker training video

Discussion Questions

- Choose one of the methods presented in this module and discuss: How can it be effective in combating hate speech in your context? What strategic considerations does it entail?

- Which methods do you think are most well-suited for reaching different audiences? (hate speakers, the audience/those reading the hateful content, tech companies, etc)?

Featured Resources

Below is a collection of resources and guides that offer different strategies to combat hate speech:

- Starting Points for Combating Hate Speech Online
- Guide: Defusing Hate
- Five Ways to Counter Hate Speech in the Media Through Ethics and Self-Regulation
- Ten Ways to Fight Hate

- ACLU: Fight Hate Speech with More Speech
- Addressing Hate Speech on Social Media: Contemporary Challenges
- Counterspeech: A Literature Review
- Psychology and Hate Speech: A Critical and Restorative Encounter
- A Primer on Rumor Control
- Technical Brief: Creating a Real-Time Rumor Management System for COVID-19
- Countering Fake News: A Comparison of Possible Solutions Regarding User Acceptance and Effectiveness

Activity

- For this module's activity, we would like you to analyze an action you took – with your organization – that aimed to combat hate speech. This could be an action or counterspeech effort, or any other method of countering hate speech described in **module 3** and **4** (*If you have not participated in any anti-hate speech activity before, choose one of the methods discussed in **module 3** or **4** and answer the questions on the worksheet hypothetically.*)

- Use the **activity template** for your analysis.

- You will share some of the insights you gained reflecting on this action.

Content Breakdown

In this module you will find:

Category 1:   **Direct Response to Hateful Tweets or Messages**

- Report hate speech

- Counterspeech

- Countering high profile hate speakers

- Amplifying positive voices

- Countering mis- and disinformation

Category 2:   **Dialogue and Other Engagement: In-person and Digital Dialogues**

- In-person or digital dialogue

- Digital responses at scale

- "Naming and shaming"

Category 3:   **Building Response Mechanism to Prevent or Mitigate Offline Violence**

- Network of trained first responders

- Network with other stakeholders with capability to respond/react

Category 4:   **Strategic Nonviolent Action**

Submodule 1: Direct Response to Hateful Tweets or Messages

Report Hate Speech

The most common type of intervention is a "direct reaction" by censorship or responding directly to the hateful message. When you come across hate speech on social media platforms, there are steps you can take to report this speech to the social media companies. Platforms such as Twitter, Facebook, YouTube, Instagram, TikTok, etc. have tools in place that allow users to report harassment and hate speech, though too many reports tend to still go unaddressed.

1. Flag and report threatening messages online, as well as those who spread them. If these platforms review the post(s) and decide the content does in fact violate the platform's terms of use, they may delete the post, and suspend or permanently ban the person who made the post you reported.

2. Where possible, gather evidence. This will help with any investigation if you decide to report to the police or one of the other possible support organizations. The best way to gather evidence of online hate is by taking screenshots of the posts or messages and/or saving the webpage link if relevant (for example on a webpage or YouTube). You can also use the "print to .pdf" function to save copies of webpages if you are concerned that the content may be removed by the poster or site later.

Case Studies

Reporting Violent Extremism on Social Media

Fight Against Hate (FAH) is a crowdsourcing platform that gathers incidents of online extremism found on various social media platforms. FAH invites internet users who see violent extremist content on social media to report it to the platform as well as to FAH, so that FAH can monitor how serious the threat is, how long it takes for the social media platform to remove the reported content, and escalate the report to higher authorities should it take too long to remove. Internet users can report the hateful content in three simple steps: 1) copy the link of the hateful content on fightagainsthate.com, 2) select the type of hateful content that it promotes, and 3) submit the report. Reports received by FAH will be useful for NGOs and researchers as a tool to access online hate speech data; for government and law enforcement to track terrorist organizations, cyberbullying, and hate speech that may incite hate crimes and violence; and for Internet users to interact with others to discuss hateful contents.

Featured Resources

How to report hate speech
- Facebook
- Twitter
- Instagram
- Youtube
- TikTok
- WhatsApp

Questions to Examine

- How do you determine whether a message is considered as hate speech?
- What are the different ways to check the validity of claims in messages?

Counterspeech

Counterspeech, according to the [Dangerous Speech Project](#), "is any direct response to hateful or harmful speech which seeks to undermine it."

Censoring hate speech is sometimes ineffective and may promote a discourse of "persecution" on the side of the people spreading hate speech. Instead, refuting discriminatory ideas through providing facts, pointing to logical inconsistencies in hateful messages, supporting the victims and spreading neutral messages can be more effective in curbing these harms of hate speech than censorial measures.

There are a few 'types' of counterspeech:

- **Organized counter-messaging campaigns or organized collective counter-speech**

 Case Studies

Panzagar ("Flower Speech") New Responses to Hatred Online

In Myanmar, the Islamic Rohingya people are in the midst of a genocide. In April 2014, as a form of counterspeech, tech activist Nay Phone Latt launched Panzagar ("Flower Speech"). The movement started off as a meme of a cartoon woman holding a flower to her mouth, as a commitment to not engage in hate speech. The movement went viral, and the Panzagar Facebook page received thousands of "likes" within days of its creation, and many activists published photographs of themselves with a flower in their mouth, standing in solidarity with their Rohingya brothers and sisters in the movement against hate speech. This is an example of how counterspeech can have a mobilizing effect.

Collective Counterspeech: The Anti-Hate Brigade, #jagärhär

In 2016, a Swedish woman named Mina Dennert started the [#jagärhär movement](#) in 2016 to counter online xenophobic hate speech in Europe. Dennert started the movement by identifying hateful misinformation on Facebook, and responding with accurate information and inclusivity. She recruited approximately 20 friends to engage in this counterspeech, and they ultimately founded the #jagärhär (which means "I am here") Facebook group and activist movement. The group members share links to hateful posts in the Facebook group, and then the organization's members respond to the hateful post, providing factual information. The members "like" and engage with each other's responses, to boost their comments to the top of Facebook's public pages. They use Facebook's algorithms to amplify their human rights activism, spreading factual and antiracist materials with the hashtag #jagärhär, while burying hateful comments. The organization is thriving and has expanded to many other countries. 146,000 people have joined one of the Facebook groups. This is a beautiful example of citizens banding together to promote peace through counterspeech.

- **Non-organized: Spontaneous, organic responses**
  These are informally produced counter messages that contest extremist messaging but lack the support of governments, large institutions, or civil society organizations.

Case Studies

Internet users can decide to fight hate on their own, and refuse to have their online spaces taken over by extremists and messages that direct hate at other groups. In 2013, Kenya was in a contentious presidential election. iHub Research (a tech company) monitored Kenyan online speech during the election cycle for online hate speech that could incite violence. This study was conducted using Ushahidi, a computer software that allows users to crowdsource data, and the study found that hate speech was abundant in newspaper comment sections and on Facebook. In contrast, iHub found that among Kenyans on Twitter, there was dramatically less hateful and violent speech, because there was increased user-driven counterspeech. To counter hateful tweets directed at ethnic minorities, other Twitter users responded with tweets such as "please remember that we are all Kenyan" or "Is this the Kenya that we want?" The results of this informal counterspeech response were apparent. At least one of the original accounts produced an apology, others stopped Tweeting hatred, and some disappeared entirely.

Hate Speech vs. Counterspeech in Times of COVID

The spread of COVID-19 has sparked racism, hate, and xenophobia in social media targeted at the Chinese and the broader Asian communities. The Georgia Institute of Technology studied the evolution and spread of anti-Asian hate speech through the lens of Twitter. To do so, they created COVID-HATE, the largest dataset of anti-Asian hate and counterhate spanning three months (i.e., from January 15, 2020 to April 17, 2020), containing over 30 million tweets. By creating a novel hand-labeled dataset of 2,400 tweets, they trained a text classifier to identify hate and counterhate tweets and finally we identified 891,204 hate and 200,198 counterhate tweets in COVID-HATE. Using this data, they conducted a comprehensive overview of anti-Asian hate and counterhate speech on Twitter during COVID-19.

From January 15, 2020 to April 17, 2020, a total of 891,204 hate tweets and 200,198 counterhate tweets were made. Hate content is always seen to exceed counterhate in terms of the number. (See chart below)
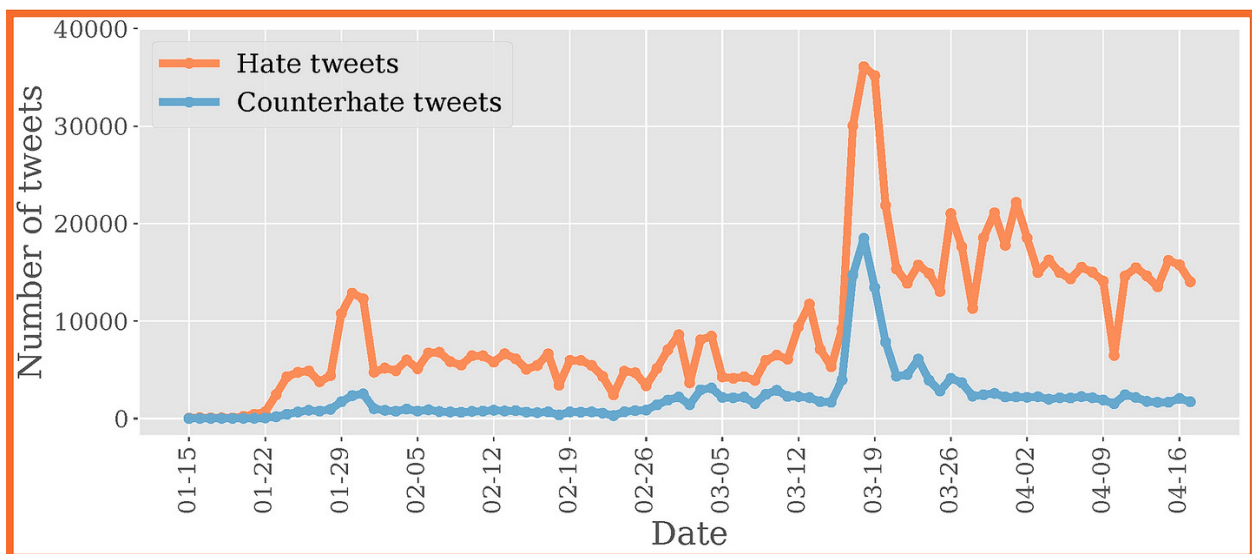
Kenyans on Twitter Counter CNN

Spontaneous organic responses can also be directed at changing perceptions about a people or a country, as Kenyans proved when they took on media giant CNN after the former had called Kenya "a hotbed of terror". This was said in the wake of then US President Barack Obama's impending visit to the East African nation. Tired of the narrative of always painting African nations as war-torn, Kenyans on Twitter used the hashtag #SomeoneTellCNN to change the discourse and showcase the many positive things about Kenya. Twitter users across the country listed the many positive aspects of the country, from its beautiful landscapes to award winning athletes and most importantly, peace-loving public.

💡 Featured Resources

- **WE CAN! Taking Action Against Hate Speech Through Counter and Alternative Narratives**
  This is a step-by-step guide to developing a counter or alternative narrative campaign.

- **Backgrounds, Experiences and Responses to Online Hate Speech: A Comparative Cross-Country Analysis**
  This study showcases young people's use of social media and their understanding of online hate speech in five European countries and how they respond and react to hate speech.

- **Counterspeech on Twitter: A Field Study and Considerations for Successful Counterspeech**
  This qualitative study of counterspeech practiced on Twitter offers suggestions on which strategies may help make counterspeech successful and the ineffective or discouraged strategies at influencing the speaker that are possibly counterproductive or harmful.

- **Thou Shalt Not Hate: Countering Online Hate Speech**
  This study focuses on counterspeech as an alternative strategy to provide a defense against hate content online, as the current policies put forward by online platforms to combat it does not seem largely successful.

Successful Counterspeech on Twitter

Lucas Wringt[4], Derek Ruths[2], Kelly P Dillion[3], Haji Mohammad Saleem[2], and Susan Benesch[1,4]
[1]Berkman Kein Center for Inernet & Society, Harvad University
[2]School of Computer Science, McGill University
[3]Department of Communication, wittenberg University
[4]Dangerous Speech Project

Overview

A study of conversations on Twitter found that some arguments between strangers led to favorable change in discourse and even in attitudes. We propose that such exchanges can be usefully distinguished according to whether individuals or groups take part on each side, since the opportunity for a constructive exchange of views seems to vary accordingly. We describe four corresponding "vectors."

What is successful counterspeech?

Counterspeech is a direct response to hateful or harmful speech, and speech is any form of expression: text or images. Success can mean either:

• The author of the original speech favorably changes beliefs or online behavior.

• The 'audience' (those who witness the exchange) are favorably influenced i.e. less likely to express themselves hatefully or harmfully online, in the future.

Four Vectors of Counterspeech

One-to-one
Exchanges, often lengthy, between two accounts. Remarkable shifts in tone were observed.

In honor of MLK day today, I'm taking a vow to use the word "nigger" as many times as possible and in the most inappropriate times.

"Let no man pull you so low as to hate him" – Martin Luther King Jr.

"I hope I get shot soon" – Martin Luther King Jr.

[...] Does your mom know you spend your time on the internet trying to hurl people?

I doubt it. She's been dead for a year and a half now so...

I'm sorry for your loss. And I hope you find a better way to honor her.

you are so nice and i'm so sorry

One-to-many
A single user responds to many others. Can also be described as multiple one-to-one exchanges.

I was prompted to write this after a recent Facebook discussion [...] when the term "abeed" (slaves) was used in a thread, in reference to a news story about an African American woman, who flashed an Arab American businessman in Detroit, during a verbal dispute. What was disturbing [...] was not simply the racist comments that were used about the unruly woman, but that some showed a profound lack of empathy when I mentioned that the term "abeed" is a hurtful word.

The basis of Islam is equality before God. So how can we tolerate derogatory, racial slurs and still call ourselves Muslim? #DropTheAWord

Many-to-one
Viral counterspeech overwhelms its target. Does it convince, or silence?

I am literally soo mad right now a ARAB won #MissAmerica

One day I hope you realize how shameful this tweet is. I hope you realize it tomorrow.

your hatred made it onto sky news! Congrats!

Ignorant/illiterate/racist/idiot

@MissAmerica sorry for being rude and "racist" and calling you a Arab please tweet back so everyone will know its real

Many-to-many
Competing hashtags, or struggle over control of/opinions expressed on a single hashtag.

It's going to be -27 without wind chill tomorrow morning and I have class at 8 #FuckPhyllis #Cunt #Bitch #Whore

In a room with Phyllis Wise, Adolf Hitler, and a gun with one bullet. Who do I shoot? #fuckphyllis

As an alum, I'm completely embarrassed with the behavior of some of you assclowns. #fuckphyllis. If I were UIUC I'd publicly shame all of u.

Oh and good luck, bitching students. your asinine and ignorant tweets are viewed by EVERYONE NOW #fuckphyllis

So happy to see the #fuckphyllis tag has been taken over by people calling out the racism that was in it.

NLP Challenges and Opportunities

Computational approaches are required in order to study and engage counterspeech efforts at scale. This has not yet been attempted.

Detection is unusually difficult, for reasons including these:

• counterspeech can involve a broad range of audience sizes - from single counterspeakers to whole communities.

• a single counterspeech act can exhibit a variety of communicative strategies.

These observations should be useful for detection of counterspeech - and hateful or abusive language - and for designing more successful counterspeech.

Future research

What are the norm-influencing capacities of hashtags around public events and controversies? How do those hashtags break through like-minded silos that prevent the exchange of ideas?

Counterspeech DOs and DON'Ts

A tool for countering online hate and harassment - without making things worse

Note to counterspeakers:

Every situation is different, and counterspeech doesn't always work. Sometimes people are determined to hurt, are convinced they're right, or both.

If you choose to engage, here are some tips for getting positive results.

At those times it might be best to disengage or use other tactics, * so use your best judgment.

Before You Start

Protect yourself – take steps to protect yourself from retaliation (see the resources at iheartmob.org/tech)

Think about how your online identity or profiles could be used against you.

Think about what to accomplish. do you want to change ther person's mind, or how they post or tweet? Stop them from attacking someone else? Change other people's mind or behavior?

Remind yourself that behind each comment – no matter how hateful – is a human being. Treat them as you would want to be treated.

Counterspeech DOs: Things you can try whn you feel safe

Stay calm. If you're upset, wait a bit before responding.
Ask questions, like "Why do you think that?" or "What do you mean?"
**Refer to potential outcomes,** like "That could hurt someone."

Try humor. If you intent is kind and you're not mocking the person, humor can soften the exchange and attract others to show their support. Conter with images that are silly, clever, or funny – not hurtful – to de-escalate.

Label the comment, not the person, like "That word comes from a racist stereotype."
Show empathy and connection with the target ("I'm Asian American too, and...") or with the speaker ("I'm angry about this too, but...").
Start a supportive hashtag like #LoveForLeslieJones.
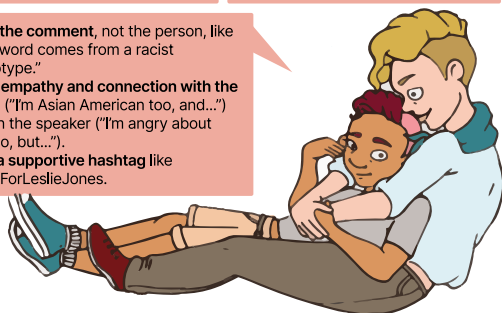
Counterspeech DON'Ts

Don't label people – for example, calling them a bigot.
Don't assume the person has bad intensions.
Don't be hostile, insulting, or aggressive – it can escalate the conflict.
Don't talk down to the person – it can shut down communication.
Don't nitpick or correct spelling or grammar. Use a civil tone and link to a source if you want to correct false information.
Don't silence the person with threats, social exclusion, or other punishment.

Find our full resource at CounterspeechTips.org.
For more tips on counterspeech on - and offline harrasment, visit ihollaback.org/resources/bystander_resources.

- Have you ever replied to a hateful comment with correct facts and debunked what had been posted? How did you do that?

- Which one is more effective: engaging in meaningful discussion with people who spread hate speech or ignoring them or deleting their comments? Why?

Countering High Profile Hate Speakers

High profile persons (e.g., political or religious leaders, celebrities and journalists) have a particular responsibility when it comes to hate speech. Because of the influence, authority and reach their position or status affords them, hateful messages they spread have higher potential of causing substantial online and offline harm.

The Dangerous Speech Project notes that influential speakers can make violence seem acceptable and necessary, but they can also favorably influence discourse through counterspeech. The most direct way this approach can succeed is to have a positive effect on the speaker, convincing him or her to stop speaking dangerously now and in the future. It can also succeed by having an impact on the audience – either by communicating norms that make dangerous speech socially unacceptable or by 'inoculating' the audience against the speech so they are less easily influenced by it.

Here are are three suggested strategies:

1. **Pressure to retract statement**

   Community actors and activists can work to pressure high profile hate speakers to retract hateful messages by making an official statement. On this occasion they may also want to raise the public's and the hate speaker's awareness about the consequences such narratives can have.

2. **Ask other high profile persons to make statements and exercise pressure (politicians, celebrities, or positive influencers)**

   Working together with other high profile individuals can help counter hate speech by drawing people's attention to hate speech and its negative effect on the communities. They can also exercise peer pressure on the high profile speaker. Social media influencers often have a solid base of followers, and spreading positive messages that condemn hate will encourage more followers to practice these words.

3. **Government agencies working to ensure high profile individuals don't use hate speech words or inciting statements**

   Leading up to periods that may bring tension such as campaign periods, government agencies that monitor elections may choose to outlaw words and statements that would be considered hate speech. This would potentially reduce the incidences in which the high profile individuals use words that can cause harm. A limitation of this approach is that it can be prone to abuse by governments, and some high-profile individuals who would use impunity to get away with hate speech.

Case Studies

Countering High Profile Hate Speakers in Nigeria

To raise the awareness of the impact of hate speech amongst society, it is important to reach out to hate speakers, and explain how dangerous speech is one of the main reasons behind incidents of violence due to its broad reach online and offline. This study highlights the effectiveness of direct outreach to these speakers through social media platforms.

And it shows that although they might show resistance at first, using a constructive and a friendly tone to influence those who engage in dangerous speech usually influences them to engage in meaningful conversations.

Former Governor of Katsina State in Nigeria, Ibrahim Shehu Shema, once described his political opponents as "cockroaches worthy of being killed." The Centre for Information Technology and Development (CITAD) responded by mobilizing several civil society organizations and initiated three joint press conferences on the matter ensuring a wide public reach. This publicity led the Governor to respond by claiming he was quoted out of context thus discrediting his comments and removing any sense of associated authority.

Use of "Hate Speech" Laws and Monitoring of Politicians on Social Media Platforms

The National Cohesion and Integration Commission of Kenya was created from a need to mitigate ethno-political competition and ethnically motivated violence, that characterized the 2007 / 2008 Kenyan post election violence. In Kenya senator Mithika Rinturi was arrested on January 9, 2022 for his utterances that evoked memories of past episodes of the 2007/2008 pre and post-election ethnic violence where certain communities were attacked for not supporting a candidate from the larger host community.

As much as hate speech laws and monitoring of high profile speakers might be seen as an efficient technique to prevent harm emanating from speech, it may also be an avenue for some governments to suppress the opposition, media representatives, civil society actors, and the general public for legitimate speech and dissent; in some cases high-level hate-mongers also escape with impunity.

Featured Resources

The UN has highlighted the important role played by religious leaders in combating (and not contributing to) hate speech.

- Plan of Action for Religious Leaders/ Actors from the Asia-Pacific Region to Prevent Incitement to Violence That Could Lead to Atrocity Crimes

  This action plan was adopted by more than thirty religious leaders and actors representing different religions and faith traditions from the Asia-pacific region to develop a strategy to prevent incitement.

- #Faith4Right Toolkit: Incitement to Hatred

  The #Faith4Rights toolkit suggests prototypes of peer-to-peer learning modules, exploring the relationship between religions, beliefs and human rights. Module 7 of the toolkit specifically addresses the issues of incitement to hatred.

- Ten Ways to Fight Hate Guide: Pressure Leaders

  Elected officials and other community leaders can be important allies. But some must overcome reluctance—and others, their own biases—before they're able to take a stand.

Questions to Examine

- Tell us an example of high-profile individuals who have been countered for using hate speech terms in your country! What did they do to counter the hateful speaker?

- What strategy would you use to help counter hate speech committed by high profile speakers in your community?

Amplifying Positive Voices

In addition to countering hateful messages, amplifying positive and unifying voices can have a positive effect and has shown to have an impact. Raising other people's voices is critical in advancing discussions around countering hateful messages. Some of the ways in which you can amplify positive voices are:

1. Sharing the work of those working to curb hate speech. This could be sharing their websites, blogs, podcasts, videos and other creations that mitigate hate speech.

2. Recommending their work or creations to your friends and colleagues.

3. Actively promoting and supporting others in spaces in which their work is relevant, and making space for their voices and creations – within your curriculum, newsletters, websites, podcasts, social media spaces, etc.

Case Studies

A Campaign That Supports Women Peacebuilders

She Builds Peace is a global collaborative campaign that supports women peacebuilders to soar, by ensuring their safety and protection, fulfilling obligations to make peace and security inclusive, and appreciating and resourcing the critical work women peacebuilders do to build a future in which all can flourish. It is a campaign by The International Civil Society Action Network (ICAN) and the Women's Alliance for Security Leadership (WASL). The campaign includes a campaign kit that has tools and materials to help one implement the She Builds Peace campaign. The Participation, Protection and Funding Frameworks each consist of an analytical brief, operational guidance, pledge, and action tool.

EU Coalition of Positive Messengers to Counter Online Hate Speech

The Positive Messengers project aimed to build a EU-wide "coalition of positive messengers" trained to step in when needed to assist individuals affected by online hate speech and work towards making the internet a safe and inclusive environment for all. The project was built on a study and research of best practices of countering hate speech targeting refugees, minorities and migrants conducted in 10 countries, focusing on raising awareness and building capacity on positive messaging. It also provides data on the impact of hate speech on developing more effective anti-discrimination policies in the EU.

Giving Voice to the Voiceless or "Help Speech"

The Language Technology Institute (LTI) used AI to search more than a quarter of a million comments on social media to automatically identify the fraction that defended or sympathized with a marginalized group (in this case Rohingya refugees). Human social media moderators, who couldn't possibly manually sift through so many comments that quickly, would then have the option to highlight this "help speech" in comment sections, thus elevating the voices that support the Rohingya, who themselves do not have the resources to counter hate speech by themselves. Finding and highlighting these positive comments, the study suggests, "might do as much to make the internet a safer, healthier place as would detecting and eliminating hostile content or banning the trolls responsible."

Featured Resources

- How to Identity and Report Hate Speech on Social Media
- Combating Hate Speech Through Counterspeech
- Defusing Hate: A Strategic Communication Guide to Counteract Dangerous Speech

- [Counterspeech on Twitter: A Field Study](#)
- [How to Counter Dangerous Speech Online](#)
- [Hate Speech, Social Media, and Religious Minorities](#)

Questions to Examine

- What have you done to amplify positive voices online and offline?
- Are there any positive voices in your community that you can identify and would like to amplify?

Countering Mis- and Disinformation

Hateful speech is often false, and therefore responding to it requires fact-checking. False information may be spread by those who believe it to be true (misinformation), or with the intention to deceive, by those who know it is false (disinformation). The following case studies and resources describe the best practices for responding to hateful mis/disinformation.

Case Studies

#ThinkB4UClick – Raising Awareness on the Misuse of Social Media

#Defyhatenow trains citizens and community based organizations to become "positive influencers" through counterspeech, fact-checking, peacebuilding, education, and social media monitoring. The #thinkB4Uclick campaign was a one-month campaign that had the goal of creating public awareness about the dangers of misinformation, fake news and hate speech. Each week, the organization promoted content based on a specific theme. The four themes were: 1) checking and verifying the sources of information, 2) how quickly lies travel and how hard it is to retract untrue statements, 3) the importance of context, and reading before responding, and 4) how to have a positive impact.

Indivisible Truth Brigade

The Indivisible Truth Brigade is a collective effort to counter disinformation in the United States. Members are discouraged from directly engaging with disinformation; instead they share united, factual messages through their own social networks in an effort to amplify the truth. The group designs its campaigns to counter the specific disinformation narratives that are trending at the time.

Featured Resources

- [Countering Disinformation: A Guide for Promoting Information Integrity](#)
- [Countering Misinformation Via WhatsApp: Evidence from the COVID-19 Pandemic in Zimbabwe](#)
- [Tools That Fight Disinformation Online](#)

Questions to Examine

- What are some ways that you verify the truth of the content you are consuming?
- How might you discuss the problem of mis-/disinformation with someone who believes the information is true, even when presented with evidence it is false?

Submodule 2: Dialogue and Other Engagement: In-person and Digital Dialogues

The aim of dialogue in countering hate speech is to engage diverse and divided people in a constructive conversation in order to break down stereotypes and rebuild trust. The intended result of dialogue is that both the person spreading hate speech and the person confronting them gain an understanding of others' ways of feeling, thinking and expressing themselves, which then develops empathy towards each other and allows them to bring about change.

Engaging in dialogue with someone spreading hate speech does not mean one endorses their comments, or are offering them a platform to spread their beliefs, rather try to understand their motivations and explain their misconceptions and their effect on people.

In-Person or Digital Dialogue

Some initiatives work to engage directly with people who may be unintentionally posting hateful and/or dangerous messages to make them understand the true nature and impact of their speech. In his book "Conversations with People Who Hate Me," social justice activist Dylan Marron offers an insight into how dialogue can make spreaders of online hate reconsider their actions, and some even come around and realize what they are saying is wrong and a result of misinformation or misconceptions. There are many individuals, journalists and organizations that are engaging in dialogue with hate mongers to try and reduce the instances of online hate speech.

The old way of having conversations needs to expand to include a wider angle and different guests. "I imagined I'd be bringing together hater and target on bigger subjects. I'd learn more about people – and in the process, I'd learn more about myself. A few years ago I began a social experiment in which I phoned my internet detractors and engaged them in conversation. This concept grew, and soon I was moderating discussions between others who had clashed online as well." *Marron, Dylan (2022). Conversations with People Who Hate Me: 12 Things I learned From Talking to Internet Strangers; Atria Books.*

Case Studies

Reaching Out to Dangerous Speakers in Nigeria

The Nigeria Stability and Reconciliation Programme (NSRP) created a network of 'online peace ambassadors' who reach out to the hate speakers via Facebook or Twitter "explaining the danger of specific statements with requests to embrace more peaceful alternatives for expressing political opinions."

The Commons Project

The Commons Project is a response to the current, challenging political and social climate in the United States. Many people are observing and experiencing a decrease in constructive conversations, respect and open-mindedness in their everyday interactions, in the media, and in politics. Among the different factors contributing to this are the ways that social media platforms, including Facebook and Twitter, are built and shape our communication and interactions. The Commons identifies people engaged in political discussions about the USA on Twitter and Facebook, analyzes the likelihood that they are at risk of polarization, and engages them in conversation. It facilitated conversations seeking to help people understand and make different choices in their interactions,

online and offline, particularly around political differences, and offer skills and resources to promote constructive conversations, listening, and respect.

Featured Resources

- Empathy is Not Endorsement
- Australia Using Dialogue and Therapy as a Cure for Online Hate
- Digital Dialogue: Countering Online Hate Speech

Questions to Examine

- Can you give examples of which situations you think would benefit from dialogue in terms of online hate speech?
- What are some of your concerns as a peace advocate in doing in-person or digital dialogue?

Digital Responses at Scale

Social and digital media users, predominantly networked cultures on various platforms, have generated responses attempting to combat hate speech. These user-generated responses use approaches and tactics that rely on the scale and reach of social media to circulate counterspeech. Hashtag takeovers, sometimes also called *hashtag hijacking* or *keyword squatting*, is a primary tactic used to circulate counterspeech and disrupt online hate speech at scale. The process involves identifying hashtags used by accounts spreading hate and implementing a "takeover" of the hashtag by flooding it with counterspeech, including texts, images, and video clips. This started as a primarily Twitter-based tactic, but because hashtags are now ubiquitous beyond the Twitter platform, it can be used on a variety of platforms. Such campaigns can be successful in the short term (they are not often sustained over long periods of time), especially if news media report on them.

Case Studies

Stealing #StopIslam

The Stop Islam hashtag was already circulating in 2016, however is spiked in viral popularity (an increase from 1,050 tweets to over 400,000 tweets) during and directly after the terrorist attack in Brussels, Belgium that year. The initial campaign was filled with hateful and Islamophobic commentary and linked to anti-immigration/refugee sentiment in multiple nations with the primary use coming from US accounts, high use coming from accounts in the UK and Pakistan, and large numbers of non-European actors further highlighting the transnational nature of online hate campaigns. In response, a counter narrative campaign developed using the Stop Islam hashtag to resist Islamophobia, defend Muslims, and circulate counterspeech. The campaign included written statements, infographics, and memes. The counterspeech campaign was robust and eventually reported on by mainstream news media expanding, at least temporarily, the circulation and reach of the action.

Multi-Hashtag/Multi-Platform K-Pop Campaigns against Hate (Online)

During the summer of 2020, multiple online counterspeech campaigns emerged on social media, predominantly Twitter and Instagram, taking over a variety of hashtags. Initially these campaigns took over #WhiteLivesMatter, #BlueLivesMatter, and #QAnon in support of Black Lives Matter protests in response to the death of George Floyd at the hands of US police. These multi-hashtag campaigns provided counterspeech to hateful and racist comments and images. K-pop (Korean popular music) fans generated a networked social response campaign where their counterspeech consisted of flooding the hashtags with videos clips of their favorite K-pop songs. The goal of the campaign was to disrupt hate and racism

on the platforms by flooding them with tens of thousands of posts and saturating both the hashtag and the platform trends with more positive media. This campaign was also amplified when it was picked up and circulated by the mainstream news media.

Featured Resources

- [Eight Massive Moments when Hashtag Activism Actually Worked](#)
- [Understanding the Versatility and Necessity of Hashtag Activism](#)
- [Hashtag Activism: Tips on How to Leverage Social Media for Social Change](#)

Questions to Examine

- What could be some potential problems with user-generated responses to hate speech?
- As a peace advocate, what do you think is necessary to both encourage users and increase their capacity to respond at scale on social media?

"Naming and Shaming"

To name and shame is to "publicly say that a person, group or business has done something wrong." It is used to discourage certain activities. It is one of the common practices to highlight human rights abuses at the state level. Beyond the debate as to whether it is an effective tactic to dissuade human rights violations, in the context of combating hate speech, it raises several issues: some studies have suggested that in some cases it may not silence, change minds, or significantly influence the opinion of those who spread hateful messages. On the contrary, it may even contribute to deepening divisions and increase polarization by pushing the opposite side in its trenches. On the other hand, some actors maintain that it can have a dissuasive effect.

Case Studies

Naming and Shaming in Kenya by a National Commission

The civil society and government agencies in Kenya such as the National Cohesion and Integration Commission (NCIC) has routinely put on notice various members of parliament among them former Nairobi governor Mike Mbuvi Sonko over hate speech. In preparation for the 2022 elections the National Cohesion and Integration Commission of Kenya announced that it will have a wall of shame to stem hate speech and violence during elections.

Featured Resources

- [The Case for Naming and Shaming White Supremacists](#)
- [When the Tactic is Weaponized: What Are Your Rights if You Are 'Named and Shamed'?](#)
- [Name, Shame, and Then Build Consensus? Bringing Conflict Resolution Skills to Human Rights](#)
- [Why ISIS is Immune to 'Naming and Shaming'](#)

Questions to Examine

- Under what circumstances can "naming and shaming" work to counter hate speech?

Submodule 3: Building Response Mechanism to Prevent or Mitigate Offline Violence

Building response mechanisms to prevent or mitigate offline violence not only can include early warning and early response (EWER) operations that are set up in preparation of specific events (like elections), but also more long term programs that seek to combat and prevent offline violence or support victims of hate speech in a more systematic way.

EWER systems involve observing and monitoring social media for evidence of harmful content, such as incitement and misinformation, and monitoring whether the conversations are moving towards escalation. Monitoring provides an early warning of impending conflicts or re-escalation of ongoing ones. It helps analyze, understand, and ultimately contribute to reducing violence.

Network of Trained First Responders

These networks can include persons trained to intervene and diffuse potentially violent situations. For example, during an election, first responders are dispatched in response to online rumors of fraud or violence to diffuse the situation or re-direct voters to other polling stations.

First responders also act as regular data collectors and analysers on conflicts, by systematically monitoring and reporting conflict indicators. They alert decision makers to the potential outbreak, escalation and resurgence of violent conflict; and promote an understanding among decision makers of the nature and impacts of violent conflict. They generate results, based on quantitative and qualitative analytical methods which help formulate scenarios and response options that are communicated to decision-makers. Depending on the nature of conflict or violence some first responders may remain anonymous.

Case Studies

A Local First Approach to Early Warning and Response in Kenya

In preparation of the 2013 Kenya election, local initiatives created a nation-wide early warning system that would provide real-time information, in order to allow immediate intervention. Its approach was based on training hundreds of volunteers, providing them with cell phones, and creating a local call-in center for warning and response. The system was built in coordination with local community leaders who would respond to warnings by quickly visiting the site of the report to assess the situation and engage directly with those impacted to develop response options.

A Citizen Journalism Framework That Crowdsources and Maps Conflict Incidents in Cameroon for Rapid Response

Through the Early Warning Early Response (EWER) system, called the MUNGO, #defyhatenow in Cameroon plots, reports, and highlights violent incident patterns and hotspots in Cameroon's conflict regions and maps conflict incidents, allowing for a more rapid response. The first responders are some of the alumni who participate in a fact checking fellowship under the project in Cameroon.

The MUNGO also remains an anonymous platform due to the sensitive nature of the Anglophone Conflict in Cameroon to which access is limited to #defyhatenow trusted partners. This is to protect the identity of the responders who may otherwise be

subjected to harassment by the government or the separatists in the anglophone region of Cameroon.

Network of Stakeholders with the Capability to Respond/React

This can include networks that systematically relay information to law enforcement, or other support services for perpetrators (those who seek to exit extremist groups for example) or victims of hate speech (like psychological support).

Case Studies

An EWER Tool Showing How the State and Its Partners Can Advance Commitments to Work for Peace and Cohesion

The Kenya Community Support Centre (KECOSCE's) has created a Conflict Early Warning and Early Response (EWER) practitioners guide, a critical tool in contemporary conflict prevention and peace building. The guide shows us how the State and its partners can advance commitments to work for peace and cohesion. KECOSCE has been successful in its EWER interventions because of its strategic partnerships with relevant public institutions mandated to ensure peace, security and stability in Kenya. At the heart of the functionality of the EWER mechanism is data collection, analysis and the timely dissemination of information and knowledge to the right people and institutions.

INHOPE: Hotlines and Complaint Forms

INHOPE, the International Association of Internet Hotlines, coordinates a network of global internet hotlines, supporting the individual hotlines in addressing reports of illegal or hateful content online. The hotline exists to promote internet safety by removing illegal material from the Internet efficiently and transparently to enable investigations by law

enforcement. INHOPE provides guidance and mentorship to member organizations as they establish and operate their hotlines. Once a report is received by a hotline, the report is added to the hotline database system. Trained hotline staff members, who are well equipped to assess Internet content, determine if the content is illegal under existing local legislation. If the reported material is presumed illegal, the hotline carries out an examination to identify the origin of the material. This process enables law enforcement to efficiently receive analyzed reports on dangerous speech in their region, thereby establishing an effective response mechanism to offline violence.

Featured Resources

- Early Warning in Responses to Hate Speech and Divisive Narratives in Conflict-Affected Contexts

Questions to Examine

- How does an Early Warning and Early Response (EWER) system work to mitigate hate speech?
- How can you use EWER systems to mitigate conflict in your community?

Submodule 4: Strategic Nonviolent Action

The field of strategic nonviolent action can offer additional insights and tactics that you can add to your toolkit when combating hate speech.

Nonviolent action is the practice of achieving goals (such as political or social gains) through collective and nonviolent actions like mass protests, boycotts, strikes, sit-ins and various forms of non-cooperation. There are well over 200 recorded nonviolent tactics and the key 'ingredient' of this discipline emphasizes the strategic use of these tactics in a way that shifts the power dynamics at play on both sides of a given cause (or struggle).

Case Studies

English Sports Fraternity Boycott of Social Media to Protest Online Abuse

One of the most recent boycotts that received worldwide attention was the English football, rugby, F1, netball, and cricket clubs and players social media boycott in May 2021 to protest the abuse and discrimination their players faced online. Dubbed #StopOnlineAbuse, the boycott which lasted for four days across all social media platforms aimed to compel social media companies to do more to protect social media users from online abuse. Most of their players, especially players of color faced horrific racial abuse and attacks on their personal accounts. In this regard they asked for social media companies to have stronger preventative and takedown measures to stop discriminatory abuse being sent and seen, be accountable to the users for their safety and support law enforcement in ensuring consequences for perpetrators.

Stop Hate for Profit

"Stop Hate for Profit" is a US-based organization that exists to pressure social media companies, mainly Facebook, to remove hate speech and misinformation from their websites. The group led a successful boycott of Facebook's advertising services (Facebook's main source of revenue), and this boycott included over 1,000 advertisers, ranging from Honda to Verizon. These boycotts actively changed policies and initiated worldwide discussions. On June 17, 2020, Stop Hate for Profit encouraged businesses to temporarily pause advertising on Facebook's platforms to force Mark Zuckerberg to address the hate speech on Facebook and its influence on the Black Lives Matter movement. As the movement captured nationwide support, Zuckerberg requested a meeting for Stop Hate for Profit coalition leaders in July 2020. This example illustrates how boycotts can successfully be used to pressure companies to be more proactive on combating hate speech on their platforms.

Japanese Protestors Stomp Out Hate

One hundred demonstrators gathered outside of Twitter's Tokyo office and covered the ground with printed hateful tweets. The tweets had all been reported to the company, but it had not removed them. In protest, demonstrators displayed the tweets and people "were encouraged to tread on the printed tweets" before recycling them at the end of the protest.

Featured Resources

- **SNAP: Synergizing Nonviolent Action and Peacebuilding**

  This guide explores how the fields of nonviolent action and peacebuilding can make each other more effective.
  - Introduction Video
  - More on how nonviolent action and peacebuilding can make each other more effective
  - The SNAP Action Guide
- **Explore Nonviolent Action Tactics**
  - Beautiful Trouble Toolbox
  - 198 Methods of Nonviolent Action
  - 198 Nonviolent Methods 2.0

Crowdtangle Link Checker

Introduction

Crowdtangle Link Checker is a Google Chrome extension that searches for the URL you have specified through Facebook, Twitter, Reddit and Instagram to see if any of the accounts, Pages, or profiles in the CrowdTangle database have shared that link. This can be useful to check if and how a certain link has been shared across these social media network sites.

Getting Started

To get started with the CrowdTangle extension, go to the Chrome web store and search for the extension. Kindy note that this only works with Google Chrome as at the time of writing.

Once located, click on the Add to Chrome button to add the extension to your browser. You will now be able to find it on the extensions menu on the top right corner of the browser.
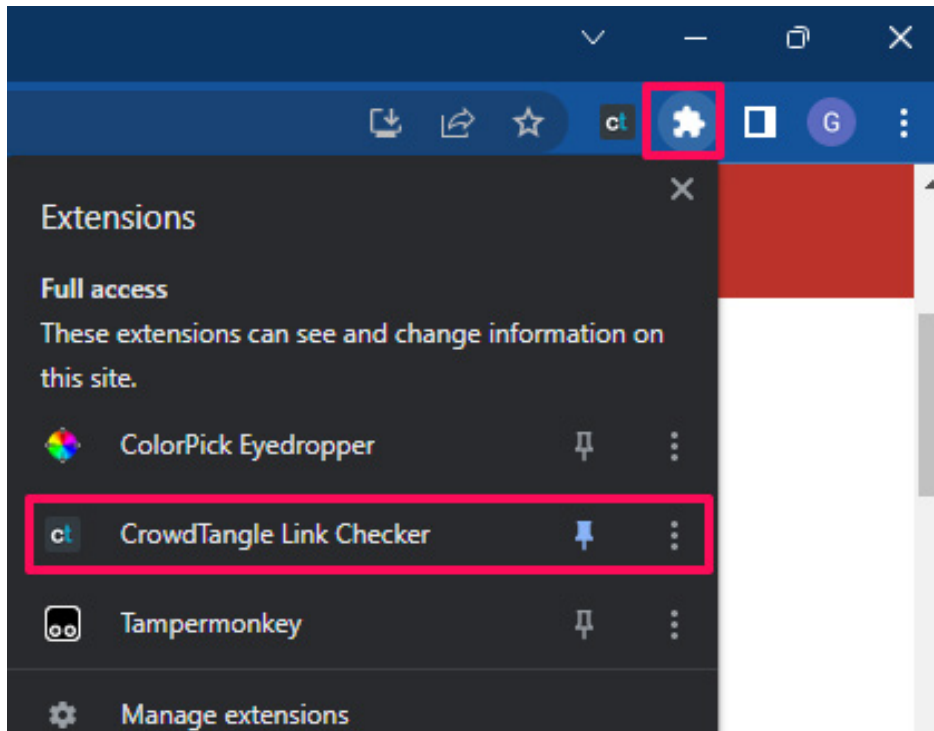
To use the extension, visit any link that you would like more information on and then open the Crowdtangle extension.
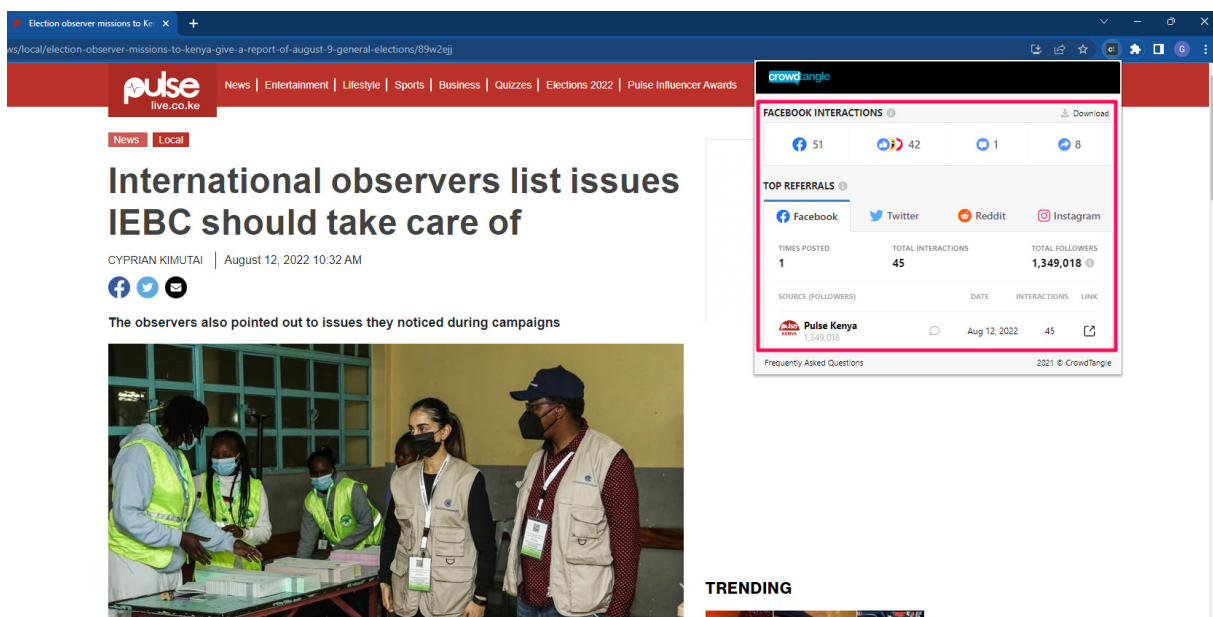
The top section shows Facebook interactions, that is aggregated data from all Facebook posts that reference this URL, including both public and private posts. An interaction consists of reactions, comments, and shares. The bottom section shows referrals, which are individual posts or tweets that mention our URL, and provides a post preview, post date, and total interaction count for that particular post.

🔧 Tech Tool Practice

Find an article that contains disinformation and use Crowdtangle to track the article through social media and find out how far and wide it has been shared.
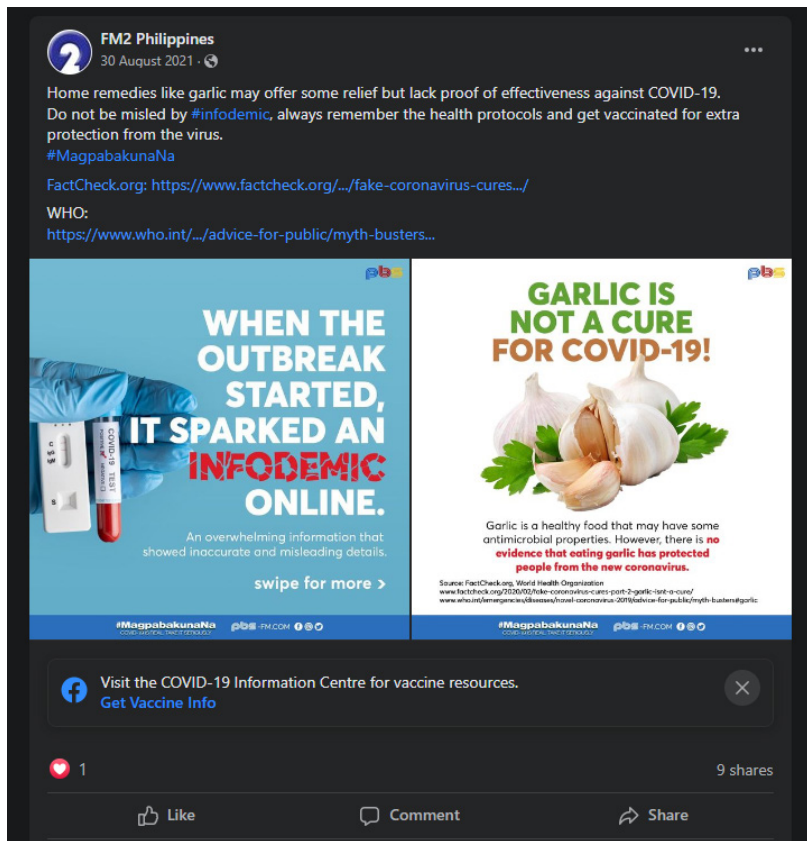
1. Open a link with disinformation, for example: ([https://www.factcheck. org/2020/02/fake-coronavirus-cures-part-2-garlic-isnt-a-cure/](https://www.factcheck.org/2020/02/fake-coronavirus-cures-part-2-garlic-isnt-a-cure/)).

2. Open the Crowdtangle link checker extension and analyze the numbers.

3. From our example article, it has had 1,205 interactions on facebook, most of which are reactions. It has also been posted on Facebook 61 times.

4. Clicking on the link button for each Facebook post takes you to its location on the site.

5. On Twitter, Instagram and Reddit, Crowdtangle couldn't find anything related to our article. This can tell us which social media site is most affected by the disinformation.
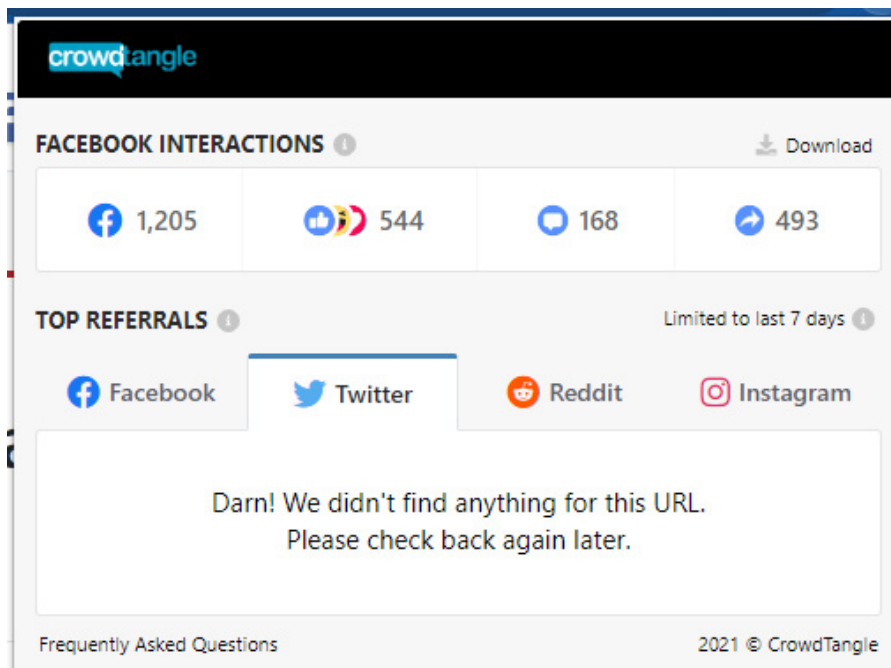
Module 4
Combating Hate Speech: Strategies and Practices – Part 2

Effectively combating hate speech is as complicated as the nature of the problem. In **Module 3**, we explored a first set of strategies, tactics, and practices that organizations and stakeholders all over the world have developed to address this phenomenon and mitigate or prevent its worst consequences. We will now look at further strategies and practices aimed at combating hate speech. These are bigger in scope when it comes to implementation and reach, but they may run longer and have broader impact.

It is important to consider the potential benefits and challenges as well as the practicalities of developing larger-scale campaigns to combat hate. You will need to determine whether offline or online strategies, or in some cases blended strategies (e.g., an offline campaign paired with an online or social media component), are suited to the local context in which the campaign is conducted. In the case of online campaigns this may include factors such as web-hosting costs, technical knowledge and skills, as well as the intended audience's level of access to the selected digital platforms (social media, web, blogs, video streaming, podcasting, etc.). For example, running a Twitter campaign when most of a community uses Telegram will likely have less impact than desired. For offline campaigns, this may include factors such as access to financial and community support, organizing and coordinating capacity, printing materials, as well as the nuances of institutional or governmental resistance to the campaign (threat of arrest, legal threats, etc.). In both offline and online contexts planning campaigns should consider and build-in mechanisms to ensure campaigner safety (see Module 2).

Discussion Questions

- Having gone through the approaches to addressing hate speech (Modules 3 and 4), identify at least two approaches that you think would be effective in your context. Why do you think these approaches could be effective in your context?

- Next, begin basic planning for how you would implement these two approaches by answering these questions:
  1. What would be your goal and who would be your audience?
  2. What would you need?
  3. Who would you work with?
  4. Where would you begin?
  5. How would you go about it?
  6. How would you measure success?

Featured Tech Tools

- Vicinitas is an analytics platform that enables users to fetch Twitter data related to a particular hashtag, keyword or user.

  Here is the Vicinitas trailer video

  Here is the Vicinitas training video

- Trendsmap is a mapping tool with a niche for visualizing Twitter insights around the globe.

  Here is the Trendsmap trailer video

  Here is the Trendsmap training video

Featured Resources

Below is a collection of resources and guides that offer different strategies to combat hate speech:

- No Hate Speech EU Campaign
- Responding to Hate Speech with Positive Measures: A Case Study from 6 EU Countries
- Ten Ways to Fight Hate: A Community Resource Guide (SPLC)

Activity

- For this module's activity, we would like you to continue thinking about the action you discussed in the Module 3 activity, but analyze how you might help your organization convert it to a larger-scale action. Use your reflections from the Module 3 activity to think about how you might adapt that action to one of the strategies in this module.

- Select one of the large-scale strategies from Module 4 to assess how you might adapt your action from the Module 3 activity to a larger-scale strategy.

- Assess the utility of "scaling up" to this new approach: Is it useful to address the problem your organization wants to address? Is it feasible to plan, implement, and manage this type of strategy within your context? Does scaling up help your organization achieve its goals, why or why not?

Content Breakdown

In this module you will find:

Category 5: **Inform or Educate the Wider Public**

- Hate speech literacy and education

- Promotion of media literacy

- Media/public awareness raising campaigns

- Youth counterspeech initiatives and resilience building

- Adoption of a hate speech charter or manifesto

Category 6: **Advocacy**

- Advocacy with community leaders and/or public figure

- Advocacy with social media platforms

- Legal advocacy

Category 7: **Combating Hate by Addressing the Root Causes and Drivers of Conflict**

- Addressing drivers of hate speech

- Addressing conflict via "peace promotion"

- Social cohesion initiatives and promotion of intercultural and inter religious dialogue

Submodule 1: Inform or Educate the Wider Public

Hate Speech Literacy and Education

Hate speech literacy instruction is an important aspect of civic and peace education and should be an integral part of the discourse for combating hate speech and as a preventive measure. With the rise in online hate speech there is a need to address media literacy among internet users to assist them in identifying hate speech and taking care not to engage in it. Some online users may find themselves unknowingly involved with groups or sites that engage in hate speech, and it is important to be able to recognize defamatory speech and how it affects the targeted population. Teaching online users to recognize prejudice and hate speech helps them to   protect themselves and others against hostility and violence, and create safe environments where people can engage freely.

Case Studies

Fighting Hate Speech Through Schools and the Media in the EU

The European Federation of Journalists (EFJ) started the Silence Hate project to "combat and prevent online hate speech against migrants and refugees by developing new and creative counter-narratives." In a span of two years, it held national workshops in schools and trained young journalists. Teachers, educators, and activists learned how to counter hate speech through media literacy and intercultural dialogue. Journalists, audio visual professionals, and social activists participated in a five-day training aimed to bring together a wide network of creative minds to generate content for building a counter-narrative about migration.

Educating Leaders and the Public on Hate speech Through Media

In Côte d'Ivoire, Internews, an international non-profit, teams up with key local partners to create awareness and combat hate speech, especially around election time. Using the information they have gathered from documenting online and offline hate speech, Internews runs three programs to reduce hate speech. One program trains young people – who are heavy users of social media – in order to provide them with tools to combat hate speech. Another program focuses on training local leaders to recognize hate speech, its consequences, and its dangers with the goal of making them more responsible in their positions and public speeches. Finally, the third program focuses on public education through podcasting. As the host of "Carton Rouge", a weekly 10-minute radio and web column, Internews presents and discusses the types of hate speech that are observed locally throughout the week.

Featured Resources

- Three-day Workshop on Combating Hate Speech
- Guide for Practitioners on How to Use Sensitization to Address Hate Speech
- Parent and Educators Guide to Combating Hate Speech

Questions to Examine

- How confident are you in your ability to recognize hate speech in your normal online interactions?
- Could you give examples of words or phrases that would be considered hate speech?

Promotion of Media Literacy

Media literacy is a wider approach than hate speech literacy. It focuses on education around how to sort through information provided by the media in order to better distinguish fact from false information or opinions, and how to identify trusted sources. At the collective level, these critical thinking skills can support civil society resiliency to misinformation and hate speech. Media literacy competency equips people with a variety of skills to access, search, critically evaluate, use, and contribute to information and media content. Media and information-literate citizens are knowledgeable and discerning processors and producers of information which allows them to actively tackle hate speech, and contribute to social inclusion and peace in online and offline spaces. Media literacy instruction should start as early as possible because misinformation is targeted at users of all ages. As an important aspect of civic and peace education, media literacy should be an integral part of the discourse for combating hate speech, and as a preventive measure.

Case Studies

MediaSmarts: Promoting Media and Digital Literacy

MediaSmarts is a Canadian non-profit organization, and is a pioneer in the field of developing public media literacy and digital literacy. They have produced education and awareness programs and resources, through their working partnership with Canadian and international organizations, and through speaking to audiences across Canada and around the world.

Featured Resources

- Checkology

  This interactive online course teaches users how to recognize false or misleading information. Checkology also explains how to differentiate the various types of information people are exposed to on the internet, including information that informs, provokes, entertains, persuades, documents and sells.

- Poynter Fact-Check

  Poynter is dedicated to educating journalists and the public. Their fact-checking tool is searchable and covers a variety of specific claims.

- Africa Check

  Africa Check identifies important public statements, interrogates the best available evidence and publishes fact-checking reports to guide public debate.

Questions to Examine

- Think of all the sources of news you come across in your day. What makes you trust the sources that you do? Or not trust a source?

- Do you check information before you share something online? From all sources? What about information from a friend? Why or why not?

Media Campaigns/Public Awareness Raising Campaigns

Information campaigns raise concerns about hate speech issues in public spaces to spread awareness, foster discussion about what hate speech is, and to inform people about a specific topic or issue. The intention of such campaigns is influencing people's attitudes, behaviors, and beliefs towards the achievement of a defined purpose or goal in respect to combating hate and hate speech. Campaigns can mobilize the power of public opinion in support of an issue and thereby increase pressure on and influence the political will of decision makers. The campaigns can also raise specific issues or highlight the situation of targeted communities.

There are multiple activities that one can use as part of information campaigns to raise public awareness. They include issuing press releases, disseminating reports and publications, convening conferences and workshops, as well as creating and contributing to educational materials. Information may be disseminated via traditional media formats (e.g., radio, television, newspapers, newsletters, leaflets, poster campaigns and arts) or via "new" digital media formats (e.g., social media, websites, blogs, podcasts, and chat groups). Visual tools such as stickers, logos, t-shirts, armbands, bracelets and banners may also be used.

Case Studies

BRICkS - EU Campaign Against Online Hate Speech

Building Respect on the Internet by Combating Hate Speech (BRICkS) was a campaign against online hate speech targeting migrants and minorities in Europe. The campaign included multiple actions, including: an online component focused on a public debate about the big issues of circulation of information in the era of digital world, workshops for youth aimed at identifying and react to hate speech manifestations with web tools, a series of research on the representation of migrants and minorities on online media outlets, participatory meetings and exchanges involving web experts and social media managers, and development of training module and multimedia toolkit as the result of the meetings and exchanges.

A Counter Narrative Campaign in France

To fight the rise of Islamophobia and hate speech against migrants in France, La Cimade, an NGO that advocates for refugees and migrants rights, launched a counter narrative campaign called Festival Migrant'scène. La Cimade has addressed various oppressive narratives through a series of initiatives bundled in the festival since 2009. In its planning, after assessing the oppressive narrative and designing the counter narrative, La Cimade engaged the media as a vehicle to spread the counter narrative messages.

No Hate Speech Video Competition in Spain

Spanish Youth Institute held a No Hate Speech Video Competition, targeting young people to encourage them to defend human rights. Young people aged 15-29 were invited to create and submit a one-minute video that reviews hate narratives or expressions of hate speech and develops alternative narratives of human rights and dignity online. The creators of the 10 most voted-for videos won a prize. This is an example of a campaign that used video as a medium to counter hate speech.

The Blackout Tuesday Campaign Overshadowed the Black Lives Matter Movement in the United States

At one point during the Black Lives Matter (BLM) protests in 2020, Instagram users shared images of a black square to represent social media break and use the time to learn and reflect on Black people's struggles and police brutality in the country. Rather than only tagging the posts with the #BlackoutTuesday hashtag, a lot of users used the #BlackLivesMatter hashtag. This created a backfire effect on the BLM movement, because the #BlackLivesMatter hashtag is where people usually find vital information on the movement, organize protests, and document police violence. As a result, the black squares dominated the #BlackLivesMatter hashtag and it was difficult for people to find the information they needed. So, when conducting similar campaigns, it is essential to express a clear framework for how people should participate, and to understand that when people respond at large scales online, it may shape the resulting campaign in ways that were not intended in the original plan.

#KickItOut: Tackling Discrimination on Social Media

Kick It Out, English Football's Equality and Inclusion Organization, ran a campaign throughout Euro 2016 to raise awareness of social media discrimination and highlight reporting procedures. They produced a guide of learning activities, and used social media to educate fans about the consequences of posting discrimination online.

Featured Resources

- Defusing Hate: A Strategic Communication Guide to Counteract Dangerous Speech

    This detailed guide provides a step-by-step process on how to design a strategic communication campaign to counteract dangerous speech and offers a toolbox with techniques for conducting each phase of campaign design.

Questions to Examine

- Think about an effective campaign that you have seen to raise awareness to counter hate speech. What made it successful? What could have been done differently?
- For your community, what is the most effective approach?

Youth Counterspeech Initiatives and Resilience Building

A newer, but highly promising focus of research and programmatic approaches to countering hate focuses on youth engagement to build media literacy, resilience against hate and radicalization, as well as generating the capacity for direct interventions (also see examples of youth programs in case studies above). This area of focus is essential because teens and young adult audiences are highly engaged online and therefore subject to regular experiences of hate speech, cyberbullying, and recruitment/radicalization into hate ideology. In some cases hate groups specifically target youth for radicalization and participation in hateful online activities. So, a focus on youth education and participation is increasingly being recognized as a necessary component of campaigns for change.

Case Studies

The Game Changer Project – Engaging Youth to Counter Hate

One program from Europe, The Game Changer program, has created a "peer to peer," youth-engaged program which works through NGOs to train selected youth. The program started in Greece, Poland, and France and is now running in eight European countries. The youth trained by the program, called Ambassadors of Change, select topics for intervention, create social media campaigns, and also act as community ambassadors in offline contexts. For example, the program in France partners with the NGO "L'association Artemis to run campaigns aimed at encouraging young people in Paris from different social and religious backgrounds to interact with and understand each other. As part of the Game Changer project, Artemis is developing campaigns to help young people to identify and counter hate speech and explore sensitive issues from a variety of perspectives." Along with the Ambassadors of Change program, the Game Changer project also contains an element that involves the use of both Offline Social City Games and Online Social RPGs (role player games), which promote tolerance, civic engagement, diversity and inclusion through gamification. The Game Changer project has a tools library that includes access to their various trainings, NGO "Camp" programs, and the "games" elements.

PERIL: Addressing Youth Radicalization and Extremism to Build Resilience

This is another youth-oriented program that focuses on addressing hate speech, risk, and media literacy through capacity and resilience building. This program is a joint effort of the Polarization and Extremism Research Innovation Lab (PERIL) and the Southern Poverty Law Center (SPLC). Together these groups created a Parent and Caregivers' Toolkit for middle school-aged youth (11-13 years of age) to work against the increasing threat of youth online radicalization. The toolkit comprises a guide that offers parents and caregivers strategies and tips to recognize the warning signs of youth radicalization as well as new risks in the COVID-19 era. Its aim is also to help parents, caregivers, and teachers understand the drivers and grievances that create susceptibility to extremist rhetoric which allows them to discuss these problems with the young people they care for and, if necessary, intervene more effectively. The guide was launched in summer 2020 with a series of webinars including presentations from academics and practitioners in a variety of topical areas related to the guide along with post-presentation discussion sessions for the parents, caregivers, and teachers in attendance.

Featured Resources

- Game Changer: Build a Campaign (tool box)
- Building Resilience and Confronting Risk (PERIL/SPLC Guide)

Questions to Examine

- What issues might youth in your context be able to identify that others cannot? How might you successfully engage interested and concerned youth in your campaigns?

- What factors might you need to take into consideration or what extra precautions would you take when working with young people (e.g., media literacy, ground work on hate / hate speech, parental consent, emotional and wellbeing support)?

Adoption of a Hate Speech Charter

Journalists and other public figures adopt anti-hate speech charters in which they publicly recognize the problem of hate speech in political or other public debates, and vow to act against it. These charters are like a 'code of conduct' against hate speech and serve to raise awareness and create momentum around a hate speech campaign.

Case Studies

Women's Peace Conference in Cameroon to address the Anglophone Crisis

In July 2021 over 1,500 women from all regions of Cameroon came together to call on the parties to the Anglophone conflict to find peaceful solutions in the first ever Women's Peace Conference in Cameroon. The conference highlighted the facts that women and children are primary victims of armed violence and that women are also often primary caregivers to those harmed in armed conflict. At the end of the conference, the delegates jointly adopted a "Women's Appeal for Peace", which was handed over to the Cameroonian government in the presence of many media representatives.

Media Workers in Central and West Africa Call on the Media to Address Hate Speech and Reaffirm Ethical Standards

In two regional meetings in Central Africa and West Africa, media workers from various African nations urged the media to take steps to address hate speech and to hold higher ethical standards in journalism. The outcome of the meetings included updated ethical codes across the region to adequately deal with the

growing threats of hate speech, terrorism, and violent extremism. Participants vowed to work with media groups to create election reporting guidelines and hate speech glossaries ahead of the 2018 elections. They also agreed to put forward a program focused on improving media ethics, governance, self-regulation, and reporting of terrorism and violent extremism, as well as developing holistic media literacy programs.

"Manifesto of Venice" Against Gender-Based Violence in Italy

"Manifesto of Venice" is a declaration signed by 800 journalists and launched by the Italian journalists federation. The declaration is aimed at improving gender equality and information accuracy in reporting. The manifesto includes points regarding the appropriate language to use when describing gender-based crimes committed against women and the importance of avoiding images or symbols which are stereotypical representations of women or that reduce women to "objects of desire."

Featured Resources

- Rights for Peace, Hate Speech in International Law

Questions to Examine

- Do anti-hate speech charters help counter hate speech?
- Are there examples of anti-hate speech charters you are aware of in your community? If not, would one be useful? How might one be developed?

Submodule 2: Advocacy

Advocacy with Community Leaders and/or Public Figures

Advocacy is an umbrella term that describes collective and strategic engagement with certain stakeholders to get their attention, support, or move them to certain actions. When it comes to hate speech, and depending on the context, advocacy can aim to raise awareness of the existence of a problem, or to get decision makers or key actors to take certain commitments or actions. Involving community leaders and public figures in advocacy can give campaigns a powerful boost. Gaining the commitment of decision-makers for a campaign to counter hate speech may increase the likelihood that the public will also counter hate speech in their communities. Decision-makers can influence the public's response to hate speech by openly supporting the campaigns, emphasizing countering hate speech, or by addressing communities directly.

Case Studies

Eradicate Hate Global Summit

The Eradicate Hate Global Summit was born out of the tragic hate-based antisemitic mass murder at Pittsburgh's Tree of Life Synagogue in October 2018. The organizers and Tree of Life community refused to merely be defined as victims of hate. They used the attack as motivation and allocated funds sent by the public (to support the community in the wake of the massacre) to launch the most significant anti-hate, rule of law initiative to date. The initial Summit brought together a variety of multi-disciplinary experts and leaders committed to the global eradication of all forms of hate speech and violent extremism. Importantly, the focus of the Summit is to move beyond dialogue to focus on the collective pursuit of solutions by uniting experts and leaders from

around the globe. Using an ongoing working group structure, the organizers hope to develop practical solutions and measurable change.

Mapping Hate Groups Across the United States

Southern Poverty Law Center (SPLC) is a legal advocacy organization that defends racial justice and ensures the promise of the civil rights movement becomes a reality for all based in the Southern part of the United States. Every year since 1990, the SPLC has conducted and published a census on hate groups within the country. The results of the census are considered as an indicator of the level of hate activity in the country. The hate map visualizes identified hate groups' approximate locations gathered from the census and the result of the activity of the hate groups monitored in the previous year. The map has been used for various purposes, including advocacy efforts to push lawmakers to take action about hate in communities. It is a great advocacy tool because it makes the problem and its scale more visible and illustrates the gravity of it.

Mapping Intolerance Utilizing Social Media Conversations in Italy

The "map of intolerance" is a joint project between VOX–Osservatorio Sui Diritti, an Italian NGO, and three universities, focused on mapping out hate speech messages that targeted women, people with disabilities, LGBTQI people, and other minority groups on Twitter. Using the geolocation of the tweets, the team determined the areas where tweets came from and created heat maps that corresponded with the level of intolerance in those areas.

Featured Resources

- 10 Ways to Fight Hate (SPLC Community Advocacy)

Questions to Examine

- What types of advocacy are happening in your community? How could you connect?
- Can you think of a community leader who is frequently involved; what is it that they do? How do they connect the community?

Advocacy with Social Media Platforms

Some civil society organizations work with social media companies to 1) verify content, 2) flag false information (e.g. COVID, election rumors), and 3) redirect the user to more trustworthy sources or psycho-social services. When it comes to hate speech, they can coordinate to request immediate take down of worst cases of hate speech, draw attention to specific cases or developments in national context that will impact the level of hate speech targeting certain groups or persons or ask for further clarification of platforms monitoring rules and decisions.

- **Content removal**

    Taking down hateful content from the internet is sometimes necessary to ensure internet users are not exposed to harmful rhetoric. Both governments and technology companies often have specific policies that govern hate speech, but institutional (government or company) reviews of posted content can take a long time and may not yield the desired results. In such cases, civil society actors (individuals or groups) may take up an advocacy role, seeking stronger policies or putting pressure on companies and internet sites to take down offensive content.

Case Studies

Taking Down Fake Content on Social Media

One such initiative is being driven by Odipo Dev, a Kenyan digital strategist startup which says there's been growing concern from their clients about how to respond to fake news. Before the last elections, the company tracked false news consumption and employed a robot to analyze the presidential debate. The organization says that interested parties usually want them to find fake domains or social media handles, censor out fake items from feeds through browser extensions, or warn followers not to use certain pages that contain false content by using targeted ads or bots (automating the warnings).

Online Campaign to Take Down Hate Speech on the Internet

A Canadian collective called the The Canadian Anti-Racism Education and Research Society runs online campaigns organizing internet users to flag particular sites to take down hateful content or disassociate themselves from sites that feature or do not take down hateful content. One campaign focused on Youtube, sought to force the company to implement hateful word and username filtering, as well as IP blocking.

- **Draw attention to specific cases/issues**

    Civil society organizations use social media to create awareness of hate speech against targeted populations, often those specifically using the same platforms. Most people may not have access to the hateful content and may not know what the target group is going through, and this serves to expose the hatemongers and ask for support in combating the highlighted hate speech.

Case Studies

Bringing the World's Attention to the Hate Speech Against the Rohingya and Muslims in Myanmar

The Human Rights Center at the University of California Berkeley School of Law and Reuters collected and analyzed over 1,000 hate speech posts on Facebook that contained anti-Rohingya and anti-Muslim messages, most of them in Burmese. The collected and analyzed content included posts, comments, images, and videos attacking the group verbally as well as in an imagery way. As the outcome of the analysis, the two entities published the investigative write up to draw attention to this specific hate speech issue going on the platform against the Rohingya and Muslims in Myanmar.

Ethiopian Network Against Hate Speech

The Network Against Hate Speech, a volunteer group tracking hate speech in Ethiopia, tracked posts on Facebook that incited violence and genocidal attacks against the Amhara people. The group then called on Facebook to take actions on the moderation of its hate speech and incitement content. Facebook's Community Standards are not available in Ethiopia's two main languages and the company has no full-time employees in the country. Facebook is instead relying on activists and a network of grassroots volunteers to flag content and keep them abreast about what's happening on the ground. Activist groups joined forces and wrote an open letter to Facebook to take actions to save Ethiopians from violence-inciting speech.

- **Clarify monitoring rules and decisions to the wider public**

   The growing problem of social media companies wielding so much power and control over information consumed by the public has seen civil society organizations call for regulation by both the government

and the companies involved. The policies governing misinformation and hate speech on social media are dynamic and the public may not be aware of the constant changes and updates, so actors across civil society organizations should take it upon themselves to advocate for the public on relevant policies.

Case Studies

Civil Society Urges Governance of Social Media Leading up to Myanmar's 2020 Election

Prior to the 2020 Myanmar general election, civil society organizations like Koe Koe Tech urged social companies like Facebook to work with representatives from civil society organizations, multilateral organizations, and researchers. The goal of such multi-stakeholder oversight is to ensure independence, collaboration, and accountability in the regulation and enforcement of speech standards, algorithms, human reviewers, privacy practices, and internal policy processes of social media platforms. As a follow up, Facebook published an article to clarify and reiterated its monitoring rules and misinformation policy. Some of the rules and policy and examples of the decisions that it pointed out include those pertaining to the removal of verifiable misinformation and unverifiable rumors; its policy against hate speech and the removal of such content; their third-party fact-checking program; and finding and stopping coordinated campaigns that seek to manipulate public debate across the platform.

Featured Resources

- "Social Media and the Activist Toolkit" by Youmans and York

Questions to Examine

- Have you seen cases of hate speech being highlighted on social media to create awareness? Have you ever used social media to advocate for hate speech victims?
- What are your thoughts on government regulation of social media companies vs self-regulation?

Legal Advocacy

This strategy focuses on creating or using legal or regulatory leverage to advance the cause against hate speech. It may consist of engaging with (or pressuring) the national government to adopt a hate speech law (if there isn't one), improving existing law, improving the implementation of that law, improving regulations, and strengthening a company's own community guidelines. Legal advocacy may also use the judicial system to advance particular goals for change.

Legislation to combat hate speech and incitement has its limitations. Hate speech legislation in some national and regional contexts may be viewed as an efficient and appropriate means to prevent harm emanating from hate speech and/or dangerous speech. It places an obligation on governments to prohibit different forms of incitement. This can lead to two problematic issues. First, laws can be vague with limited use in terms of practical protections and enforcement capability. Second, government regulation of speech can also be weaponized against political outgroups and minoritized populations in highly polarized political environments.

- **For the adoption of national hate speech legislation**

Case Studies

Organizations Persuade the Malaysian Government to Implement an Anti-Hate Speech Regulation

Currently Malaysia guarantees freedom of speech, but this freedom is very loosely regulated. Hatred and threats against Rohingya refugees have proliferated since the beginning of the pandemic. The refugees are blamed for bringing the virus into the country. Due to the exacerbating situation, numerous organizations called on the Malaysian government for addressing violent threats and hate speech against the refugee community by passing a comprehensive equality/anti-discrimination legislation that is aligned with international human rights standards and making restrictions on free speech to protect the rights of the refugees lawful.

Rwanda, Sectarianism, Divisionism, and Genocide Ideology

Rwanda signed and ratified the ICCPR, the Genocide Convention and the ICERD and is thus obliged to implement the international criteria prescribed by those conventions. The Rwandan government introduced a series of laws of which 'Discrimination and Sectarianism (2001) and 'Genocide Ideology' (2008) earned most criticism. Amnesty International describes the law as vague and ambiguous, and therefore prosecution has become a matter of interpretation. This danger of these laws can be seen in the case of two female journalists, Agnes Nkusi Uwimana and Saidati Mukakibibi of the private bi-monthly newspaper Umurabyo who were sentenced to 17 and seven years of detention respectively in 2011. The journalists had covered sensitive policy issues in several opinion pieces from mid-2010 onward. Based on this work, they were charged with a mix of alleged crimes including incitement to violence, promoting ethnic division, Genocide denial, and defamation.

Surveillance Laws Propagated by Hate Speech Laws

The damage caused by hate speech legislation includes more than just wrongful convictions. Hate speech legislation can also be misused to legitimize state surveillance activities. In Kenya, the Communications Commission of Kenya (CCK) drafted guidelines that put the responsibility of filtering out inflammatory text messages on mobile phone service providers which meant the providers would prospectively be held accountable for hate messages. In 2012, they further announced plans to install Network Monitoring Software for internet traffic. By 2013, the government had implemented an "observation regime" with about 100 monitors hired to watch social media content. The rapid set-up of those surveillance mechanisms raised concerns about the invasion of personal privacy, especially the extent to which the state should be allowed to conduct such surveillance in a country with no data protection law.

- **For the adoption of BETTER hate speech legislation**

Case Studies

Myanmar: Government's Approach to 'Hate Speech' Fundamentally Flawed

Article19, a human rights organization that focuses on the defense and promotion of freedom of expression and freedom of information, released a policy brief suggesting the implementation of anti-hate speech law under consideration in Myanmar "should avoid criminalizing expression as a primary means of combating hate speech." According to Article19, a previously drafted version of the law included vague language that potentially could infringe on the freedom of expression. The law was finally passed in 2020, but it only encourages civil servants to denounce and prevent hate speech without any specific consequences noted on the law. Civil society organizations pointed out that the lack of any specific regulations only means that it is unlikely to have any impact.

Forming a Working Group to Lobby for an Amendment in the Hate Speech-Related Criminal Code in Georgia

Multiethnic Georgia (Mnogonatsionalnaya Gruzia) and a number of other NGOs set up a working group to lobby for an amendment of the existing hate speech-related Criminal Code in Georgia. Their goal was to improve state representatives' accountability for acts of hate speech, and to raise awareness about the initiative through an informational campaign. This initiative served a dual strategy to prepare the amendment with the aim of getting it adopted by the Parliament and to encourage public participation in the process.

- **Better implementation of EXISTING hate speech laws or community guidelines**

Case Studies

Facebook Faces a Lawsuit Over Hate Speech and Disinformation in France

Reporters Without Borders, a press-freedom organization known as RSF, filed a complaint in March 2021 against Facebook in France due to "the large-scale proliferation of hate speech and false information on its networks." RSF argues that this issue is caused by Facebook's negligence to effectively implement its own community standards which promise a "safe" and "error-free" online environment. RSF chose to file the lawsuit in France because of its law that is suitable to scrutinize Facebook on this issue. The organization is currently considering opening similar complaints in other countries as well, as Facebook's terms of service are the

same worldwide. The goal of this initial lawsuit is to open a preliminary probe, which hopefully will make a global impact after a court ruling in France set a precedent.

💡 Featured Resources

- The Inherent Danger of Hate Speech Legislation
- Elements of a Policy Advocacy Campaign: Guidance for Civil Society Organization
- Developing Effective Advocacy Campaigns

❓ Questions to Examine

- Reflect on potential advantages and challenges of hate speech legislation?
- What are the unintended consequences of hate speech legislation?
- What are the challenges of existing hate speech legislation in your country?

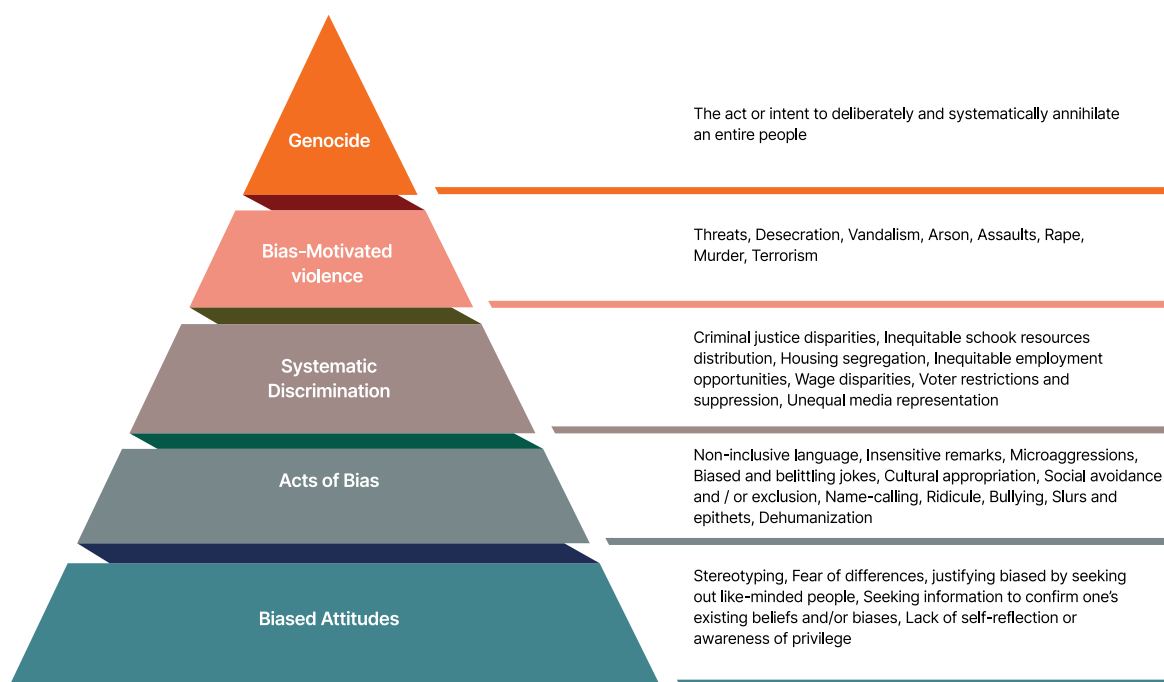# Submodule 3: Combating Hate Speech by Addressing Its Root Causes and Drivers

A broader approach to combating hate speech is driven by the insight that hate speech often has deep underlying social, cultural, and/or political causes. Especially in contexts experiencing conflict or political polarization, addressing these drivers of hate, conflict, and societal discord can be a more thorough and long lasting approach. Hate speech does not occur in a vacuum, and is often a symptom of deeper and larger social problems.

Addressing Drivers of Hate Speech

Stopping hate in its earliest stages, or preventing it from developing altogether, is essential to avert the worst hate-based harms including violence and genocide. Addressing bias, stigma, and the "otherizing" of groups and minorities is essential to early disruption. Moreover, building community cohesion and resilience against hate and violence is rooted in prevention work such as efforts aimed at building and reinforcing positive cross-cultural and cross-group engagement and the promotion of peace.

The Pyramid of Hate is an infographic created by the Anti-Defamation League (ADL). The Pyramid of Hate illustrates the prevalence of bias, hate and oppression in our society. It is organized in escalating levels of attitudes and behavior that grow in complexity from bottom to top. Like a pyramid, the upper levels are supported by the lower levels; unlike a pyramid, the levels are not built consecutively or to demonstrate a ranking of each level. Bias at each level reflects a system of oppression that negatively impacts individuals, institutions and society. Unchecked bias can become "normalized" and contribute to a pattern of accepting discrimination, violence and injustice in society. So, if a society treats lower-level behaviors as acceptable or "normal", it results in the behaviors at the next level becoming

Level	Description
Genocide	The act or intent to deliberately and systematically annihilate an entire people
Bias-Motivated violence	Threats, Desecration, Vandalism, Arson, Assaults, Rape, Murder, Terrorism
Systematic Discrimination	Criminal justice disparities, Inequitable schook resources distribution, Housing segregation, Inequitable employment opportunities, Wage disparities, Voter restrictions and suppression, Unequal media representation
Acts of Bias	Non-inclusive language, Insensitive remarks, Microaggressions, Biased and belittling jokes, Cultural appropriation, Social avoidance and / or exclusion, Name-calling, Ridicule, Bullying, Slurs and epithets, Dehumanization
Biased Attitudes	Stereotyping, Fear of differences, justifying biased by seeking out like-minded people, Seeking information to confirm one's existing beliefs and/or biases, Lack of self-reflection or awareness of privilege

Pyramid of Hate © 2021 Anti-Defamation League

more acceptable. Ultimately, normalization of hate behaviors working up the pyramid can lead to violence if not addressed.

Case Studies

Addressing and Challenging the Stigmatization and Prejudiced Stereotyping as Root Causes for Social Exclusion of Roma

The European Roma Grassroots Organization Network (ERGO) initiated the 'Typical Roma?' campaign that addressed stigmatization and stereotypes against the Roma that are the root causes of social exclusion of the Roma group in Europe. The campaign aimed to raise a positive image of the Romas and encourage active citizenship that constructs an "ALL in ONE" society. The campaign—held in Macedonia, Albania, Bulgaria, Moldova, and Romania—consisted of a series of activities, including the "Which community has the most Roma-friendly mayor" competition, the set up of "Inclusion Zones" where Romas and non-Romas can interact and learn about each other, and a summit in which 40 Roma youth will undergo a training.

NoRa (No Racism) Campaign: Public Discourse and Antiracism Training

In Finland, youth trainers involved in online youth work were asked about how to deal with online racism in discussion threads targeting Finnish Somali or Finnish Roma. In response to this issue, a number of organizations developed a project that investigated and developed strategies to counter online hate speech. The outcome of this research was a project to advocate for antiracism and to provide training for online youth workers about hateful terms. The training emphasized providing a chance for people to reflect on their own attitudes and how to respond to these hateful terms should they come up online. The overall goal of this activity was to support online youth workers to develop anti racism strategies.

#MyFriend Campaign

In Myanmar, ethnic and religious division has been a main driver of intergroup violence. The #MyFriend campaign was started to raise awareness about the strength of diversity by promoting love and friendship between diverse people in Myanmar. The campaign encouraged people to post pictures of themselves with friends of different ethnicities and religions using the hashtags #myfriend and #friendshiphasnoboundaries with the goal of reducing "all forms of discrimination, hatred, hate speech, and extreme racism based on religion, ethnicity, nationality, colour and gender" in Myanmar.

Addressing Conflict via "Peace Promotion"

Peace promotion involves strengthening peace and non-violence through education, advocacy and media, amongst other activities. Digital campaigns and activities can be incorporated into already existing works to promote peace which includes on and offline activities.

Case Studies

Friend Me 4 Peace: Creating Opportunities for People Who Support Peace to Connect and Break Stereotypes

The objective of the "Friend Me 4 Peace" project was to promote peace between social media users who are on "different sides" of multinational conflicts in the Middle East. Using Facebook, Peace Factory matched social media users to become online friends. The assumption was that once connected, they could see each other's posts and would realize that they have more things in common as human beings than differences. Some participants did become real friends. The connections made through the "Friend Me 4 Peace" project improved mutual understanding among participants and promoted peace.

This policy brief encourages peacebuilders to first seek to understand various narratives around the hate speech issue that they want to work on. They can do this by building their narrative competency first, so that they can conduct narrative analysis. Understanding narrative is imperative in the process of peace promotion, because you cannot promote peace without engaging with different narratives to co-construct new narratives for peace with others.

Social Cohesion Initiatives and Promotion of Intercultural and Interreligious Dialogue

Social cohesion allows us to understand the interactions between populations. It creates a space for shared advocacy by people within a common process of development. It also enables a better framework for responses to the issues involved in intercultural and interreligious dialogues. Social cohesion is based on individuals' abilities to interact with others to the benefit of society as a whole.

An activity used to foster social cohesion is the Social Identity Wheel. The wheel activity encourages participants to identify and reflect on the various ways they identify socially, how those identities become visible or more keenly felt at different times, and how those identities impact the ways others perceive or treat them.

The worksheet prompts participants to fill in various social identities (such as race, gender, sex, ability disability, sexual orientation, etc.) and further categorize those identities based on which matter most in their selfperception and which matter most in others' perception of them. The wheels can be used as a prompt for small or large group discussion or reflective writing on identity.

The chart features a circle that is separated into 11 sections. Each section is labeled: (starting at the top and moving clockwise around the circle) ethnicity; socioeconomic status; gender; sex; sexual orientation; national origin; first language; physical, emotional, developmental (dis)ability; age; religious or spiritual affiliation; race.

In the center of the circle, there are five numbered prompts:

1. Identities you think about most often
2. Identities you think about least often
3. Your own identities you would like to know more about

Social Identity Wheel
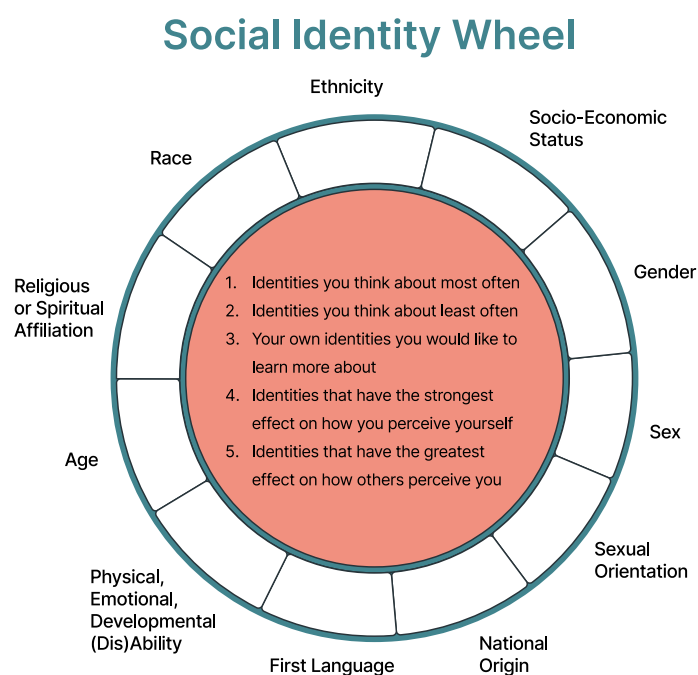
"Social Identity Wheel". *Inclusive Teaching at U-M,* https://sites.lsa.umich.edu/inclusive-teaching/wp-content/uploads/sites/853/2017/04/Screen-Shot-2017-04-20-at-10.29.12-AM.png. Accessed 13 January 2023.

4. Identities that have the strongest effect on how you perceive yourself

5. Identities that have the greatest effect on how others perceive you.

Goals of the exercise:

1. To encourage participants to consider their identities critically and how identities are more or less keenly felt in different social contexts. Different religious beliefs or lack thereof can be highlighted as a context as a way to approach questions on barriers to inclusion.

2. To sensitize participants to their shared identities with their peers as well as the diversity of identities in the community while building and encouraging empathy.

Case Studies

Peacebuilding After the Genocide

On the 20th anniversary of the Rwanda Genocide, the country ran a 'Peace-building after Genocide' initiative which was a mobile exhibition showcased around the country that used storytelling and dialogue methodologies to educate people about the 1994 genocide. It examined what causes violence and sent messages of peace and social cohesion. In the stories, messages that contributed to social cohesion and peace-making were told, along with those that explored the causes of genocide and mass violence.

Using Textbooks to Spread Information About Both Sides of the Conflicts

As a peacebuilding effort between Israel and Palestine, authors worked with Israeli and Palestinian school teachers to co-create a school history textbook. This was done to weave together two competing narratives (during the period of conflict, each side developed their own narratives and held on to it as the absolute

truth) into a common narrative. The co-created narrative was included in a textbook to help students learn to understand and respect the history of Israel and Palestine from their own, and each others' perspectives. This is a narrative approach that can be applied in conflicts centered around history and memory.

UN Adopts Resolution to Combat Hate Speech with Dialogue

The United Nations General Assembly adopted a resolution calling for the use of interreligious and intercultural dialogue to combat hate speech. Adopted in July 2019, the resolution is part of the United Nations' multilateral effort to combat rising xenophobia and racism. It acknowledges the vital importance of interreligious and intercultural dialogue as a means to foster tolerance, as well as the need to partner with religious leaders to promote diversity and human rights. The UN Fez Plan of Action is the first document to focus on the role of religious actors in preventing incitement to violence and the first to develop specific regional strategies with this objective.

Featured Resources

- Teaching Tolerance Countering Online Hate Speech
- Building Resisliance to Genocide through Peace Education
- Ten Ways to Fight Hate: Teach Acceptance
- Countering hate speech through arts and arts education – Addressing intersections and policy implications

Questions to Examine

- In your context, what strategies are used for peace promotion, both online and offline?
- How do you connect to any existing online and offline campaigns?

Technology Tools Corner

Vicinitas

Introduction

Vicinitas is an analytics platform that enables users to fetch Twitter data related to a particular hashtag, keyword or user. On the free tier you can download tweets posted within the last 7 to 10 days that are related to any topic you may be researching. With the paid tier, this gets bumped up to all the tweets available in the last couple of years, as well as real time tweets and insights. However, Vicinitas does offer discounts for students and non-profit organizations.

Getting Started

To get started with the free tools on Vicinitas, visit https://www.vicinitas.io/ and scroll down to the section named **Download and export into Excel for free**.

You can then input the hashtag or keyword of interest and click on the **Search** button to start the retrieval of tweets. This may take a few minutes. Kindly note that you will be prompted to connect to your Twitter account, so having access to an account on Twitter is required for working with this tool.

Once it is done searching, you will be presented with the option to download an Excel file with the tweet data.

Right below that, under the **Analytics** tab, there is a section that has some statistics on the keyword you are searching for. This section contains a word cloud of the most common phrases from the tweets that were fetched, aggregations of user posts and engagements, timeline graphs as well as pie charts showing various tweet classifications.
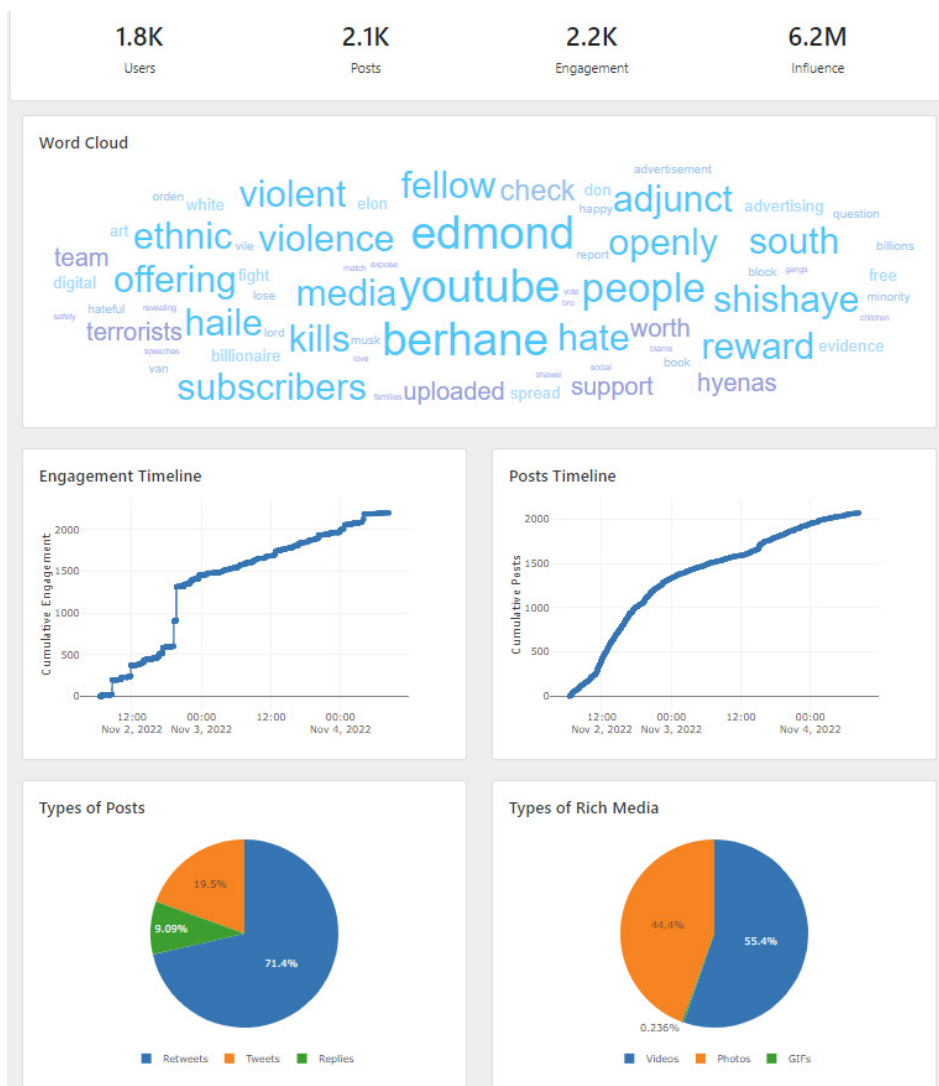
You can also perform this search for user tweets as well as user followers to get the number and content of tweets or number of followers from a particular Twitter handle.

Download User Tweets for Free

Export into Excel for free up to 3,200 recent tweets of any Twitter user

○ Hashtag / Keyword Tweets   ● User Tweets   ○ User Followers

@nohate_speech   Search

Total number of tweets downloaded: 3155

Export To Excel

Tech Tool Practice

For this tool, we will practice by finding all tweets related to the hashtag **#Antisemitism** which has been trending lately. To do this, go to the Vicinitas home page (https://vicinitas.io/) and type in the hashtag under on the search box. This search criteria should be under the *Hashtag/Keyword Tweets* option.

VICINITAS   Real-time Tweets   Historical Tweets   Pricing   Free Tools ▾   Sign In

Understand how Twitter users engage with your content

Vicinitas helps track and analyze real-time and historical tweets of your social media campaigns and brands on Twitter.

Sign up for real-time tracking   Quote for historical tweets

Not ready to commit yet? Try our free tools.

Download and export into Excel for free

● Hashtag / Keyword Tweets   ○ User Tweets   ○ User Followers

#Antisemitism   Search

No credit card required. All you need is a Twitter account.

Upon clicking on the *Search* button, Vicinitas should start the search. It should take a few minutes to complete. Once complete, be sure to download the Excel file to view the tweets in detail. You can also browse these tweets by navigating to the *Browse* tab.

Under the Analytics tab, be sure to check out some statistics of tweets with the #Antisemitism hashtag.

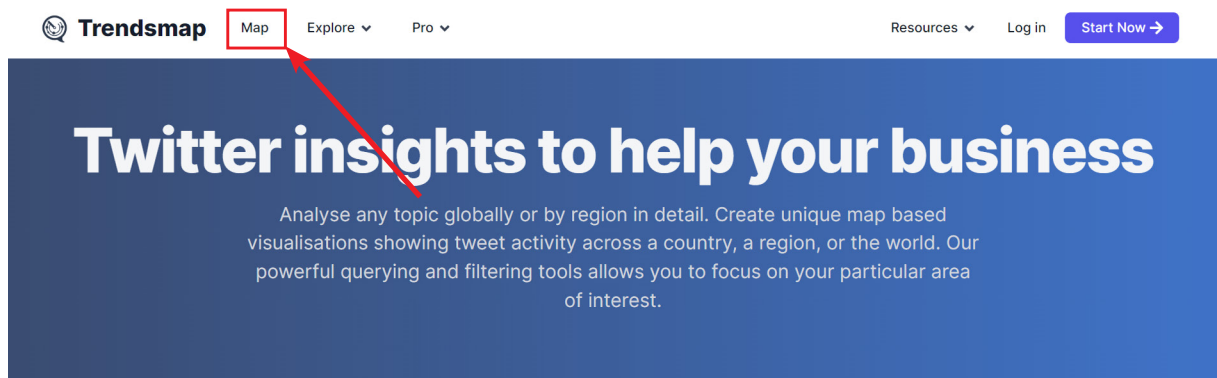Now you are ready to do your reporting on the #Antisemitism hashtag.

Trendsmap

Introduction

Trendsmap, at its very core, is a mapping tool with a niche for visualizing Twitter insights around the globe. You can use Trendsmap to analyze any topic, globally or by region, in detail. The tool also allows you to create a unique map based on visualizations showing tweet activity across a country, region or globally. Unfortunately, a lot of Trendsmap's features are placed behind a paywall so we will focus on what we can do on the free tier.

Getting Started

To get started with the free tools on Trendsmap, visit https://www.trendsmap.com/ and click on the **Map** tab at the top of the page. You will need to sign up for an account first.

Trendsmap   Map   Explore ⌄   Pro ⌄   Resources ⌄   Log in   Start Now →

Twitter insights to help your business

Analyse any topic globally or by region in detail. Create unique map based visualisations showing tweet activity across a country, a region, or the world. Our powerful querying and filtering tools allows you to focus on your particular area of interest.

Unique and Powerful Tools

Our tools allow you to analyse Twitter activity from today, through to months ago.

Analytics
Get unique and detailed insights into any topic on Twitter using our powerful query capabilities. Precisely filter the data or go global. Download data to complement your research in other tools.

Visualizations
Create stunning visualisations to display data in a range of unique and engaging ways. Highlight the spread of a topic across the country, or across the globe. Show how multiple topics unfold over time.
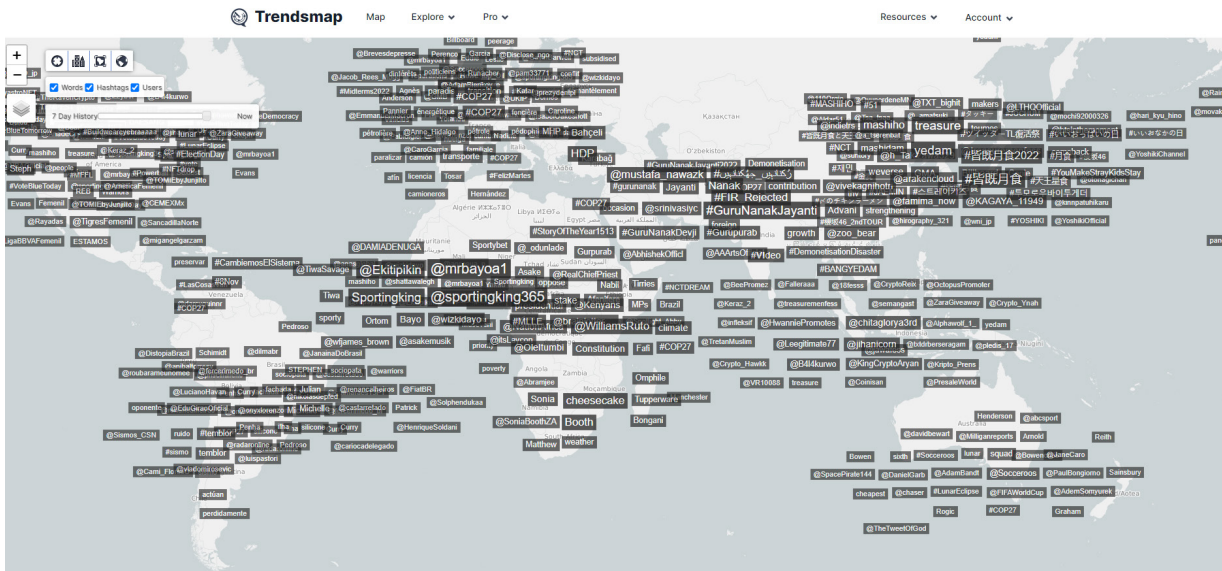
Alerts
Stay up to date with proactive email and Slack based alerts for any topic so you know when something happens on Twitter.

Explore Trends
The original 'Trends Map'. Explore the trends happening now using our interactive map view in any part of the world

On this page, you can find a map of the globe showing hashtags, keywords and users that are currently trending.

Zoom in to a particular area on the map to get more information about that area. You can also click on any particular keyword or hashtag to learn more about what is trending.
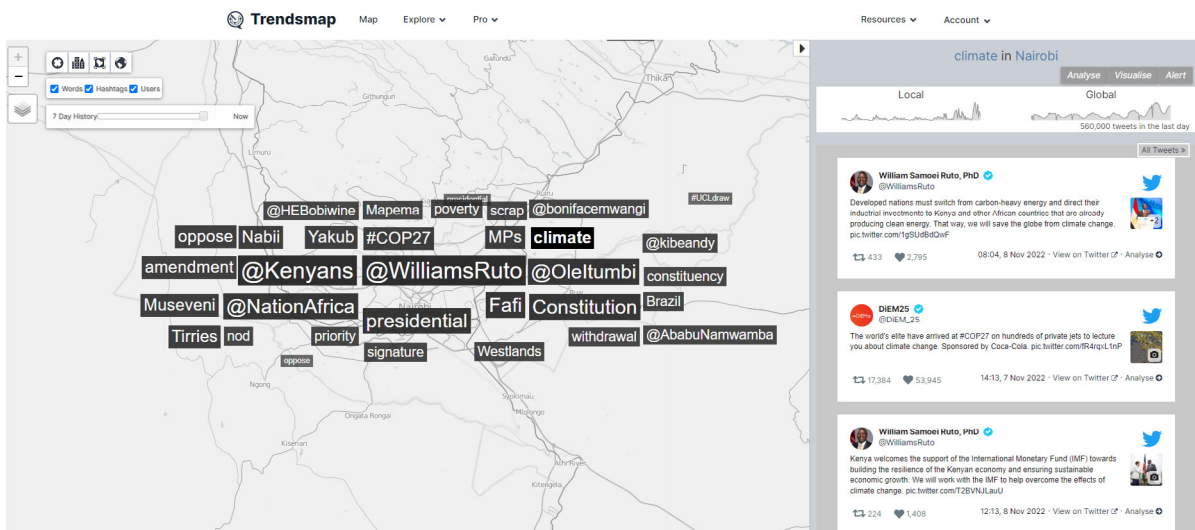
Under the **Explore** tab, you will find another button to take you to the map section, as well as top tweets, top users, and tweets in your local area.
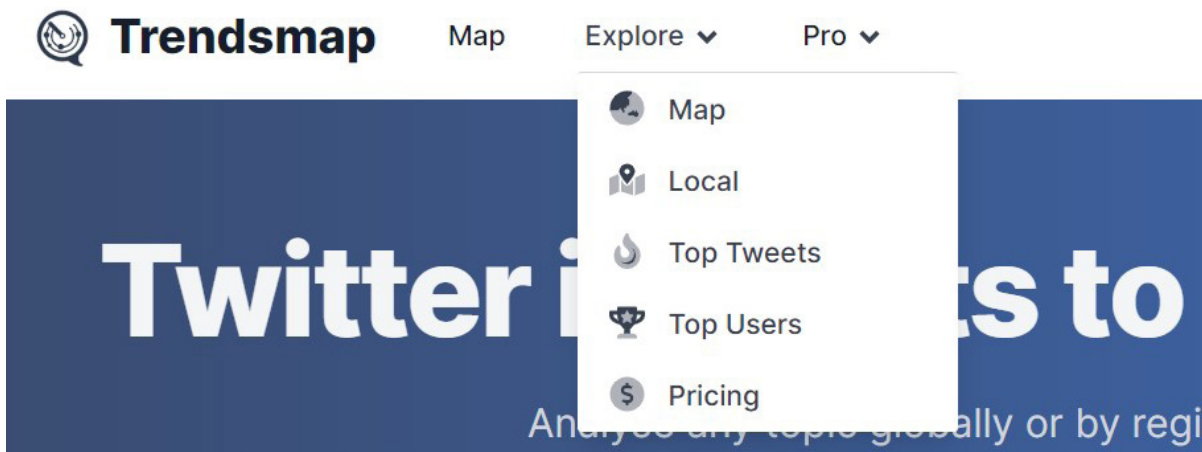
The *Pro* tab is where you will want to go to perform searches for specific topics of interest, particularly with the *Analytics* and *Topic Search* buttons.
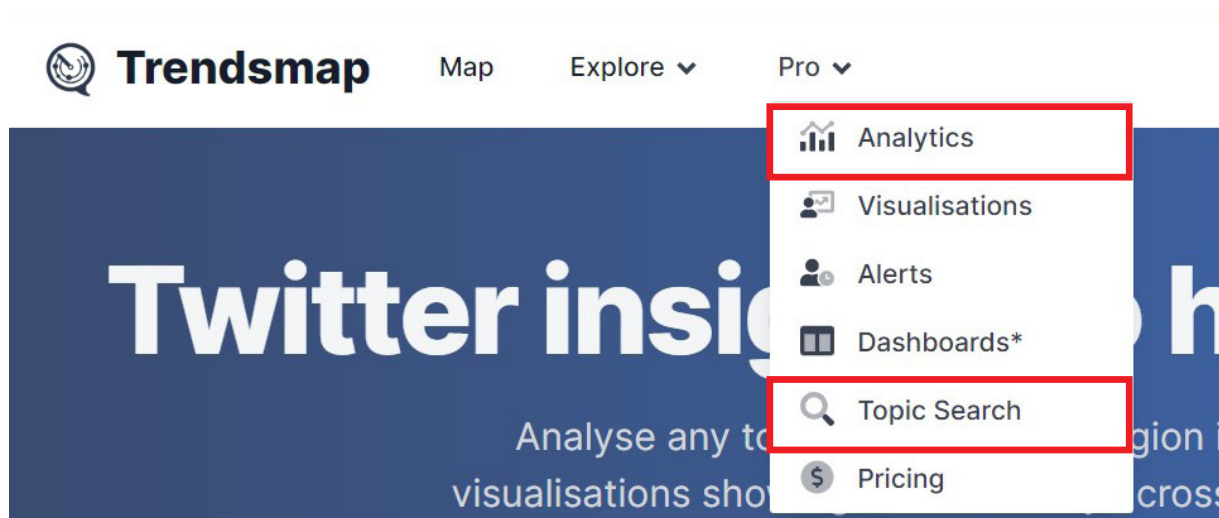
Under *Analytics*, a search window is presented where you can type in your keyword or hashtag of choice. On the free tier, we are limited to a history of 1 week.

Once you have input a keyword, username or hashtag, hit the update button and wait for the analytics dashboard to appear. This may take a few seconds.

The analytics dashboard will have a lot of useful information and statistics that you can use for your reporting. This information includes a summary of tweets, a map (including an optional heatmap), a timeline of the volume of tweets, data on countries, cities, languages, associated hashtags, emojis, and source apps all related to your search keyword.
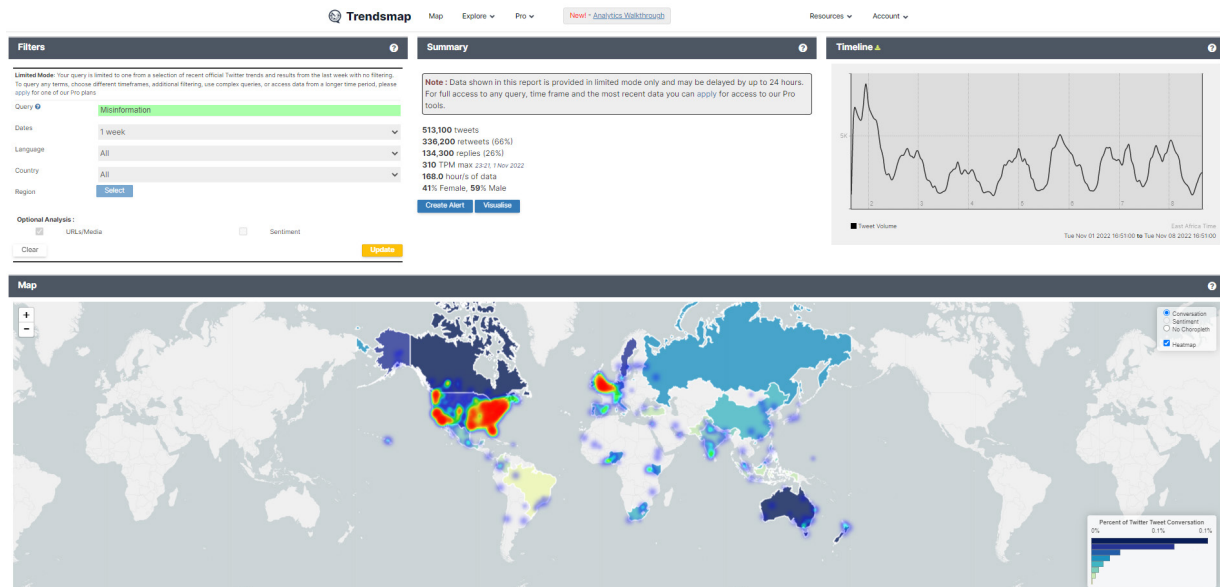
You can download each individual analysis by clicking on the download icon, however this is a paid feature. You can also find out what each chart represents by clicking on the help icon to the right of the chart.
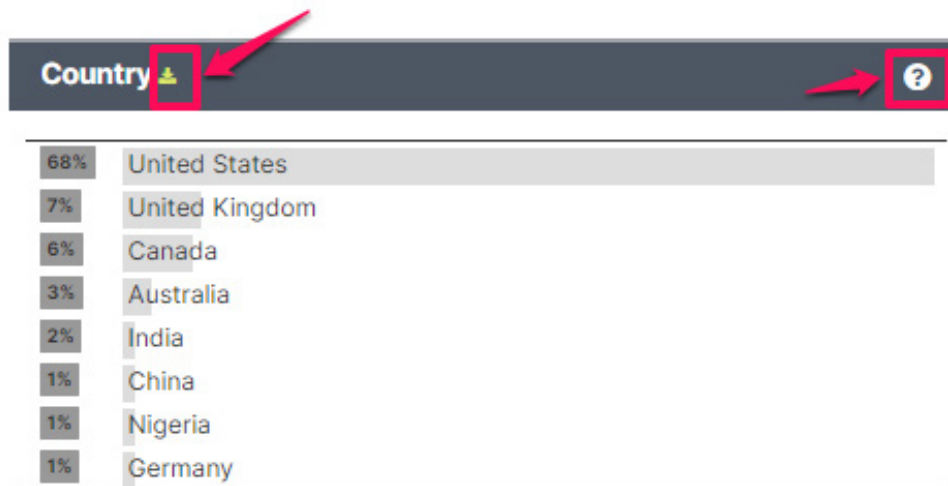
The **Topic Search** tool works similarly to the analytics tool, except that it has one extra step. Upon clicking on the **Topic Search** tool, you will be directed to a page with a search window.

Topic Search

Use this tool to search Twitter for hashtags, words, and users that contain a specific term.

This allows you to find even more relevant Twitter content by using these results to expand your search queries in our Analytics, Alerts and Visualisation tools.

Keyword to match : *(eg bbc, trump, #trump, brexit, vote)*

Result types :
- ☑ Hashtags
- ☑ Words
- ☑ Users

[Search]

No results

Input a key word onto the search box, specify the result type you would like to return and hit the **Search** button. The results will be in a bar chart format with the most relevant result being at the top with a corresponding score.

Topic Search

Use this tool to search Twitter for hashtags, words, and users that contain a specific term.

This allows you to find even more relevant Twitter content by using these results to expand your search queries in our Analytics, Alerts and Visualisation tools.

Keyword to match : *(eg bbc, trump, #trump, brexit, vote)*

hate speech

Result types :
- ☑ Hashtags
- ☑ Words
- ☑ Users

[Search]

Found 50 results

Score	Keyword	Actions
100	whatever	Analyse Visualise Alert
17	haters	Analyse Visualise Alert
15	hates	Analyse Visualise Alert
10	hated	Analyse Visualise Alert
8	hater	Analyse Visualise Alert
7	@thatericalper	Analyse Visualise Alert
7	hateful	Analyse Visualise Alert
5	@amirhatef3	Analyse Visualise Alert
5	@hatedbyallofyou	Analyse Visualise Alert
4	hatenablog	Analyse Visualise Alert
3	hatena	Analyse Visualise Alert
2	@thatesselguy	Analyse Visualise Alert
1	@godhatesnftees	Analyse Visualise Alert
1	chateada	Analyse Visualise Alert
1	@chatelaiine	Analyse Visualise Alert
1	ngehate	Analyse Visualise Alert
1	chateado	Analyse Visualise Alert
1	#godhatesnftees	Analyse Visualise Alert

For each of the results, you can select an action from the right end. The actions include Analyze, Visualize and Alert. The **Analyze** action will build an analytics dashboard similar to the one showcased in the **Analytics** section. The Visualize action will create a live map showing the appearance of the keyword or hashtag globally over a period of 1 week.
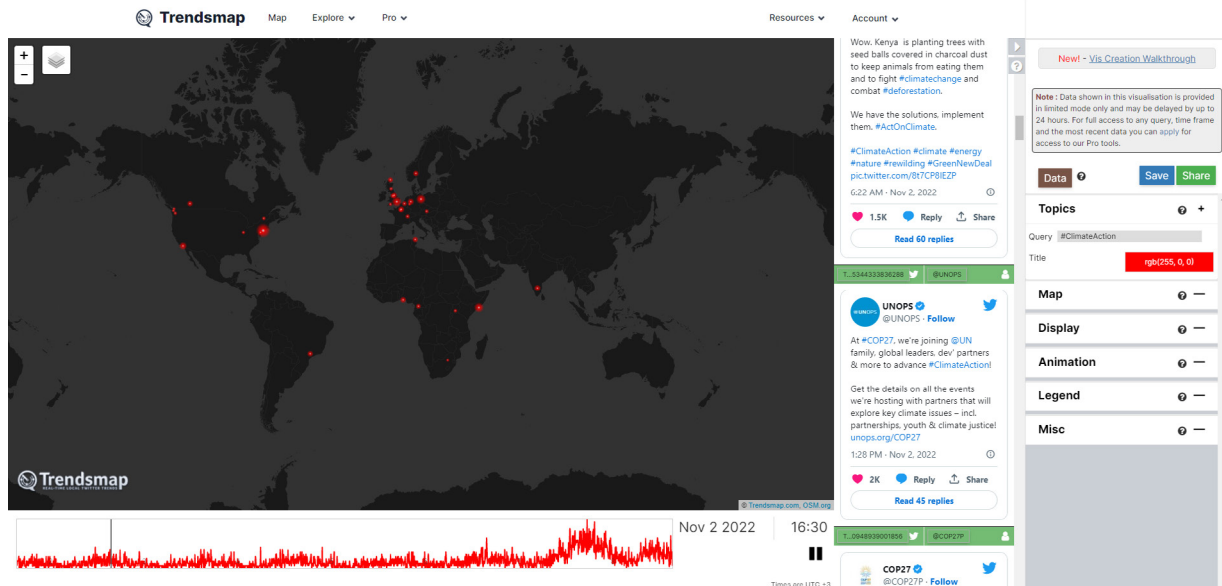
You can use the time filter at the bottom of the map to skip to a particular time or date and see how the hashtag or keyword was trending at that point in time.

The **Alert** function allows you to configure an email alert system where you will get notified about that particular topic. This functionality, however, requires a paid subscription to Trendsmap.

Tech Tool Practice

For this tool, we will practice by searching for a particular keyword of choice and building an analytics dashboard around it. For my example, I will search for the keyword 'Disinformation'

1. Under the **Pro** tab, click on **Analytics**
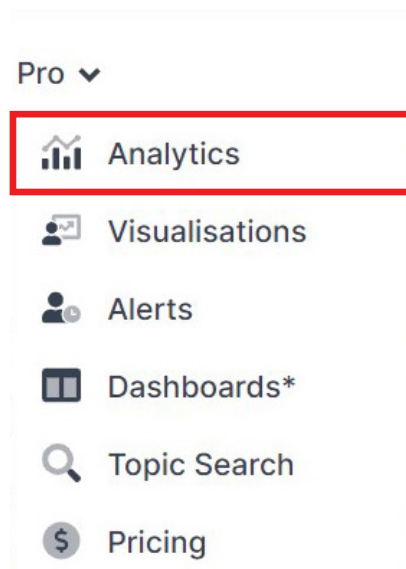
2. On the search box, we will put in our hashtag of choice and click *Update*.

3. Observe the resulting analytics dashboard and discuss any interesting insights you may encounter.
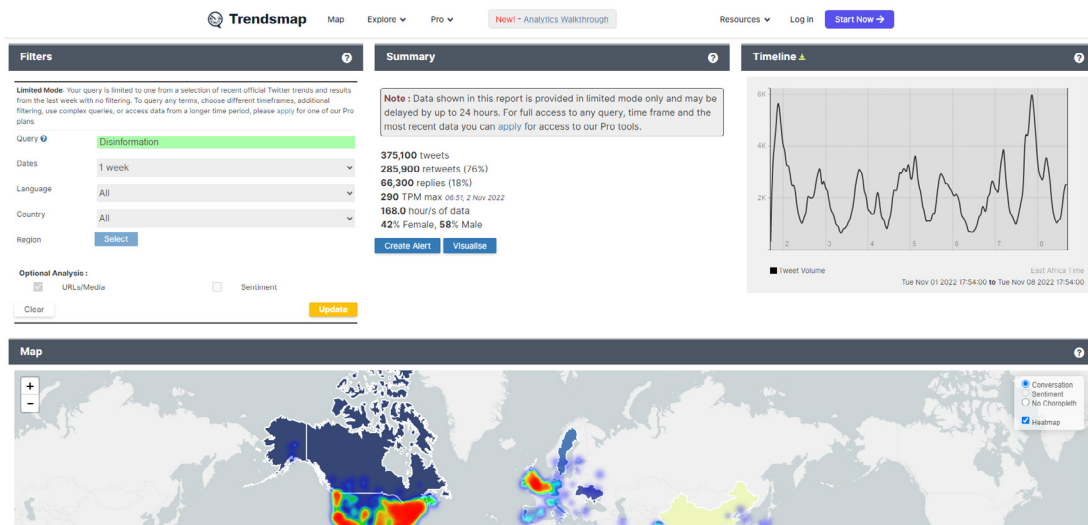
Module 3 and 4
Activity Template

Module 3 and 4 Activity Template

Analyze an action you took – with your organization – that aimed to combat hate speech. This could be an action or counterspeech effort, or any other method described today.

(If you have not participated in any anti-hate speech activity described today before, choose one of the methods and answer the questions on the worksheet hypothetically.)

To prepare, answer the following questions:

1. What did the action consist of? Describe in a few sentences. (If the action you took involved several steps, list them in chronological order.)

2. What was the stated purpose of the action?

3. Reflecting on how you implemented the action - consider the two questions below. What worked well? What would you do differently if you had the chance to re-do the action? *Write in what you can for each column.*

+ What worked well?	△ What would you change next time?

4. Now reflect on the type of method your action was in general - **What are the strategic advantages of this method in your context?** *(for example: low effort, everybody could participate.....)* and, **What are the strategic drawbacks of this method in your context?** *(for example: it was complicated to implement...). Write your answers in each column.*

+ What are the strengths of this method?	△ What are the weaknesses of this method?

▷ **Some of the Strategic Evaluation Questions you may want to consider above include:**

  – High level of participation v. low level of participation by the general public;

  – High level of reach v. low level of reach in terms of visibility for your cause;

  – High level of security/safety risk of this method v. low level of security/safety risk of this method in your context;

  – High level of complexity v. low level of complexity of the tech tool used;

  – High level of effort v. low level of effort from you team for the implementation of the action;

  – High costs involved v. low costs involved;

  – Etc ...

What other consideration do you think should be on this list?

COUNTERING HATE SPEECH STRATEGIES TRAINING MANUAL

FACILITATOR'S GUIDE

Training Overview

Participants will be introduced to some of the "best practices" for countering hate speech and will be trained on low-cost, easy-to-use technologies to address hate speech. Participants will also have the opportunity to present their experiences with countering hate speech, learn from one another on addressing hate speech – and its root causes – in their communities.

Training Objectives

By the end of this program, participants will have a working knowledge of…

- What constitutes hate speech and frameworks for understanding hate speech
- What constitutes counterspeech and frameworks for using counterspeech to address hate speech in a variety of contexts
- Ways of monitoring and analyzing hate speech in both online and offline contexts
- Gain facility with a variety of strategies for addressing hate speech in both online and offline contexts
- How to develop a response plan to hate speech arising in their context that is suitable to the direct situation and available resources.

Suggested Training Guidelines

For the learning sessions to be successful, guidelines for the program group include the following general agreements to ensure a safe, respectful learning environment.

"We, the training participants, agree to..."

- Follow Chatham House rules

  "When a meeting, or part thereof, is held under the Chatham House Rule, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed."

- Ask for permission and consent before referencing other people's experiences outside of the group, whether attributed or unattributed
- Be respectful
- Follow the schedule and timelines
- Collective Responsibility: create conducive learning environment in the online space
- Be responsive/reactive when people want to connect
- Share and exchange information, knowledge, and experiences
- Ask questions
- Avoid assumptions

Structure of the Workbook

Each content module is structured to contain learning objectives, activities for the module, selected resources, discussion questions, and case studies. The modules are broken down into submodules that cover specific sub-sections of the module content. The workbook is structured for reading, but the delivery of the training may need to be ordered differently. For example, the "Tech Tools Corner" sections of the workbook have the training information for the included tools, apps, and platforms which will need to be taught prior to completing the activities for most sections.

How to use this facilitator's guide

This guide provides information for individuals or teams facilitating training to counter hate speech both online and offline. The facilitator's guide is laid out to assist you with presenting the information in each module. It provides an overview of the workbook information and the orientation of the training program, as well as specific information for each module. While the creators of the program envision facilitation of the entire workbook content, it is possible to teach individual modules (or to teach different modules to different participant groups) based on the needs of your organization. Each section outlines one module (inclusive of all related submodules) from the workbook that can be used individually or in coordination.

To support whatever facilitation schedule is needed, this guide includes suggested session structures, learning objectives, module content outlines, and main concepts for each module. These features are included to allow facilitators to easily absorb the content to be covered in each module of the workbook and to adapt it to their needs. For example, if facilitators would prefer to use slides to present the workbook materials, the "Main Concepts" sections below can be easily used to produce presentation slides for each module or submodule. Additionally, in the "Module Content Outline," the inclusion of the case study names with hyperlinks to the supporting documents/background information will allow facilitator's to quickly familiarize themselves with the specific case studies in the workbook prior to facilitating any aspect of the program. Similarly, the inclusion of each module's activity details allows facilitators to review the activity steps and prepare for their respective sessions.

Feedback from the Workbook creators' facilitation experience:

- Engaging participants early in each session through ice breakers and discussion questions, especially those that incorporate the participants personal experiences, increases the productivity of each session.
- If utilizing presentation slides to facilitate the program, it is important to maintain a strong focus on the participation (interactive) aspects of the materials. For example, our facilitators found that limiting slide content to main concepts only and the inclusion of a case study (select one from those available) for discussion worked well.
- **Modules 3** and **4** include content around specific strategies for countering hate speech. These are content-heavy modules and may require more time to facilitate than **Modules 1** and **2** which provide background, context, and preparation for the later modules.
- Conducting the technology trainings may sometimes take longer than initially planned. It may be beneficial to either set aside extended time or to conduct the technology onboarding (install, account setup, etc) prior to the content sessions. If the technologies are very new for participants, it may even be beneficial to do the technology training separately from the content provision to ensure that participants can focus on each aspect fully.

Module 1: Frameworks for Defining Hate Speech

Ice Breaker

[Use 10 minutes to get the participants actively engaged and talking!]

Group Discussion

*[Begin the discussion of Hate Speech – **discussion questions in workbook**]*

*Review of Tech Tools [Content in "Tech Tools Corner" **section at end of module**]*

*["How To" videos can be found as **links in the workbook**]*

Jamboard is used for **Module 1** activities. Jamboard is a Google application. It can be accessed by anyone with a Google account. We suggest that the facilitator set up the "Jam" pages for the activity and share the link in advance of the class with instructions. This will provide time to manage any access issues before the session begins.

Example Session Activity: Jamboard Discussion

*Read Discussion Questions [**included in workbook**]*

*Explain Activity [activity details and Jamboard instructions **included in workbook**]*
   Activity Steps:
   1. Break into Discussion Groups (2-4 people per group depending on number of session participants)
   2. Research and Report (provide 20 minutes "research" and Jam board entry time for the groups to think through the questions. Then provide roughly 5 minutes per group (as many as is practical) to "report" their findings (share their work).

*[Discussion questions and exercise **instructions included in workbook**]*

MODULE CONTENT OUTLINE:

Learning Objectives:

By the end of this module, participants will be able to...

- Understand the landscape of hate speech definitions and frameworks, and know how to situate your work in it;
- Define and discuss the dangerous speech framework and contrast it with the hate speech definitions and frameworks;
- Discuss issues that come up around hate speech laws, social media company's policy, and free speech;
- Understand what mis- and disinformation are, and how they relate to hate speech.

SUBMODULE 1: Defining Hate Speech

[Main Concepts]

- The hate speech landscape includes a wide array of definitions determined by the mission, laws, and regulations of stakeholders (nations, NGOs, tech companies, law enforcement, etc.).
- Conceptually, most agree that at its core, hate speech is *any form of expression that seeks to promote hate and targets individuals, or groups based on shared and innate characteristics*.
- Hate speech is also widely recognized as a stepping-stone and precursor to the perpetration of mass atrocity crimes, such as genocide.
- Because of its connection with violence and crime, legality is another way of categorizing hate speech.
- A broader lens incorporates "insulting and inflammatory terms and phrases." This framework focuses on impact, rather than a strict definition or the intent of the speech, allows a nuanced approach to mapping and monitoring the phenomenon in its specific contexts.
- It is important that you or your organization positions itself and its work in this field and adopts a clear definition and frame of reference for hate speech.

*[Featured Resources – **included in workbook**]*

*[Discussion Questions – **included in workbook**]*

SUBMODULE 2: Dangerous Speech

[Main Concepts]

- Susan Benesch proposes an alternative framework of "dangerous speech" which highlights the effects and impacts of inflammatory as well as hateful communication.
- Based on this definition, the Dangerous Speech Project offers a detailed framework to identify and understand the impact of dangerous speech.

*[Case studies – details **included in workbook**]*

- President Rodrigo Duterte's Dangerous Speech in the Philippines

*[Featured Resources **included in workbook**]*

*[Discussion Questions – **included in workbook**]*

[Main Concepts]

- The content and application of national hate speech <u>laws and regulations vary from country to country</u>.
- Social media companies – as private, for-profit ventures – set their own conditions and standards of use for their platforms.
- Many states are adopting laws and regulations designed specifically to target online content or to force social media companies to take more aggressive measures to contain hateful and toxic content.
- Combating hate speech brings up important issues around freedom of speech because laws and regulations on hate speech can be misused to silence dissent, target political opponents, and crack down on minorities.

*[Case Studies – details **<u>included in workbook</u>**]*

- <u>Responding to 'Hate Speech': Comparative Overview of Six EU countries</u>
- <u>Turkish Government Wants Silicon Valley to Do Its Dirty Work</u>
- <u>Report of the Special Rapporteur on the Situation of Human Rights in Myanmar (09/2020)</u>

*[Featured Resources – **<u>included in workbook</u>**]*

*[Discussion Questions – **<u>included in workbook</u>**]*

[Main Concepts]

- Mis- and Disinformation are <u>not synonymous</u> terms. Misinformation is when the person spreading the message believes it is true. Disinformation, on the other hand, is spread when the person sharing false information knows that what they are sharing is **not** true.

- Even though the intention behind the spread of mis-/disinformation may be different, both types of false information can do a lot of harm. Just like hate speech, mis-/disinformation can be, and is increasingly, inflammatory.

- Hate speech and mis-/disinformation can be closely intertwined, using both direct and indirect messages of hate, promoting narratives that stoke existential fear, intensifying the sense of division between communities, and ultimately enabling calls for mass violence.

*[Case Studies – detail **included in workbook**]*

- [Coronajihad: Islamophobic Hate Speech and Disinformation During the First Wave of the COVID-19 Outbreak in India](#)

*[Featured Resources – **included in workbook**]*

*[Discussion Questions – **included in workbook**]*

[TECHNOLOGY TOOLS CORNER: Jamboard guide and screenshots **included in workbook**]

Module 2: Hate Speech Monitoring: Social Media Monitoring, Data Analysis, and Narrative Analysis

Ice Breaker

[Use 10 minutes to get the participants actively engaged and talking!]

Discussion

*[Begin the discussion of how familiar participants are with the four-step process – **Questions in workbook**]*

*Review of Tech Tools [Content in "Tech Tools Corner" **section at end of module**]*

*["How To" videos can be found as **links in the workbook**]*

KoBo Toolbox collects data using powerful form creation engines that cater for internet unavailability and instability. Get your data all in one place and as clean as possible.

Example Session Activity

For this module's activity, the focus is on the identified aspects of creating an action plan using the four-step approach outlined in the workbook. Use Jamboard again to have participants share their thoughts and record their ideas (this will provide tool learning reinforcement).

Explain activity, break participants into small groups, and assign each group a step in the process to sketch as part of a plan for their organization.

Activity Steps:

- **Step 1:** Identify a hate based problem
- **Step 2:** List its features
- **Step 3:** Select one step from the process (Online Safety, Data Collection, Media Monitoring, or Data Analysis) for each group to focus on and sketch a plan
- **Step 4:** Have participants post their sketch outline of the plan section to Jamboard
- **Step 5:** Discuss

MODULE CONTENT OUTLINE

Learning Objectives

By the end of this module, participants will be able to...

- Understand the potential risks of working to combat hate speech online and begin to develop a "safety" plan for themselves and their organizations.

- Understand the different steps of effective monitoring and data analysis, and how it can inform project and program design combating hate speech

SUBMODULE 1: Online Security and Safety

[Main Concepts:]

- When responding to hate speech in online contexts, counter speakers may experience mental and emotional responses to hateful content, and they may become targets of hate proponents.
- Before participating in counter speech dialogues or campaigns online it is important to assess your online "footprint" – all the sites where you maintain a presence online, have posted online materials, and posted material about you.
- Doxxing" (dox is short for documents) or publishing your personal information in public or private online forums with the intent of having others use the information to harass targets.
- Be prepared and have a response plan including documenting the harassment, reporting the harassment, making accounts private, or muting responses.
- It is important to find ways to decompress from online engagement in order to sustain your mental and emotional well-being to ensure your capacity to continue working against hate.

*[Case Studies – details **included in the workbook**]*

- Doxxing in Tunisia
- 'It Gets to You.': Extremism Researchers Confront the Unseen Toll of Their Work

*[Featured Resources – **included in workbook**]*

*[Discussion Questions – **included in workbook**]*

SUBMODULE 2: Gathering Information and Monitoring

[Main Concepts]

- Technology and data tools can substantially enhance your information gathering and monitoring capability, especially when you are looking to monitor online hate speech.
- It can be helpful to take a broader view and seek to understand the 'landscape' of hate speech before seeking to address the phenomenon itself to understand the channels of circulation, narratives, and actors involved.
- A robust information gathering, and monitoring system should be tailored to your context. It can associate low tech and high tech, as well as offline and online methods, especially when it comes to gathering information.
- Information gathering can be done by people using techniques such as in-person surveys, text messaging apps, and interviews. It can also be done using automated systems for data gathering, media monitoring, and data visualization. Each of these

methods presents advantages and limitations. It is also important to be strategic by having a clear idea what information needs to be gathered. [***Example survey questionnaires in workbook***]

- Browse the Technology Tools Corner in each module and look into what tools are best adapted to your and your teams' capabilities and needs!

[Case Studies – details ***included in workbook***]

- Umati: Monitoring Online Dangerous Speech During the 2015 Elections
- PeaceInsight: Monitoring Online Hate Speech to Inform Programming
- Mapping and Monitoring Hate Speech Directed at Jewish Lawmakers in The US

[Featured Resources – ***included in workbook***]

[Discussion Questions – ***included in workbook***]

SUBMODULE 3: Analyzing the Data

[Main Concepts]

- Data gathered through monitoring activities can provide a number of analytical insights. One way of looking at the data is **quantitatively**: For example, which terms are most prevalent online, on what platform, and used by what actors? Another way to analyze the data is **qualitatively**: how are hateful terms being used, what other terms are they associated with, do their meaning evolve?
- "Critical episodes" or "catalyst events" are events such as elections, violent incidents, natural disasters, assassinations, and other events that can polarize opinions at the regional or national level. They may also exacerbate tensions and conflict dynamics.
- Incitement (hate speech that openly calls for violence against a group based on a common characteristic) has a high risk of leading to offline violence. Especially when the message is spread by a person (or group) with authority and influence.
- Identifying the actors and social media accounts that are the most prolific hate speakers and have the widest reach provides important information when considering targeted actions and strategies against hate speech.
- Multiple methods can be used to analyze data qualitatively. The goal of qualitative analysis is to understand the *meaning* of the data in its context particularly when "irony," "satire," or "humor" are used as vehicles for hate because these can often be difficult to parse in simple quantitative terms.
- Qualitative analysis may involve surveys, focus groups, or narrative analysis of texts and images. Although typically not done at as large a scale as quantitative studies because qualitative analysis takes more time, it can provide nuanced and rich findings that are especially useful for creating interventions and developing counter speech.

*[Case Studies – **details included in workbook**]*

- Data Collection, Analysis and Use in Protracted Humanitarian Crises (Bangladesh & Iraq)

*[Featured Resources – **included in workbook**]*

*[Discussion Questions – **included in workbook**]*

SUBMODULE 4: Narrative Analysis

[Main Concepts:]

- Narrative analysis approaches provide a way to understand how stories function in communities and how to intervene when destructive stories circulate.
- Why is Narrative Analysis Important? Because competing narratives and histories are drivers of conflict, hate, and violence.
- Narrative Analysis approaches include Structural, Functional, and Poststructural forms of analysis which can be used separately or together in various stages of the research process. *[**Details and summary of narrative analyses included in workbook**]*
- Narrative analysis relies on one's understanding of the history of conflict and current conflict dynamics, as well as the actors and parties to the conflict(s) so it is essential to conduct a context/conflict analysis prior to conducting a narrative analysis.

*[Case Studies – **details included in workbook**]*

- An analytical framework for reconciliation processes – Two case studies in the context of post-war Bosnia and Herzegovina

*[Featured Resources – **included in workbook**]*

*[Discussion Questions – **included in workbook**]*

[TECHNOLOGY TOOLS CORNER: KoBo Toolbox guide and screenshots **included in workbook**]

Module 3: Combating Hate Speech: Strategies and Practices – Part 1

Ice Breaker

[Use 10 minutes to get the participants actively engaged and talking!]

Discussion

*[Begin the discussion of Strategies – **questions in workbook**]*

*Review of Tech Tools [Content in "Tech Tools Corner" **section at end of module**]*

*["How To" videos can be found as **links in the workbook**]*

Crowdtangle Link Checker is a Google Chrome extension that searches for the URL you have specified through Facebook, Twitter, Reddit and Instagram to see if any of the accounts, Pages, or profiles in the CrowdTangle database have shared that link. This can be useful to check if and how a certain link has been shared across these social media network sites. To get started with the CrowdTangle extension, go to the Chrome web store and search for the extension. Kindy note that this only works with Google Chrome as at the time of writing.

Example Session Activity

For this module's activity, the focus is analyzing an action the participants have previously taken (with their organization) that aimed to combat hate speech. This could be an action or counter speech effort, or any other method of countering hate speech described in **module 3** and **4** (If participants have not participated in any anti-hate speech activity before, have them choose one of the methods discussed in **module 3** or **4** and answer the questions on the worksheet hypothetically.)

Activity Steps:

Prior to Session: Participants download and fill out the activity template for their analysis *[**template included in workbook materials**]*

During the session: Participants will share some of the insights they gained reflecting on the action   they took (the format is flexible – large group, small groups, or even additional Jamboard reflections).

Learning Objectives

By the end of this module, participants will...

- Become familiar with practical methods to combat hate speech
- Identify methods that will address the problem they have identified in their context

[Main Concepts]

- To manage harmful content is to take action. Effectively combating hate speech is as complicated as the nature of the problem. Different strategies, tactics and practices have been developed to address this phenomenon and mitigate or prevent its worst consequences.

*[Discussion Questions – **included in workbook**]*

*[Featured Resources – **included in workbook**]*

*[Content Breakdown – **list of strategies included in workbook**]*

SUBMODULE 1: Direct Response to Hateful Tweets or Messages

- **Strategy: Report Hate Speech**

*[Case Study – detail **included in workbook**]*

- Reporting Violent Extremism on Social Media

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

- **Strategy: Counterspeech (Organized and Organic)**

*[Case Studies – Organized – details **included in workbook**]*

- Panzagar ("Flower Speech") New Responses to Hatred Online
- Collective Counterspeech: The Anti-Hate Brigade, #jagärhär

*[Case Studies – Organic – details **included in workbook**]*

- Umati: Counterspeech during Kenya Elections
- Hate Speech vs. Counterspeech in Times of COVID
- Kenyans on Twitter Counter CNN

*[Featured Resources – included in workbook: **links and helpful infographics**]*

*[Questions to Examine – **included in workbook**]*

- **Strategy: Countering High Profile Hate Speakers**

  [Three suggested tactics]

  1. Pressure to retract statement
  2. Ask other high profile persons to make statements and exercise pressure
  3. Government agency / institutional intervention

  *[Case Studies – details **included in workbook**]*

  - Countering High Profile Hate Speakers in Nigeria
  - Use of "Hate Speech" Laws and Monitoring of Politicians on Social Media Platforms

  *[Featured Resources – **included in workbook**]*

  *[Questions to Examine – **included in workbook**]*

- **Strategy: Amplifying Positive Voices**

  [Three Suggested Tactics]

  1. Sharing the work of those working to curb hate speech (websites, blogs, podcasts, videos and other creations).
  2. Recommending their work or creations to your friends and colleagues.
  3. Actively promoting and supporting others countering hate within your curriculum, newsletters, websites, podcasts, social media spaces, etc.

  *[Case Studies – details **included in workbook**]*

  - A Campaign That Supports Women Peacebuilders
  - EU Coalition of Positive Messengers to Counter Online Hate Speech
  - Giving Voice to the Voiceless or "Help Speech"

  *[Featured Resources – **included in workbook**]*

  *[Questions to Examine – **included in workbook**]*

- **Strategy: Countering Mis- and Disinformation**

  *[Case Studies – details **included in workbook**]*

  - #ThinkB4UClick – Raising Awareness on the Misuse of Social Media
  - Indivisible Truth Brigade

  *[Featured Resources – **included in workbook**]*

  *[Questions to Examine – **included in workbook**]*

SUBMODULE 2: Dialogue and Other Engagement: In-person and Digital Dialogues

◦ **Strategy: In-person or Digital Dialogue**

*[Case Studies – details **included in workbook**]*

- Reaching Out to Dangerous Speakers in Nigeria
- The Commons Project

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

◦ **Strategy: Digital Responses at Scale**

*[Case Studies – details **included in workbook**]*

- Stealing #StopIslam
- Multi-Hashtag/Multi-Platform K-Pop Campaigns against Hate (Online)

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

◦ **Strategy: "Naming and Shaming"**

*[Case Studies – details **included in workbook**]*

- Naming and Shaming in Kenya by a National Commission

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

SUBMODULE 3: Building Response Mechanism to Prevent or Mitigate Offline Violence

◦ **Strategy: Network of Trained First Responders**

*[Case Studies – details **included in workbook**]*

- A Local First Approach to Early Warning and Response in Kenya
- A Citizen Journalism Framework That Crowdsources and Maps Conflict Incidents in Cameroon for Rapid Response

◦ **Strategy: Network of Stakeholders with the Capability to Respond/React**

*[Case Studies – details **included in workbook**]*

- An EWER Tool Showing How the State and Its Partners Can Advance Commitments to Work for Peace and Cohesion
- INHOPE: Hotlines and Complaint Forms

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

SUBMODULE 4: Strategic Nonviolent Action

*[Case Studies – details **included in workbook**]*

- English Sports Fraternity Boycott of Social Media to Protest Online Abuse
- Stop Hate for Profit
- Japanese Protestors Stomp Out Hate

*[Featured Resources – SNAP and Explore – **included in workbook**]*

[TECHNOLOGY TOOLS CORNER: Crowdtangle Link Checker guide and screenshots **included in workbook**]

Module 4: Combating Hate Speech: Strategies and Practices – Part 2

Ice Breaker

[Use 10 minutes to get the participants actively engaged and talking!]

Discussion

*[Begin the discussion of Strategies - **questions in workbook**]*

 Review of Tech Tools *[Content in "Tech Tools Corner" **section at end of module**]*

*["How To" videos can be found as **links in the workbook**]*

- Vicinitas is an analytics platform that enables users to fetch Twitter data related to a particular hashtag, keyword or user.
- Trendsmap, at its very core, is a mapping tool with a niche for visualizing twitter insights around the globe.

Example Session Activity

For this module's activity, the focus is analyzing the action discussed in the **Module 3** activity and thinking about how participants might help their organization to convert it to a larger-scale action. Participants will use their reflections about the action they discussed in the **Module 3** activity to guide their thoughts about one of the strategies in **Module 4** (below) to think about whether "scaling up" is possible or useful. The participants have previously taken (with their organization) that aimed to combat hate speech. (If participants have not participated in any anti-hate speech activity before, have them choose one of the methods discussed in **module 3** or **4** and answer the questions on the worksheet hypothetically.)

Activity Steps:

Select one of the large-scale strategies from **Module 4**

Assess the utility of "scaling up" to this new approach: Is it useful to address the problem your organization wants to address? Is it feasible to plan, implement, and manage this type of strategy within your context? Does scaling up help your organization achieve its goals, why or why not?

Learning Objectives:

By the end of this module, participants will be able to...

- Become familiar with more practical methods to combat hate speech
- Identify the methods that are the most adapted for your context

[Main Concepts]

- Some strategies and practices are bigger in scope when it comes to implementation and reach, but they may run longer and have broader impact.
- It is important to consider the potential benefits and challenges as well as the practicalities of developing larger-scale campaigns to combat hate.

You will need to determine whether offline or online strategies, or in some cases blended strategies (e.g., an offline campaign paired with an online or social media component), are suited to the local context in which the campaign is conducted. In both offline and online contexts planning campaigns should consider and build-in mechanisms to ensure campaigner safety (see Module 2).

*[Discussion Questions – **included in workbook**]*

*[Featured Resources – **included in workbook**]*

*[Content Breakdown – **included in workbook**]*

SUBMODULE 1: Inform or Educate the Wider Public

◦ **Strategy: Hate Speech Literacy and Education**

*[Case Studies – details **included in workbook**]*

- Fighting Hate Speech Through Schools and the Media in the EU
- Educating Leaders and the Public on Hate speech Through Media

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

◦ **Strategy: Promotion of Media Literacy**

*[Case study – details **included in workbook**]*

- MediaSmarts: Promoting media and digital literacy.

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

◦ **Strategy: Media Campaigns/Public Awareness Raising Campaigns**

*[Case Studies – details **included in workbook**]*

- BRICkS - EU Campaign Against Online Hate Speech
- A Counter Narrative Campaign in France
- No Hate Speech Video Competition in Spain
- The Blackout Tuesday Campaign Overshadowed the Black Lives Matter Movement in the United States
- #KickItOut: Tackling Discrimination on Social Media

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

◦ **Strategy: Youth Counterspeech Initiatives and Resilience Building**

*[Case Studies – details **included in workbook**]*

- The Game Changer Project – Engaging Youth to Counter Hate
- PERIL: Addressing Youth Radicalization and Extremism to Build Resilience

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

◦ **Strategy: Adoption of A Hate Speech Charter**

*[Case Studies – details **included in workbook**]*

- Women's Peace Conference in Cameroon to address the Anglophone Crisis
- Media Workers in Central and West Africa Call on the Media to Address Hate Speech and Reaffirm Ethical Standards
- "Manifesto of Venice" Against Gender-Based Violence in Italy

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

SUBMODULE 2: Advocacy

◦ **Strategy: Advocacy with Community Leaders and/or Public Figure**

*[Case Studies – details **included in workbook**]*

- Eradicate Hate Global Summit
- Mapping Hate Groups Across the United States
- Mapping Intolerance Utilizing Social Media Conversations in Italy

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

◦ **Strategy: Advocacy with Social Media Platforms (Content removal, Draw attention to specific cases / issues, Clarify monitoring rules and decisions to the wider public)**

*[Case studies – details **included in workbook**]*

- Taking down fake content on social media
- Online Campaign to Take Down Hate Speech on the Internet
- Bringing the World's Attention to the Hate Speech Against the Rohingya and

Muslims in Myanmar
- Ethiopian Network Against Hate Speech
- Civil Society Urges Governance of Social Media Leading up to Myanmar's 2020 Election

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

- **Strategy: Legal Advocacy (National legislation, Better legislation, & Better implementation of existing legislation)**

*[Case Studies – details **included in workbook**]*

- Organizations Persuade the Malaysian Government to Implement an Anti-Hate Speech Regulation
- Rwanda, Sectarianism, Divisionism, and Genocide Ideology
- Surveillance Laws Propagated by Hate Speech Laws
- Myanmar: Government's Approach to 'Hate Speech' Fundamentally Flawed
- Forming a Working Group to Lobby for an Amendment in the Hate Speech-Related Criminal Code in Georgia
- Facebook Faces a Lawsuit Over Hate Speech and Disinformation in France

*[Featured Resources – **included in workbook**]*

*[Questions to Examine – **included in workbook**]*

SUBMODULE 3: Combatting Hate Speech by Addressing its Root Causes and Drivers

- **Strategy: Addressing Drivers of Hate Speech**

*[Case Studies – details **included in workbook**]*

- Addressing and Challenging the Stigmatization and Prejudiced Stereotyping as Root Causes for Social Exclusion of Roma (Europe)
- NoRa (No Racism) Campaign: Public Discourse and Antiracism Training (Finland)
- #MyFriend Campaign

- **Strategy: Addressing Conflict via "Peace Promotion"**

*[Case Studies – details **included in the workbook**]*

- Friend Me 4 Peace: Creating Opportunities for People Who Support Peace to Connect and Break Stereotypes (Middle East)
- Engaging with Narratives for Peace

- **Strategy: Social Cohesion Initiatives and Promotion of Intercultural and Interreligious Dialogue**

*[Additional Content – details **included in workbook**]*

The Social Identity Wheel: This activity encourages participants to identify and reflect on the various ways they identify socially, how those identities become visible or more keenly felt at different times, and how those identities impact the ways others perceive or treat them. – Adapted from "Voices of Discovery."

*[Case Studies – details **included in workbook**]*

- Peacebuilding After the Genocide
- Using Textbooks to Spread Information About Both Sides of the Conflicts
- UN Adopts Resolution to Combat Hate Speech with Dialogue

*[Featured Resources (For Entire Module) – **included in workbook**]*

*[Questions to Examine (For Entire Module) – **included in workbook**]*

[TECHNOLOGY TOOLS CORNER: **Vicinitas** and **Trendsmap** guides and screenshots included in workbook]

NOTES

NOTES

NOTES