MODULE FIVE:

Non-violent Communication

Module Objective: Understand and encourage the need for and practice of non-violent communication in digital communities

Module Dilemma: My group members are insulting each other, encouraging hate speech and bullying

Understanding Online Hate Speech

WHAT IS HATE SPEECH?

The thing about hate speech is that it does not have a uniform definition in human rights law.

Hate speech regulations vary significantly by jurisdiction, particularly in how they define what constitutes hate speech and to what extent they differ by speech that is offline versus online.

There is a need for clear and narrowly circumscribed definitions of what is meant by the term “hate speech”, or objective criteria that can be applied. Over-regulation of hate speech can violate the right to freedom of expression, while under-regulation may lead to intimidation, harassment, or violence against minorities and protected groups.

Here are a couple of international definitions by various organisations.

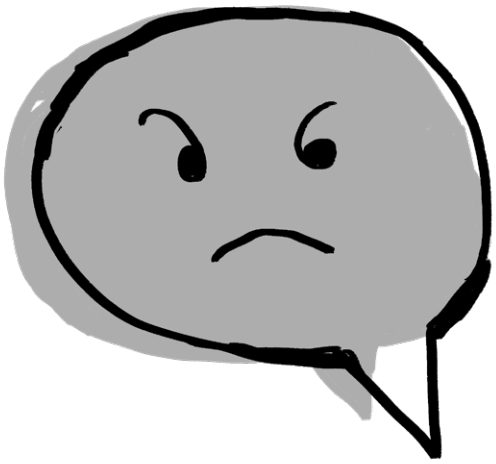
The International Covenant on Civil and Political Rights (ICCPR) Article 20 (2): Any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence shall be prohibited by law.

Article 4(a) of the International Convention on the Elimination of All Forms of Racial Discrimination: Dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin, must be declared an offence that is punishable by law.

United Nations Strategy and Plan of Action on Hate Speech: Any kind of communication in speech, writing, or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factors.

While the third definition is not a legal definition and is broader than the notion of “incitement to discrimination, hostility or violence” - prohibited under international human rights law - it highlights three important attributes: Hate speech can be conveyed through any form of expression, including images, cartoons, memes, objects, gestures, and symbols and it can be disseminated offline or online.

- Hate speech is “discriminatory” - biased, bigoted, intolerant - or “pejorative” - in other words, prejudiced, contemptuous, or demeaning - of an individual or group.
- Hate speech makes reference to real, purported, or imputed “identity factors” of an individual or a group in a broad sense: “religion, ethnicity, nationality, race, colour, descent, gender,” but also any other characteristics conveying identity, such as language, economic or social origin, disability, health status, or sexual orientation, among many others.



You can access a list of targeted groups and relevant resources for each group here. This list includes national, ethnic, religious, and linguistic minorities; migrants and refugees; women and girls; LGBTQI+; vocational targets such as journalists and activists.

HATE SPEECH VS FREEDOM OF EXPRESSION

Free speech refers to the right to seek, receive and share information and ideas with others. But this freedom must be used responsibly and can be restricted when considered as threatening or encouraging hateful activity.

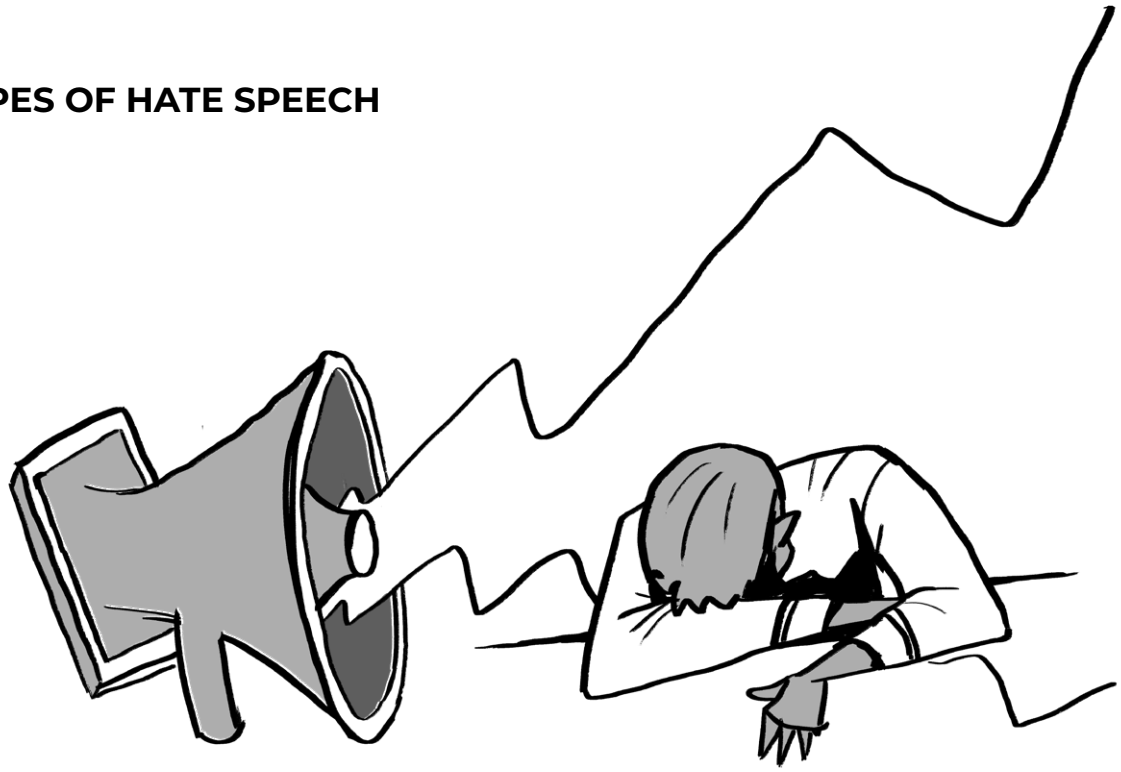
Hate speech, particularly online hate speech, targets particular groups of people – often minorities and dehumanises them. Hate speech perpetrators often see “the other” as enemies and have a tendency to connect all issues in society to these targeted communities.

Remember the key difference: Free Speech is a way to exchange, teach, learn and challenge each other’s perspectives,, whereas Hate Speech is targeting particular groups with malicious intentions and insulting individual identities. All internet platforms can be places where people post hateful content, whether as words, videos, photos, or memes, and cause great harm. It is up to all of us to consider our own online content and make sure we are not crossing the line from free speech to hate speech.

A good way to distinguish this difference between free speech vs hate speech is to use the constitution. For example, this is the case in Sri Lanka:

Some people might believe that they are within their right to say what they want – both on online and offline spaces – due to the freedom of speech and expression guaranteed by our constitution. If this is the case, then it must be also pointed out that the same constitution also guarantees that all persons are equal before the law and entitled to equal protection of the law and that no citizen will be discriminated against on the grounds of race, religion, language, caste, sex, political opinion, place of birth or any such grounds. Explain the relevance of both Article 14(1)(a) (Freedom of Expression) and article 12 (right to equality) of our constitution so that our participants understand that hate speech is not only immoral and unethical but also goes against the law.

TYPES OF HATE SPEECH



TYPE	DESCRIPTION	EXAMPLE
Disagreement	This involves disagreeing with the ideas or beliefs of a particular group.	Feminism does not exist. All feminists are wrong
Negative Actions	This highlights 'negative' nonviolent actions associated with the group.	Pro-choice activists want to ruin the future of our country by supporting abortions
Negative character	This includes negative characterization or insults towards a particular group.	All homosexuals are pedophiles.
Demonising and dehumanising	This involves belittling groups and equating them to subhuman entities.	Muslim people are pigs or homosexuals are monsters.
Violence	This outrightly calls for violence against the specific group.	Let's kill all Asians. Let's drive them out of our country.

Understanding Non-violent Communication

HOW IS DIGITAL COMMUNICATION DIFFERENT FROM IN-PERSON COMMUNICATION?

1. **Scope and Scale:** One person can send out a hateful or false message that is seen by millions of people all over the world. Hundreds of people can respond via a comment section.
2. **Anonymity:** Digital platforms allow people to communicate more easily with people we have never met and do not know.
3. **Less nonverbal cues:** Nonverbal cues such as eye contact, facial expression, hand gestures, and posture offer more opportunities for humanising and feeling empathy for others. These are missing in digital dialogue.
4. **Less Context:** In-person dialogue often relies on context cues, including ambiance, to help set a positive tone. These are missing in digital dialogue.
5. **Shorter messages:** Social media platforms emphasise short communication. Twitter limits messages to 280 characters. TikTok limits the amount of time to 60 seconds. With less space, people simplify their message to explain what they believe but rarely explain what experiences have led them to those beliefs or any complexity on the issue.
6. **More Emotional:** Emotional content spreads more rapidly. Comments or stories that evoke anger are more likely to receive engagement with "likes" or emoji markers. People may speak in more dramatic terms on social media to make up for a lack of nonverbal cues.
7. **Easier to Leave a Discussion:** Digital dialogue is easier to walk away from when discussions become tense. It can be harder to physically leave an in-person dialogue, so more people may "stay through the hard times."
8. **More Public Witnesses and Less Privacy:** Digital communication involves silent onlookers, witnesses, or bystanders. A post with communication between 2-10 people who leave comments and respond to others is common on social media. What is distinct is that the post may have hundreds or thousands of silent witnesses who read and observe the interaction.
9. **More Shaming, Humiliation, and Dehumanization and Less Dignity:** People communicate in ways on social media that are rarely seen in physical interactions. It is easier to speak harshly to shame, judge, humiliate, and dehumanise strangers on social media than it is in the physical world. People may be openly attacked and experience humiliation from public shaming on social media.
10. **Bots**: In the physical world, people do not wonder whether they are talking to a robot. On social media, there are thousands of robots ('bots') pretending to be people. Social media bots are created for a variety of reasons, both good and bad. When thousands of bots begin sharing a piece of false information, it gives others the false impression that the information is popular.

The above explanation is based on a Toda Institute report on Digital Peacebuilding Communication Skills - Beyond Counter Speech (Schirch 2020).

WHAT ARE COMMON DIGITAL RESPONSES TO PROBLEMATIC SPEECH?

There are at least eight broad patterns of responses in comment sections on social media. These strategies work better or worse depending on different audiences and contexts. They can be used together or on their own. Each can be done publicly or privately.

SILENT BYSTANDERS AND CONFLICT AVOIDERS

Conflict avoidance or choosing to be a silent bystander are by far the most common approaches people take when they encounter tense, conflicted conversations on religion, politics, health, or other issues. Silent bystanders watch but do not intervene in the digital conflict, hate speech, or disinformation. In the physical world, people may rarely witness abusive behaviour toward others. But online, the scale of conflict, hate, and false information is so great that some may feel overwhelmed by the idea of responding to it. Bystanders may choose silence for fear of making the situation worse or being implicated and pulled into the conflict.

FACT-BASED RESPONSES

Some respond to social media comments spreading false information or conspiracy theories with fact-based arguments. Fact-checking can inadvertently increase the number of people who see false information. But face checking can work if it creates doubt in some observers so that the sharing of false information declines or is deleted. It seems to work best when done with a group of supportive fact-checkers who reinforce each other.

DISTRACTION, HUMOR, OR POSITIVE RESPONSES

Another type of response to problematic speech is using humour to lighten the mood or even mock a hateful or false comment. Research on counter-speech suggests this may be helpful in some cases. It may also result in more conflict. This sensitive strategy requires care.

SHAMING AND EMOTIONAL RESPONSES

Shaming is a form of “negative counter speech” in which someone observes a comment that they perceive is harmful to others, and they shame the speaker by denouncing the values or harming them. Shaming may mock or ridicule the speaker’s beliefs, demonstrate inconsistencies in a speaker’s thinking, question their goals, or highlight the negative impact of their speech on other people. In their review of organic examples of counter-speech on Twitter, researchers found that rebuking hate speech often led to apologies or deleting the original content.

PRIVATE OR PUBLIC REQUESTS TO REMOVE OR EDIT

Another form of response is to write a public or private message to the person who wrote a problematic comment on a social media platform to make them aware that the comment is viewed as offensive and to explain why it is offensive, and then to ask them to remove or edit their comment.

UPSTANDERS

Upstanding refers to bystanders who are witnessing harassment or hate speech to intervene on behalf of the person being harassed or victimised by hate speech. Upstanders or “cyber-Samaritans” is someone who models upstanding by dissenting to harmful posts by challenging the bully or supporting the victim. When this happens, other people are more likely to join in to support.

HOW TO PRACTISE NON-VIOLENT COMMUNICATION?

If "violent" means acting in ways that result in hurt or harm, then much of how we communicate—judging others, bullying, having racial bias, blaming, finger pointing, discriminating, speaking without listening, criticizing others or ourselves, name-calling, reacting when angry, using political rhetoric, being defensive or judging who's "good/bad" or what's "right/wrong" with people—could indeed be called "violent communication."

Nonviolent Communication (NVC) is sometimes referred to as compassionate communication. Its purpose is to strengthen our ability to inspire compassion and to respond compassionately to others and to ourselves. NVC guides us to reframe how we express ourselves and hear others by focusing our consciousness on what we are observing, feeling, needing, and requesting.

Observations - NVC emphasises observation without judgement. This means presenting the simple facts we have observed. For example, instead of saying, "You have abandoned our group and never post anything anymore," you can say, "I noticed that you don't participate in the group as much as you used to."

Feelings - NVC involves taking responsibility for your feelings. This requires a change in perspective of how others' words and actions affect our feelings. In NVC, what others say and do is considered the stimulus, but never the cause of feelings. When faced with a negative message from someone else, NVC illuminates four options. To illustrate these options, let's use the example of criticism, "You're so selfish":

- Take it personally: "I really am selfish..."
- Fight back: "I'm not selfish; you're selfish!"
- Consider your own feelings and needs: Say something like: "When I hear you say that I am selfish, I feel hurt because I need some recognition of the effort I make to consider your preferences." B
- Consider the other person's feelings and needs: Ask something like: "Are you feeling hurt because you need more consideration for your preferences?"

Needs - Taking the next step, NVC makes the connection between feelings and unmet needs in the individual. These needs are common and fundamental to all human beings. The outer expression of feelings, such as anger and frustration, are seen as indicators of needs, such as love and acceptance, that are unfulfilled. For example, if a moderator of a group is feeling angry at the interactions of some group members, we need to dig deeper and think about what unmet need is causing this feeling. Is the moderator unsupported by the other members? Are they overwhelmed by the responsibilities? Are they not seen and appreciated enough?

Requests - NVC's final step is to make specific, doable requests for things that enrich the requester's life. They are made in such a way that it enables the person to respond compassionately to the request. Requests are never demanded. NVC considers demands always to be violent, intimidating, and forceful – the source of many ineffective and unhelpful communication exchanges. Requests in NVC are positive. This means requesting what you want, rather than what you don't want. An example of this would be saying: "I'd like you to support me in moderating some of the content", rather than "I don't want you to ignore me and let me do all the work by myself'.

Managing Conflict in Digital Groups

HOW DO WE MANAGE CONFLICT IN SOCIAL MEDIA GROUPS?

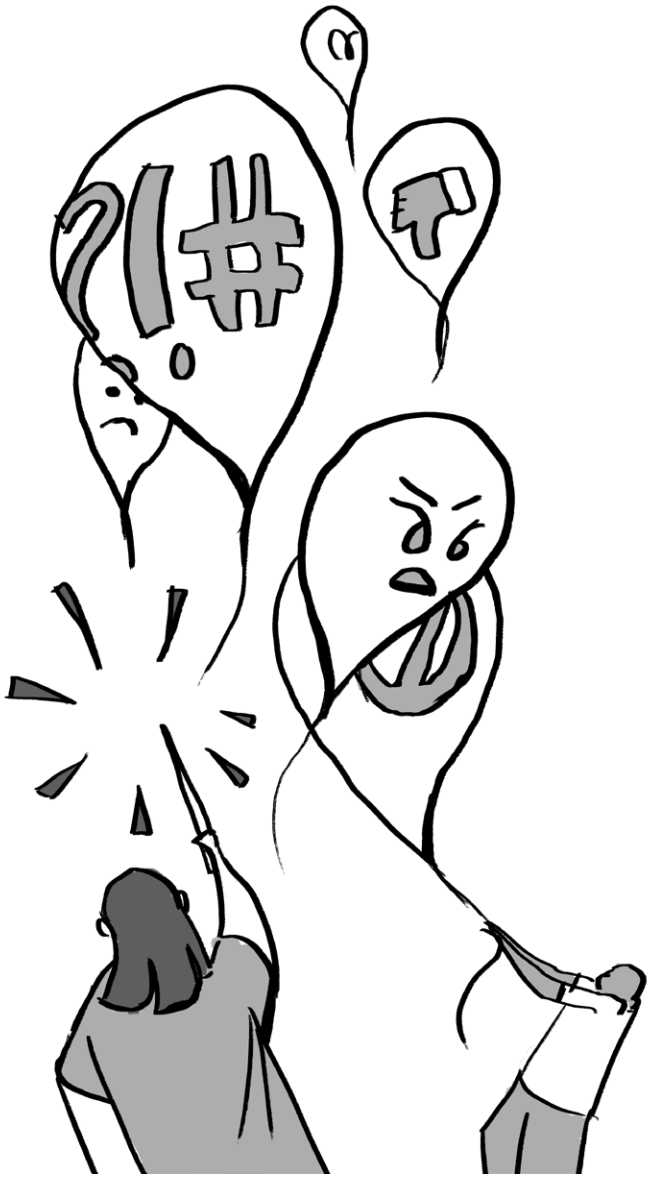
PREVENTING CONFLICT - A QUICK THROWBACK TO GROUP RULES

Many admins say preventing conflict starts with writing great rules, which serve as guidelines for member behaviour. Experienced admins recommend posting these rules long before you think you need them. Clear guidelines are useful for overall group culture when their tone is positive. Describe the behaviours you want to encourage, rather than listing only things members should not do.

REACHING OUT PRIVATELY

Experienced admins recommend privately reaching out to members involved in conflict situations. Taking the time to have a private conversation can make members feel heard and respected. Many admins share that after a private conversation, people have returned to the group as valuable and supportive members. Here is what the digital community stewards can do.

- Get to know the people involved to understand the source of conflict.
- Reach out privately to the people involved in a conflict situation and remind them of the rules.
- Help them understand how they can positively contribute to the group.



MANAGING CONFLICT BETWEEN MEMBERS

From time to time, a social media group may experience conflict between members. Conflict can arise from a difference of opinion, misunderstanding, confusion, or controversy. Respectful disagreement and debate can be part of a healthy community. If a disagreement turns disrespectful or breaks group rules, admin actions can help get the group back on the right track.

When conflict occurs, such as a heated conversation in a thread, experienced admins recommend acting quickly. Get involved before the situation escalates. Speedy action helps reset the tone of the conversation and lets group members know admins care by being present.

- If there is a thread that has gone off track, experienced admins recommend the following:
- Engage with the member in the comments and try to reset the tone of the conversation.
- If needed, restate the rules in the comments of the thread and turn off commenting.
- Leave the thread visible so others can learn what happened.

Some admins temporarily turn on post approvals to calm things down. Once things have cooled down, they'll turn post approvals off again.

If moderation isn't working, consider temporarily muting the member. Allow some time for the discussion (and the member) to cool down. When all else fails, admins should feel empowered to remove members from your group.

Many admins we spoke to told us that they were initially unsure about removing people from the group, but realised it was sometimes necessary. As an admin, you uphold the culture and rules of your community. Members appreciate your moderation, including removing other members who aren't following your community's rules.

In extreme cases, such as something that goes against Community Standards (nudity, hate speech, or threats of violence), you or someone on your team can report the post to the platform.

It is unlikely that admins can watch your group all the time. Experienced admins recommend enlisting members' help by asking them to report heated conversations to an admin so they can take appropriate action.

MANAGING DIFFICULT MEMBERS

A little bit of conflict is inevitable (even healthy) in most groups but managing conflict can be especially challenging when you have a bad actor in your group. Often this can be prevented by establishing great rules and screening new members carefully, but sometimes you'll need to take action against group members who are causing conflict. This could mean working with your team or other group members to resolve the issue, or in some cases, it may be necessary to remove someone from the group.

MANAGING CONFLICT AND NON-VIOLENT COMMUNICATION

A tool that is highly useful when practising NVC - especially with group members who might disagree with admins - is the Change Conversation Pyramid.

- **Comfort** - Make them feel safe enough to talk with you
- **Connection** - Earn their trust so they will take risks
- **Comprehension** - Learn their point of view, so they feel heard
- **Compassion** - Show you care so they will listen to your perspective
- **Cognition** - Gently encourage rethinking so they can update their beliefs

Here are some digital communication tools that will allow you to handle this situation:

1. **Work to prevent conflict before it happens.**
No group is conflict-proof, but if you are proactive, you can work to make sure people know what's allowed and that when conflicts do arise, they're solved in a consistent way. The first step is having clear rules that are very visible in your group. Also, consider keeping a list of group members you're concerned about among your team so you can watch for problem behaviour.

2. **Recognize problems before they get worse.**
Once you have a foundation of rules, the next step is working with your community and your admin team to keep conflict from escalating. Encourage your group members to report bad actors to you or your team. This will allow you to get involved early and contain most problems.

3. **Understand their point of view and diffuse the situation.**
Reach out to members who have broken the rules quickly. One good tactic is to contact these members privately, using chat or even a call, to remind them of the rules. Sometimes a simple misunderstanding can escalate because members feel cornered or ganged up on. Use active listening and try to get to the core feelings behind their statements. Repeat their feelings back to them so you can get to an understanding, i.e., "what I hear you saying...".

4. **Work with your team.**
Your team is there to back you up when things get tricky, so don't forget to use them. When dealing with a bad actor, alert your team as soon as you can. That way, they know that you're on top of it, and they can be there to offer support.

5. **Mute or remove someone and inform them about the rules they have violated.**
Sometimes, if someone is acting out in your group, they just need a little time to calm down. You can use the 'mute' feature to temporarily stop someone from posting or commenting in the group. You'll be able to set the duration of time they're muted, and you can specify to them which group rule they broke. If muting doesn't do it, you can remove them from the group.

6. **Address the situation.**
For tough situations that have escalated in your group, you can make sure everyone understands the actions you've taken by addressing them directly. You can clear the air and control any rumours by going Live, posting, or leaving a comment for the group. Be prepared to answer questions.

Here are some technical tools on Facebook Groups that will allow you to manage bad actors:

1. **Post approvals**
Post approvals will allow you and your moderation team to screen incoming posts. This is a great way to proactively manage the content that's shared in your group.

2. **Maintain an admin activity log**
Admin activity log helps you to keep track of admin and moderator activity. Here you can filter and view notes on different actions taken by your team.

3. **Set up member questions**
Setting up member questions can help you better screen incoming members to your group.

4. **Turn on keyword alerts**
If you know that a certain language is banned or indicative of conflict, you can flag those keywords and get alerted when they come up in your group.
Dealing with bad actors can be a drain on your time and emotional energy. You can often de-escalate a situation by approaching it proactively and with empathy. But at the end of the day, if someone isn't right for your group, you should always feel free to remove them.