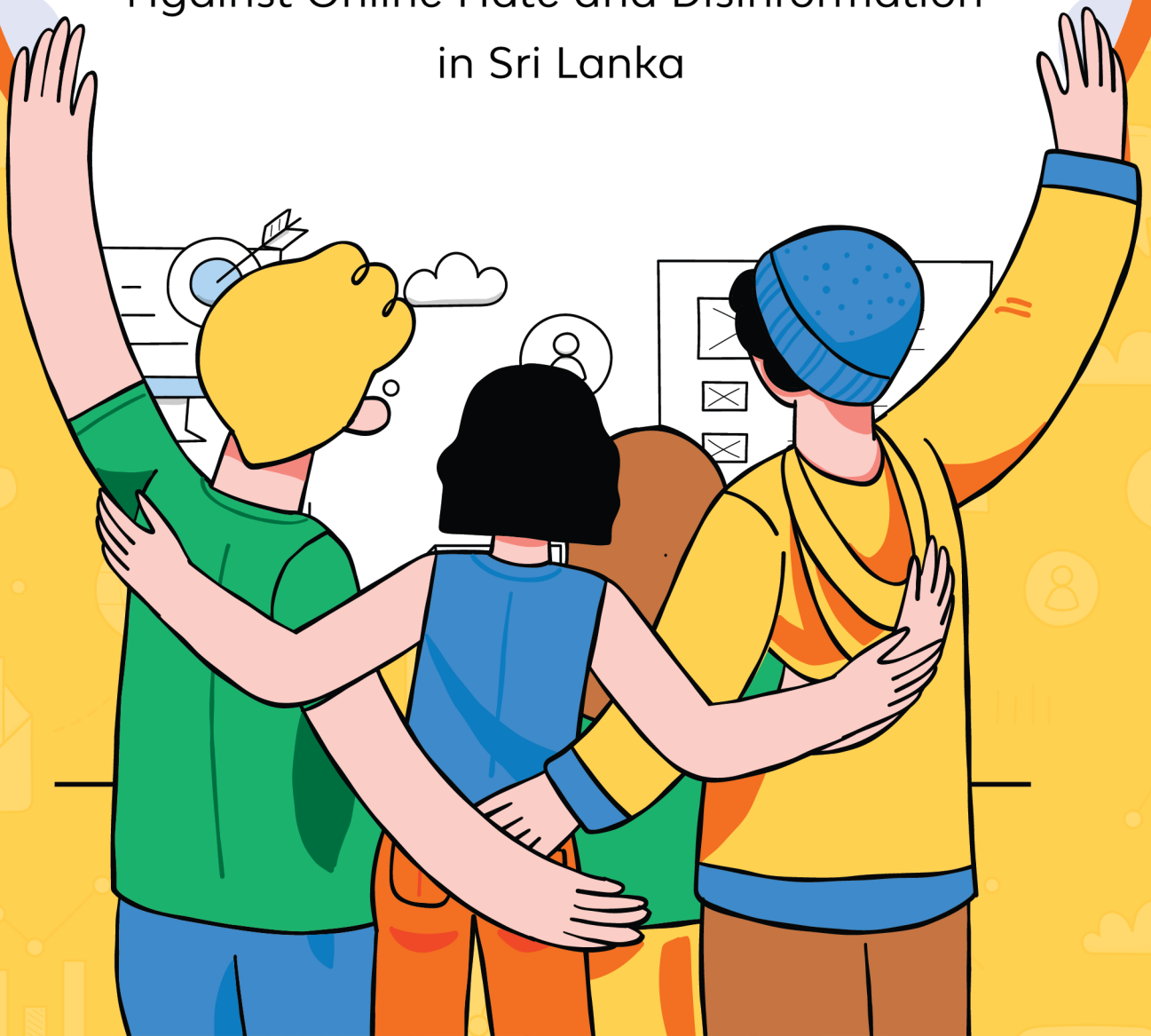


A Strategy for Effective

# Counterspeech

Against Online Hate and Disinformation  
in Sri Lanka



## **FIRST PUBLISHED IN AUGUST 2022**

### **PUBLISHED BY:**

National Christian Evangelical Alliance of Sri Lanka (NCEASL)

P.O. Box 113, Dehiwala, Sri Lanka.

Phone: +94 11 5 511358

Email: [info@nceasl.org](mailto:info@nceasl.org)

Web: <http://nceasl.org> and <https://www.minormatters.org>

**LEAD CONSULTANT AND WRITER:** Nalaka Gunawardene

**GENDER CONSULTANT:** Saritha Irugalbandara

### **PROJECT COORDINATION BY:**

NCEASL team comprising Yamini Ravindran, Sarah Arumugam,

Shalomi Daniel, and Sanjula Pietersz

**COVER DESIGN:** Shenal Jesudian

**PAGE LAYOUT AND DESIGN:** Shenal Jesudian

**PRINTED BY:** NAME TO COME

### **FUNDING SUPPORT FROM:**

United States Agency for International Development (USAID)



**DISCLAIMER:** The analysis and opinions expressed in this publication are those of the Consultants and are not necessarily those of the National Christian Evangelical Alliance of Sri Lanka, and do not commit the organisation.

## EXECUTIVE SUMMARY

The proliferation of mass media and online media during the first two decades of the 21st century has produced mixed results. It has certainly enhanced access to information and broadened the opportunities for learning, self-expression, networking and community mobilisation. Unfortunately, some individuals are abusing these same spaces and technologies for spreading hatred, misogyny, disinformation and violent extremism. The big challenge is how to optimise the digital benefits while minimising the digital risks in ways that do not unduly restrict or violate anyone's right to freedom of expression.

In tackling harmful speech and disinformation online, criminal law should not be the first resort, but the last. While laws are necessary to deal with extreme abuses, criminalisation of any speech risks state over-reach, selective law enforcement and censorship (by states or tech companies). For this reason, there is an urgent need to enhance non-legal responses by citizens, civil society groups, professionals, mass media and IT/ICT industries. Such responses include reforms in formal education and mass media sectors to enhance everyone's media and information literacy – so that there is more responsible expression and everyone becomes more discerning in consuming news, other information and views.

Such long-term interventions need to be accompanied by some short-term responses. One type of quick responses is counterspeech – the art of countering problematic speech creatively. This entails countering hate speech with expressions promoting harmony, tolerance, respect and more inclusive societies. Disinformation may be countered by fact-checking and by promoting public discourse based on facts, evidence and reason.

Because hate speech and disinformation have well-documented gender dimensions, all responses also need to be anchored in gender-sensitive language, with special attention to hate speech and disinformation that target girls, women and sexual or gender minorities.

While hate speech and disinformation have been circulating in Sri Lanka for decades, their volume and spreading speed have increased with rising levels of internet use (by 2021, roughly half the population was online). Youth (those below 35 years) are the predominant users of online and digital services. They also hold the key to responsible internet use.

Some producing and spreading of problematic content appear to be getting more organised and also devious in avoiding platform level monitoring and self-regulation by social media companies. This requires counterspeech efforts to also go beyond random reactions and become more strategic, better coordinated and sustained over longer periods.

*This Strategy, prepared through a consultative process, identifies some approaches and interventions in counterspeech in the context of Sri Lanka's socio-political realities. It is meant to help streamline and scale up various efforts already being pursued by different individuals or groups. The focus is on promoting strategic thinking and strategic communication in deciding when (and if) to respond to problematic speech, and how best to do so. Investing sufficient time in thinking through these early, important steps can help make subsequent material production and dissemination activities more effective.*

## **CONTENTS**

	Executive Summary
page 2	1. Introduction: Why this strategy?
page 8	2. Consultative process
page 12	3. Situation analysis: Dealing with problematic speech
page 18	4. Key concepts and challenges
page 32	5. Counterspeech
page 36	6. A Strategy for Effective Counterspeech
page 44	7. Strategies used by other counter-speakers
page 48	8. Counterspeech and social media platforms
page 52	9. Safeguarding the counter-speakers
page 56	10. Implementing Counterspeech: Some practical advice
page 62	11. After Strategising, Action!
page 66	12. Conclusion: The way forward





## 01. Introduction: Why this strategy?

“No one is born hating another person because of the colour of his skin, or his background, or his religion. People must learn to hate, and if they can learn to hate, they can be taught to love, for love comes more naturally to the human heart than its opposite.”

- Nelson Mandela, in Long Walk to Freedom

Rising levels of hate speech and disinformation are endangering inter-ethnic and inter-religious harmony in multicultural Sri Lanka. Such expressions – which are a misuse of the basic human right to freedom of expression – can also threaten the integrity of elections, endanger public health at times of pandemics, and undermine citizens’ trust in their government.

Hate speech and disinformation are not new phenomena: they have been around for centuries and are today found in both physical space as well as cyberspace. With rising levels of internet use, such problematic content is able to spread faster and wider.

This strategy is meant to support the Sri Lankan society’s on-going quest for interventions to counter hate speech, disinformation and misogynistic content. The scope of this document is limited to their online manifestations (on websites, social media and instant messaging services).



**The strategy is anchored in the conviction that laws and regulations alone cannot deal with these challenges which abuse the right to freedom of expression cherished by all open and democratic societies. In many countries, where governments legally criminalised hate speech and/or disinformation, such laws have been misused by governments to suppress political criticism and dissent.**

Therefore, more nuanced and multi-pronged interventions are needed.

On the one hand, it is necessary to minimise the reach and speed of truly harmful items of hate speech, disinformation and misogynistic content while allowing the freedom for all kinds of legitimate speech (including political criticism, satire, parodies, critiques of religious teachings and self-expressions by minority or marginalised groups).

On the other hand, because hate speech and disinformation threaten social cohesion, they should be countered in systematic and sustained ways.

One type of quick responses is counterspeech – the art of countering problematic speech creatively. This entails countering hate speech with

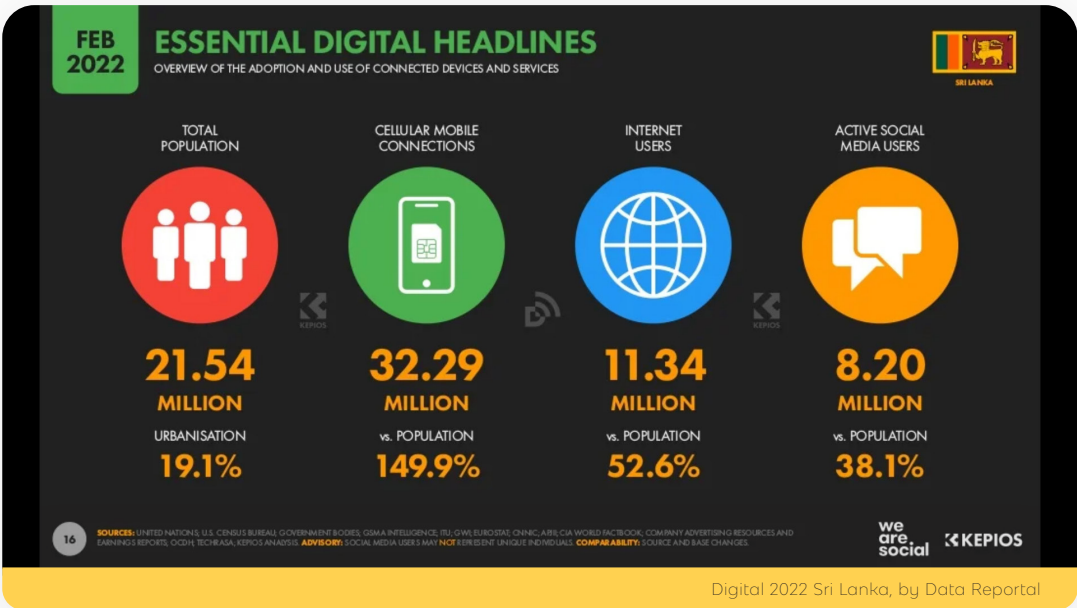


expressions promoting harmony, tolerance, respect and more inclusive societies. Disinformation may be countered by fact-checking and by promoting public discourse based on facts, evidence and reason.

Other non-legal responses -- such as strengthening independent media and reorienting formal education - are also available even though their results will take longer to emerge.

**The strategy identifies some approaches and interventions in counterspeech in the context of Sri Lanka’s socio-political and communications realities after a dozen years have passed since the civil war ended in 2009.** This post-war period has seen a deepening polarisation of society along ethno-religious and political lines despite many attempts aimed at reconciliation.

From around 2008 onward, there has also been a steady increase in the use of internet and digital devices (especially smartphones) in the country that, in turn, has enabled much easier and faster content generation and sharing - especially through social media.<sup>1</sup> While there are clear economic, educational and societal benefits of such technologies and services, there are also some negative effects.



1 There were 11.34 million internet users in Sri Lanka in January 2022, which worked out to 52.6% of the total population of 21.54 million. It was estimated that there were 8.2 million active social media users in the country by January 2022, or 38.1% of population. Source: <https://datareportal.com/reports/digital-2022-sri-lanka>

Paradoxically, the very online spaces and digital platforms that have enabled citizens to engage in public conversations supportive of democracy, human rights, disaster responses, humanitarian interventions and many other matters of public interest are being misused by some citizens to spread hatred and lies.

The negative effects have a clear gender dimension, too. Women and girls as well as sexual and gender minorities are disproportionately being targeted online. Sexual and gender-based violence (SGBV) is manifesting in various online forms, facilitating a rapidly expanding digital ecosystem that fosters deeply harmful, violent misogynistic ideology.



**Counterspeech is not a panacea that can counter or resolve all these formidable problems. But it is one tried-and-tested approach that is certainly worth considering and adopting along with other responses.**







## 02. Strategy preparation: Methodology and consultative process

The Minor Matters initiative of the National Christian Evangelical Alliance of Sri Lanka (NCEASL) facilitated the process of preparing this strategy as part of its on-going Digital Citizen campaign to raise awareness and insight on growing challenges in the digital world including hate speech, cyber bullying and disinformation.<sup>1</sup>

---

1 <https://www.nceasl.org/post/digital-citizen>

The campaign is dedicated to educating and equipping youth on digital citizenship. It has been engaging youth leaders (those between 18 and 35 years) from all ethno-religious backgrounds. Under this, NCEASL/Minor Matters has already published a Digital Citizenship Toolkit and a series of Digital Citizenship Comic Books.<sup>2</sup>

This strategy was prepared through a consultative process during the period from May 2021 to February 2022. Consultations were constrained by factors beyond anyone's control, i.e. the COVID-19 pandemic and economic crisis.

The process involved the following key steps:

- Internal discussions were held with NCEASL programme managers and other relevant staff to clarify their perspectives and objectives with the strategy.
- An informal consultative group was identified and were engaged in online discussions to elicit their views and suggestions. This group included social activists, socially-engaged lawyers, researchers on ethno-religious conflicts, and a media researcher.
- Several focus group discussions were held, using online platforms, to listen to the experiences, observations and suggestions of youth activists and social media activists.
- The draft strategy was shared for feedback from different ethno-religious groups, youth activists, peace builders, faith groups and those already engaged in counterspeech initiatives of their own.
- From November 2021 to February 2022, the draft strategy was used as the basis for conducting three online training workshops for selected groups of youth leaders (two workshops were in Sinhala and one was in Tamil). The participating youth discussed counterspeech interventions proposed in this strategy and gave other feedback - these have been used to finetune the strategy.

---

2 <https://www.minormatters.org/en/toolkit-guides>

- Youth leaders who completed training were encouraged to apply for a small grant for planning and carrying out a community level counterspeech training and advocacy work. Two groups received such grants and their reports after 3 months of community level activities (March to May 2022) have also been studied when finalising the strategy.







### 03.

## Situation analysis: Dealing with problematic speech

“The internet’s unparalleled facilitation of instantaneous, worldwide communication is a double-edged sword in terms of hateful speech. The internet not only makes it easier than ever to convey hateful messages; it also makes it easier than ever to rebut them. What’s more, the internet makes it easier to measure the extent and impact of counterspeech. Although the field is still young, there have been promising online counterspeech initiatives and studies of their efficacy.”

- Nadine Strossen, Professor of Constitutional Law at New York Law School and former national president of the American Civil Liberties Union (ACLU)

From xenophobia and racism to misogyny and homophobia, deep human prejudices have found expression for millennia. Similarly, falsehoods – including rumours, hoaxes, conspiracy theories and partisan content – have been in circulation for much of human history.

So, what is new?

The spread of internet use since the early 1990s and the rise of social media since the early 2000s have made it much easier and faster to disseminate harmful or misleading expressions. In this sense, the web has become not only a gigantic mirror of humanity's best and worst sentiments but also a space where ideas – good, bad and ugly – can quickly gain visibility and traction. Recent research in Europe suggests that there is a link between online hate speech and real-life acts of discrimination and violence.<sup>1</sup>

This reality presents a conundrum to modern societies: how to safeguard free speech while guarding against its abuses?

### Freedom of expression

covers any activity of seeking, receiving and imparting information or ideas, regardless of the medium used. The right to freedom of expression has been recognized as a basic human right by Article 19 of the Universal Declaration of Human Rights (UDHR, 1948)<sup>2</sup>

and is also guaranteed by Article 14 in the Constitution of Sri Lanka.<sup>3</sup>

**Article 19**

Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.

▶ Everyone has the right to their own opinion and thoughts, and no one can stop you.

However, freedom of expression is not an absolute right – some conditions do apply. International human rights laws (that Sri Lanka has ratified) narrowly define the allowable restrictions. Any official restrictions must pass a three-part test of legality (provided by a written law or regulation), legitimacy (protect or promote an aim considered legitimate in international law), and proportionality

1 <https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/hate-speech-and-violence>

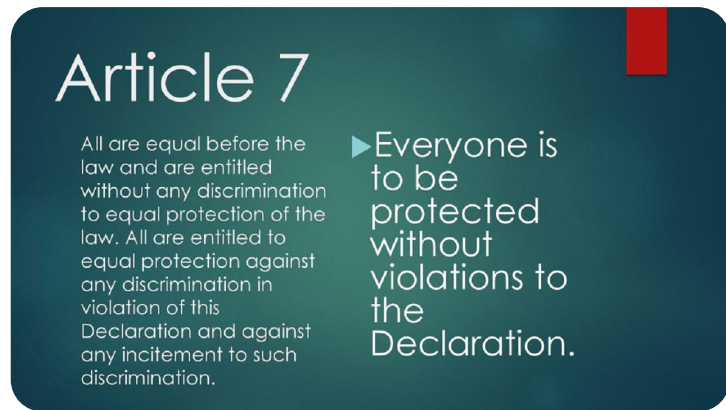
2 <https://www.un.org/en/about-us/universal-declaration-of-human-rights>

3 <https://www.parliament.lk/files/pdf/constitution.pdf>

(least restrictive means required).<sup>4</sup>

At the same time, **non-discrimination** is a central principle in all the major human rights treaties adopted over the past few decades: principles of equality and non-discrimination help form the rule of law.<sup>5</sup> All human beings have the right to equality, and the right not to

be discriminated against on the basis of their race, ethnicity, nationality, religious belief, sex, disability, educational level, social background or any other factor.



**These two are foundational rights essential for the enjoyment and protection of all human rights. In fact, they are mutually supporting and reinforcing. As human rights experts point out, it is only when coordinated and focused action is taken to promote both freedom of expression and equality that either can be fully realised.**

Sometimes, though, free expression taken to extremes can threaten the right to equality. The challenge is to balance the two rights. In other words, when and how to limit the free speech of individuals from spreading ideas that can generate deep resentment, suspicion or hatred against specified persons or groups, potentially discriminating against them?

It is not easy to agree on where free speech ends and hate speech begins. As described in 4.1 below, there is no universally agreed definition of hate speech. The task is made harder by the fact that freedom of expression covers not only "information" or "ideas" that are favourably received, but also those that may offend, shock or disturb some (see also Box 1).

---

4 A helpful explanation is found at: <http://www.law-democracy.org/live/wp-content/uploads/2015/02/foe-briefingnotes-2.pdf>

5 UDHR's Article 7 says: "All are equal before the law and are entitled without any discrimination to equal protection of the law..." See a non-legal explanation at: <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=23885&LangID=E>

As such, **legal limits to harmful speech need to be set and enforced very carefully.**

Those advocating action against hate speech and/or disinformation often overestimate what laws and regulations can realistically accomplish. When harmful speech is criminalised, what is legally allowed or not would depend on official definitions and interpretations – which leaves room for governments or tech companies to exclude legitimate criticism and dissent.

In contrast, many persons promoting robust public discourse tend to underestimate the power of non-legal responses, not fully realising how much their own voices of reason and other counterspeech interventions can do in tackling hatred, deception and misogyny.







## 04. Key concepts and challenges

In this section, we briefly explain the key concepts to enable a clearer understanding of the issues and challenges involved.

## 4.1 HATE SPEECH

There is no universally accepted definition of hate speech in international human rights law. The characterization of what is 'hateful' is subjective, often controversial and debated. Hate speech lies in a complex nexus with freedom of expression, individual, group and minority rights, as well as concepts of dignity, liberty and equality.<sup>1</sup>

Here are three working definitions found in key international documents that can help us to understand the concept.

Source document	How it has defined hate speech
<i>United Nations Strategy and Plan of Action on Hate Speech</i> , launched at the UN Headquarters in June 2019. ( <a href="#">Link</a> )	"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 factor."
<i>Countering Online Hate Speech</i> United Nations Educational, Scientific and Cultural Organisation (UNESCO, 2015) ( <a href="#">Link</a> )	"Hate speech refers to expressions that advocate incitement to harm (particularly, discrimination, hostility or violence) based upon the target's being identified with a certain social or demographic group. It may include, but is not limited to, speech that advocates, threatens, or encourages violent acts."

---

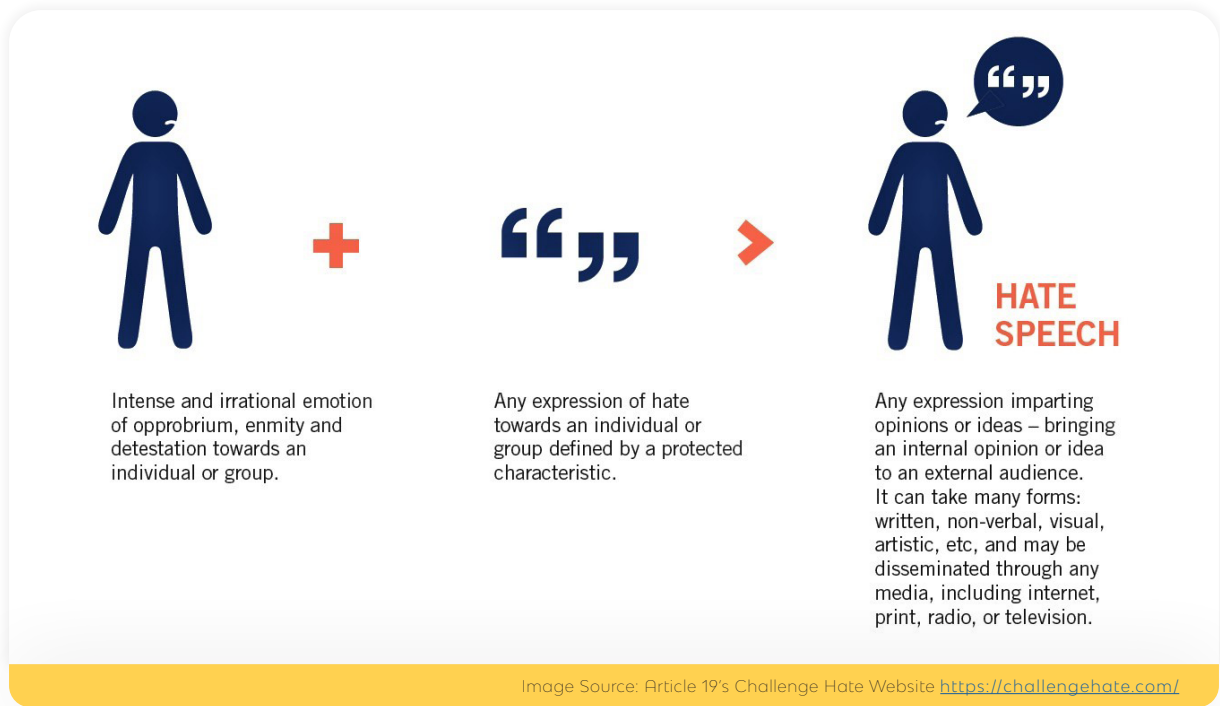
<sup>1</sup> Countering Online Hate Speech. UNESCO, 2015. <https://en.unesco.org/news/unesco-launches-countering-online-hate-speech-publication>



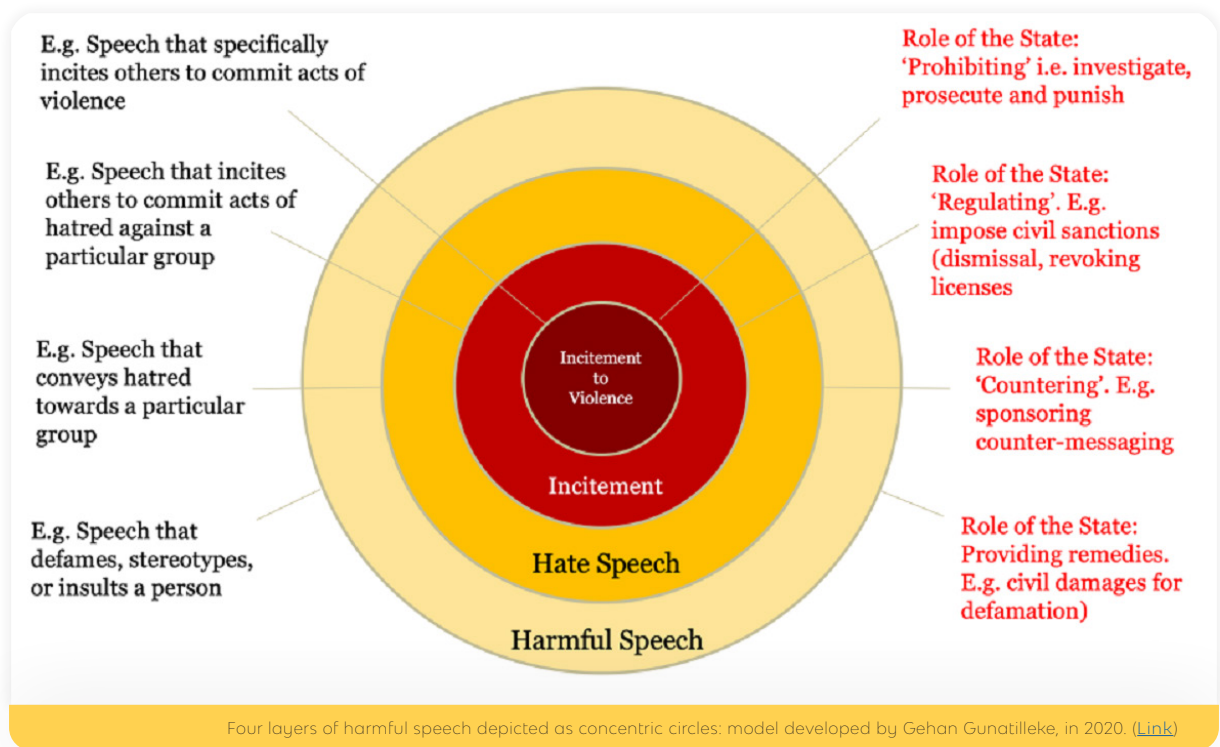
Source document	How it has defined hate speech
<p><i>International Covenant on Civil and Political Rights (ICCPR)</i>. Article 20</p> <p>This is a key international human rights treaty that Sri Lanka ratified in June 1980. (<a href="#">Link</a>)</p>	<p>“Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law”</p>

## IMPORTANT:

- Whatever definition is used, hate speech can only be generated against human beings (individuals, groups or entire communities) and not against ideas, teachings or institutions. Any criticism of ideas or teachings, including political doctrines and religious faiths, does not come within the scope of hate speech (see also Box 1).
- Hate speech is different from general abuse or insults. For hate speech to happen, specified human beings should be the target of hate, and it should be on the basis of one or more ‘protected characteristics’. Guidance for these protected characteristics comes from the Universal Declaration of Human Rights of 1948, Article 2 (Freedom from Discrimination) which says that everyone is entitled to all the freedoms listed “without distinction of any kind such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status.” The last words – “other status” – have frequently been cited to expand the list of people specifically protected.<sup>2</sup>
- In recent times, some more such protected characteristics have been recognised, including indigenous origin or identity, disability, migrant or refugee status, sexual orientation, gender identity or intersex status.



Another way to look at hate speech is to position it within a spectrum of problematic speech, each capable of causing harm but at different levels. Gehan Gunatilleke, a Sri Lankan legal scholar, has developed a model where he depicts four layers of harmful speech as concentric circles.



The following explanation is adapted from an article<sup>3</sup> he wrote in December 2020:

- **The first (outermost) layer** represents all types of harmful speech, including speech that defames, stereotypes or insults others. Falsely accusing someone of a crime, or insulting a person's faith, can fall into this category.
- **The second layer** represents 'hate speech', which is a subset of harmful speech. Speech that conveys hatred towards a particular ethnic, racial, religious, gender or other group falls into this category. Of course, the term 'hatred' needs to be defined carefully. It generally means 'intense and irrational emotions of opprobrium, enmity and detestation towards the target group'. For example, calling a community inherently 'evil' constitutes hate speech.
- **The third layer** represents 'incitement', which is a subset of hate speech. While the broader category of hate speech involves conveying hatred, incitement additionally involves compelling others to act in a hateful manner. Such acts include discriminating the target group. For example, compelling others to boycott Muslim-owned businesses constitutes incitement to discrimination.
- **The final (innermost) layer** represents incitement to violence, which is a subset of incitement. This is the worst form of harmful speech. Incitement to violence contains two elements: first, it involves the act of compelling others to perpetrate violence against a specific person or group; second, it involves a context where such violence is 'imminent', i.e. there is a real risk of violence happening because of hateful words or other communication.

**Gehan Gunatilleke says that the state's ultimate authority to prosecute and punish individuals must be confined to a narrowly defined domain of speech – i.e. incitement to violence. Other forms of harmful speech need to be responded through civil remedies, counter-messaging and certain limited forms of regulation.**

---

3 <https://groundviews.org/2020/12/17/counering-harmful-speech-why-trust-the-state/>

Box 1

## WHAT EXPRESSIONS ARE NOT CONSIDERED 'HATE SPEECH'?

It is also very important to understand what kind of speech does not amount to hate speech.

- **Deeply offensive expressions not hate speech:** International freedom of expression standards allow expressions that may be offensive, disturbing or shocking to some persons: speech limitations cannot be made solely on the basis of "offence" caused to an individual or group.
- **Blasphemy or "defamation of religion" is not hate speech:** International human rights law protects people, not concepts, ideologies or belief systems -- including religions. It distinguishes between ideas or beliefs attached to individuals, and does not protect religions or beliefs per se from adverse comments.
- **Criticism of the State or governments is not hate speech:** International standards do not permit protection of "the state" or its symbols from insults or criticism. These entities cannot be the target of 'hate speech', because they are not people.

Please note, however, that there could be laws that make some of the above expressions illegal in certain countries without them coming within the scope of hate speech. An example is blasphemy that is explicitly outlawed in some countries.

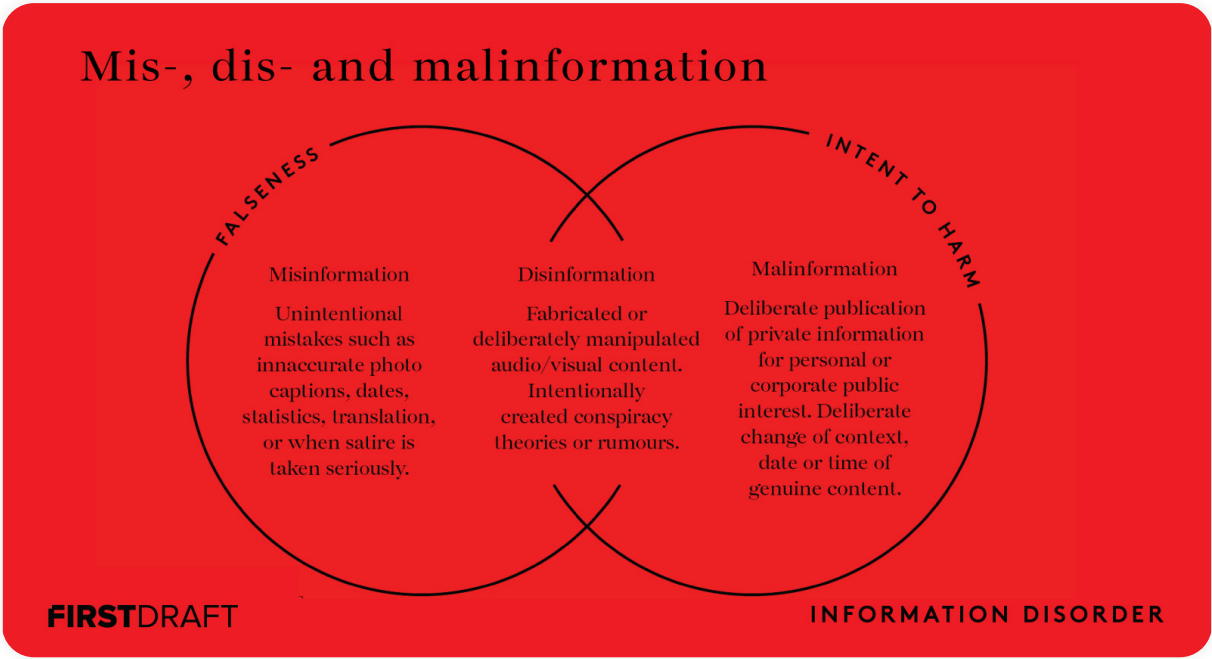
Source: <https://www.article19.org/resources/hate-speech-explained-a-summary/>

## 4.2 DISINFORMATION AND MISINFORMATION

Information manipulation is not at all new: it has been happening for centuries. Advances in communications technology and changes in the media environment have made it easier today to spread falsehoods faster and wider through the web, in the media, as well as inter-personally.

Those studying this phenomenon use the following definitions (as summed up by UNESCO):<sup>4</sup>

- Dis-information is false or manipulated information that is created and disseminated with the deliberate intention of misleading recipients. These can harm a person, entity, process or an entire nation.
- Mis-information is false or manipulated information that is shared without the intent to cause harm, usually by persons who believe it to be true.
- Mal-information is information based on reality but being shared with the clear intent to cause harm and/or to benefit the disseminator (e.g. making somebody's private information or images public without consent).

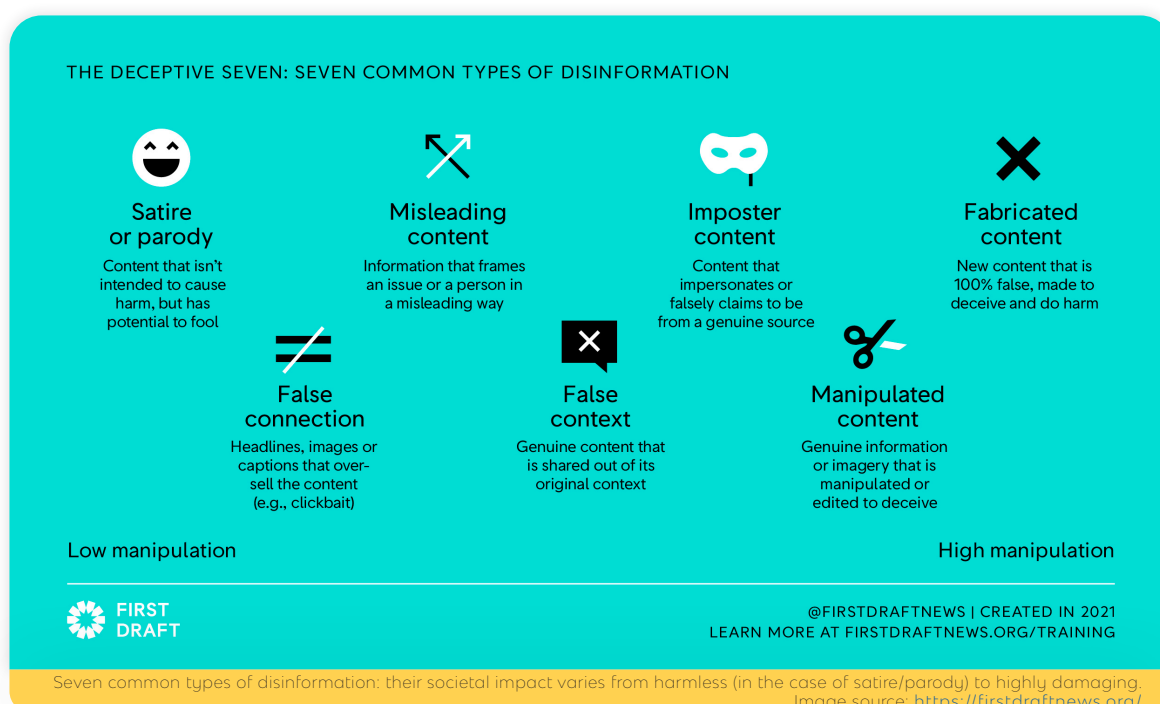


First Draft, a global non-profit organisation working on issues related to information disorder, has captured the inter-relationships of these three in this image. In their analysis, what matters most is identifying the intention to cause harm through information manipulation.<sup>5</sup>

Please also note:

4 <https://en.unesco.org/fightfakenews>  
5 <https://firstdraftnews.org/articles/information-disorder-the-techniques-we-saw-in-2016-have-evolved/>

- Falsehoods can come in many shapes and forms. Their societal impact varies from harmless (in the case of satire/parody) to moderately harmful or highly damaging.
- First Draft has identified seven common types of disinformation: satire or parody (harmless entertainment); false connection; misleading content; false context; imposter content; manipulated content; and fabricated content. These are briefly explained in the infographic.
- The term ‘Fake News’ is no longer used in serious discussions about this subject, because it does not fully capture the full range of false content. The preferred generic term is Disinformation (and its impact on society is now known as ‘Information Disorder’).



### 4.3 ONLINE SEXUAL AND GENDER-BASED VIOLENCE (SGBV)

There exists a significant gender disparity in who is targeted for harassment and violence online. While any web user can encounter cyber harassment or violence, data shows that women, girls and sexual and gender minorities are disproportionately targeted.

Sexual and gender-based violence (SGBV) is any act of sexual, physical, psychological, and mental abuse that is perpetrated against a person on the

basis of their sex, gender (including expression and identity) or sexual orientation.

The term 'online SGBV' should be used with caution. Some might think that SGBV on the internet is distinct or different from 'offline' forms of SGBV. Indeed, online SGBV is -- and should be seen as -- an extension of offline violence, and not as a separate issue. The internet is yet another modality, and the gendered harms both follow from and inform violence that happens offline.

"The internet presents a double-edged sword for women. On the one hand, it provides vital spaces for women seeking expression and opportunity. On the other, it's increasingly a vector for abusers targeting women. The COVID-19 pandemic worsened this situation as women spend more time online, increasing their exposure to threats," wrote the Economist Intelligence Unit (EIU), an independent research agency, when introducing findings of their 2020 study 'Measuring the prevalence of online violence against women'.<sup>6</sup>

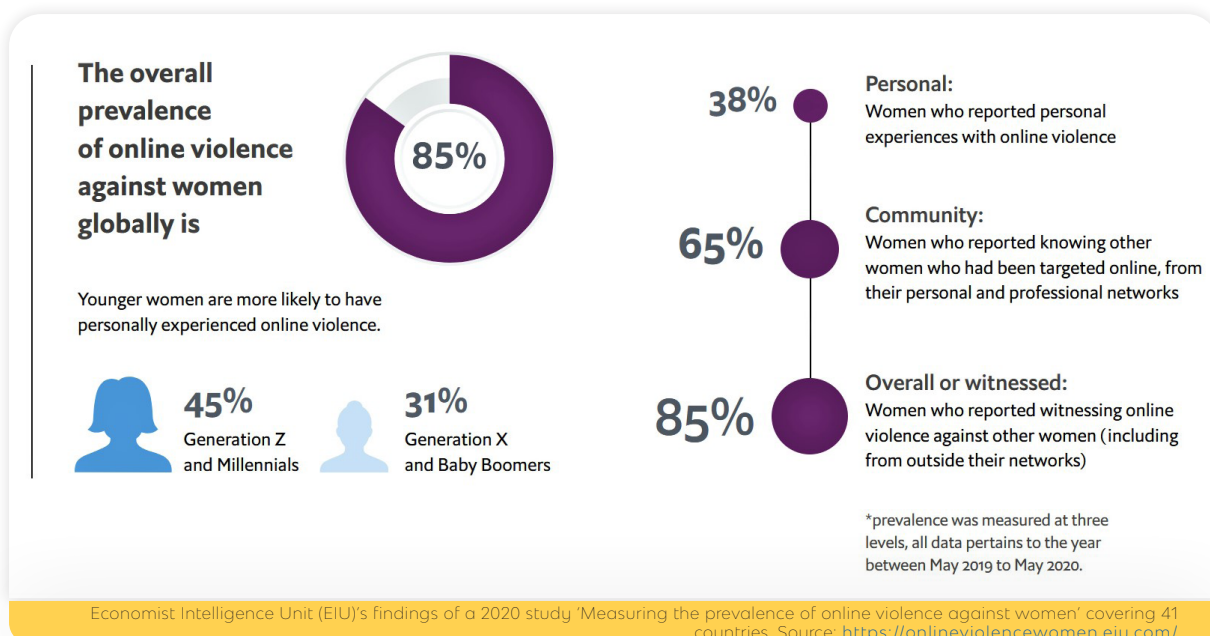
EIU's survey of women from 51 countries around the world found nine threat tactics experienced by high percentages of women online:

- Cyber harassment: Repeated behaviour using textual or graphical content to frighten and undermine women's self-esteem
- Misinformation and defamation: Spreading rumours and slander to discredit or damage a woman's reputation (often based on sexist tropes and gendered stereotypes)
- Hate speech: Misogynistic or hateful language designed to attack or humiliate
- Impersonation: Creating a false online presence in someone else's name
- Hacking and stalking: Intercepting communications and data; targets women across social media accounts and through location tracking
- Video- and image- based abuse: Recording, gathering, and sharing of (private) images without consent

---

6 <https://onlineviolencewomen.eiu.com/>

- Doxing: Publicly revealing private or identity-revealing information without consent
- Violent threats: Threats of physical or sexual harm sent through online channels



Especially targeted for online violence are women who hold elected office, and those in senior management positions, or are visible in public life as result of their professional pursuits – such as politicians, journalists, lawyers, social activists, performing artists or sportspersons. Often this harassment involves some form of sexist or misogynistic rhetoric, or sexual objectification, to undermine a woman’s suitability or credibility to hold a position of power on the basis of their sex and/ or gender.

These trends have been documented in Sri Lanka too. A 2019 study conducted by researchers from three advocacy groups -- Centre for Policy Alternatives (CPA), Ghosha and Hashtag Generation – offered some insights on how online violence against women happen in the Sri Lankan context. The study looked at how women are discussed on selected pages on Facebook, the most popular social media platform in Sri Lanka. It found a clear pattern of sexist speech that objectified, harassed or otherwise targeted women and members of the lesbian, bisexual and trans-gender communities.<sup>7</sup>

7 OPINIONS, B\*TCH: Technology-based Violence Against Women in Sri Lanka. CPA, Ghosha and Hashtag Generation, 2019. <https://groundviews.org/2019/06/27/opinions-btch-technology-based-violence-against-women-in-sri-lanka/>



## 4.4 GENDERED AND SEXUALISED DISINFORMATION

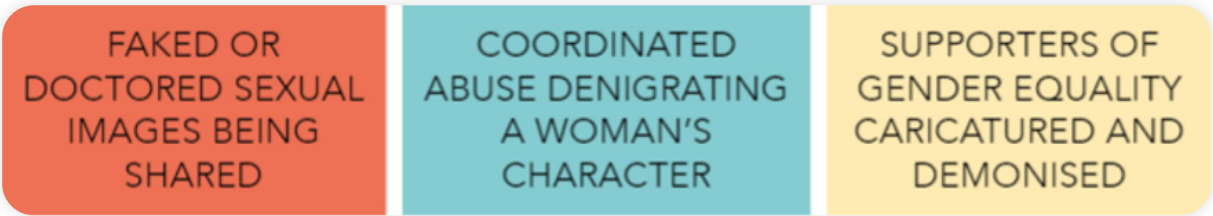
Recent studies in several countries have shown that one of the key identity factors on which disinformation is concentrated is gender. This means disinformation is targeted more at women than men, and especially at women who are playing a public role. Disinformation is also created and spread against feminist struggles and gendered discourse.

This trend is known as **gendered and sexualized disinformation**. It is aimed at women in the mainstream media, women human rights defenders, female politicians, female entrepreneurs, and many women who use social media for personal or professional reasons. It has also been found that abusers often deploy both sex- and race-based narratives, so the harassment and abuse faced by women of colour or women belonging to ethnic and religious minorities is heavily racialised and more acute.

The European Union’s Disinfo Labs, in a recent analysis (December 2020), defined gendered disinformation as the “dissemination of false or misleading information attacking women, basing the attack on their identity as women.”<sup>8</sup>

Another study, by the US-based Wilson Centre published in January 2021, looked at gendered and sexualized disinformation on women in public life, as well as its impacts on national security and democratic participation. That study defined the phenomenon as “a subset of online gendered abuse that uses false or misleading gender and sex-based narratives against women, often with some degree of coordination, aimed at deterring women from participating in the public sphere. It combines three defining characteristics of online disinformation: falsity, malign intent, and coordination.”<sup>9</sup>

Examples of gendered disinformation identified by the EU Disinfo Labs include:



8 Misogyny and Misinformation: An analysis of gendered disinformation tactics during the COVID-19 pandemic. EU Disinfo Labs, 2020. <https://www.disinfo.eu/publications/misogyny-and-misinformation%3A-an-analysis-of-gendered-disinformation-tactics-during-the-covid-19-pandemic/>

9 Malign Creativity: How Gender, Sex, and Lies are Weaponized Against Women Online. Wilson Centre, 2021. <https://www.wilsoncenter.org/publication/malign-creativity-how-gender-sex-and-lies-are-weaponized-against-women-online>

**Gendered and sexualized disinformation that is spread online can have serious impacts offline. It can deepen negative perceptions of women in society. It can also undermine women's credibility in occupying positions of power and discourage more women from participating in the public debate.**

Gendered and sexualized disinformation is a phenomenon distinct from broad-based gendered abuse [online] and should be defined as such to allow social media platforms to develop effective responses, the Wilson Centre report recommended.







## 05. Counterspeech

Counterspeech broadly means any direct response to harmful speech that is generated and spread to undermine such harmful speech. In other words, counterspeech is the careful use of communication to mitigate the negative impacts of harmful communication.<sup>1</sup>

---

<sup>1</sup> Put simply, “communication” is the act of sharing, receiving, and interpretation of messages, while “communications” is the means by which those messages get shared.

There are two types of counterspeech:

- **organized counter-messaging campaigns**, usually by an organisation or social movement
- **spontaneous responses** coming from many individuals, usually reacting to an incident

Counterspeech content and their delivery can be varied, including:

- incorporating relevant concepts and messages within the formal education system
- non-formal public education efforts through civil society, mass media and online
- public campaigns that use humour, satire and entertainment formats
- direct engagement with producers or disseminators of hate speech (if they are known)

**In relation to hate speech, counterspeech can cover** a broad range of expression, including:

- communications that directly refutes or otherwise responds to a specific hate speech
- proactive public educational initiatives promote equality, human rights and dignity
- alternative narratives advocating pluralistic and inclusive societies

**In relation to disinformation, counterspeech can cover** various responses, including:

- increasing the supply of accurate information in the news, and in other public forums
- supporting post-publication fact-checking services to verify assertions and statistics cited
- promoting a culture of using evidence for personal and professional decisions

The Dangerous Speech Project (which studies all forms of human expression that can inspire violence between groups of people), has been studying and promoting counterspeech for many years. They have defined successful counterspeech in two main ways.

- Firstly, counterspeech (text or visual) that has a **favourable impact on the original (hateful) communicator/s**, usually indicated by an apology or recanting, or the deletion of the original content online.
- Secondly, counterspeech **positively affects the audience** of a counterspeech conversation in how they look at the issue or topic involved.

### Are counterspeech efforts effective?

How do we know that counterspeech actually works? Users of social media and other online sources are exposed to many kinds of content, so it is very difficult to show a direct cause-and-effect kind of impact of specific counterspeech efforts.

However, here are some indicators of influence or impact of counterspeech:

- long online conversations that remain civil without users turning to insults or invective
- large numbers of social media or web users start following and/or sharing counterspeech
- emergence of new counter-speakers who are inspired or encouraged



“Since wars begin in the minds of men, it is in the minds of men that the defences of peace must be constructed” says the Preamble to the Constitution of the United Nations Educational, Scientific and Cultural Organization (UNESCO) founded in 1948. This important statement has been revised in recent years for it to read ‘men and women’.

[Source](#)







## 06. A Strategy for Effective Counterspeech

A strategy is a plan of action designed to achieve a long-term or overall aim or goal. What is offered below is a vision, objectives and interventions that make up a strategy for counterspeech in the Sri Lankan context.

**Note:** This document's scope is limited to non-legal responses, and therefore, various legal and regulatory responses are not explored here. For that reason, self-regulation at the level of social media platforms (which involves community standards, reporting problematic content, and seeking their removal or demotion) is not discussed in detail either.

## 6.1 STRATEGIC VISION

This is the long-term vision, i.e. what we wish to see achieved in the long term:

**Sri Lanka's multicultural society becomes more inclusive, tolerant and appreciative of its diversity through more open, truthful and respectful communications**

## 6.2 STRATEGIC OBJECTIVES

These are the objectives, or broad pathways, that can be used in pursuit of the vision:

- Enhancing, through monitoring and analysis, the current understanding of hate speech, disinformation and other key abuses of free speech
- Promoting greater public discourse on the value and efficacy of non-legal responses to hate speech, disinformation and other key abuses of free speech
- Exploring opportunities for engaging known originators of hate speech and disinformation with a view to persuading them to end such practices
- Building the capacity of individuals and groups who are already engaged in counterspeech efforts by improving their skills in messaging and campaigning
- Networking those engaged in counterspeech efforts (in respect of both hate speech and disinformation) to enable peer-to-peer learning, mutual support and collective advocacy

- Scaling up and, where possible coordinating, the currently scattered counterspeech efforts to ensure a bigger reach, better audience engagement and greater impact
- Documenting counterspeech efforts by state, professional, civil society, mass media, youth and other stakeholder groups, with a view to evaluating their reach and influence
- Safeguarding those engaging in counterspeech, especially when they counter well-organised political campaigns, and ensuring they have access to psycho-social support

### 6.3 STRATEGIC INTERVENTIONS AND ENGAGEMENTS

This being a strategy, it can only suggest different kinds of interventions and engagements, and not list out specific activities. It is extremely important to spend some time reflecting on these points before rushing to design any counterspeech campaigns or produce content.

**Note: As suggested below, sometimes the best response is not to respond in any manner (but to simply monitor and remain vigilant). Reacting too much, or too soon, can inadvertently give visibility and legitimacy to those engaged in harmful speech.**

A. Countering Hate Speech through counterspeech	
Intervention	Comment/caution
A1. <i>Watch in silence (do nothing):</i> Monitor and document a specific hate speech activity but not react or engage in any manner (this can be considered if the scale and reach are self-limited)	By ignoring scattered, low-level or sporadic hate speech activity, we avoid giving the originators/ peddlers a higher profile
A2. <i>Reach out and reason:</i> Identify the key originators of hate speech where possible, reach out and engage them in public or private conversations to persuade them to suspend/end discriminatory speech/campaigns	This can be potentially hazardous and should only be attempted by well-organised counter-speakers

## A. Countering Hate Speech through counterspeech

Intervention	Comment/caution
<p>A3. <i>Direct refuting</i>: When a hate campaign is based on a clear fallacy or distortion [Example: misinterpretations of demographic/census data claiming an 'ethnic minority could become majority'], go public with the real facts, backed by hard evidence and topic experts where possible. [In the above example, capturing and amplifying the expert analysis of demographic specialists can debunk racial misinterpretations.]</p>	<p>Ensure that facts are verified, expert opinion is unequivocal, and messaging is clear. This is not advisable for historical arguments: much of history is open to many interpretations.</p> <p>Caution: Never malign, ridicule or assign labels to hate speech peddlers.</p>
<p>A4. <i>Indirect countering</i>: Without directly contradicting or clarifying a hateful expression/campaign, generate public discussion around the same topic, bringing in many voices and perspectives – making people think.</p> <p>[Example: if a hate expression vilifies LGBTQI communities, create conversations or discussions where their status, problems and aspirations are explored, thus demystifying the topic and generating solidarity]</p>	<p>Key to success here is to allow the widest possible public discussion.</p> <p>Do not try to promote a single narrative. Also avoid getting into arguments on morality or political ideology – in both these areas, it is very hard to reach any consensus!</p>

## A. Countering Hate Speech through counterspeech

Intervention	Comment/caution
<p>A5. <i>Alternative narratives</i>: Without even indirectly referring to any prevailing polarisation, prejudices or hate, we start and sustain conversations about a more inclusive society or ideal future scenarios.</p> <p>[Example: Discussing how to achieve a common Sri Lankan identity where “Sri Lankan” is the shared identity while the diversity of 19 ethnic groups that share the island is the basis of strength for them all.]<sup>1</sup></p>	<p>This kind of communication may seem idealistic or even utopian when there is so much division in society. But such idealism is very much needed to get people to reflect beyond current realities and imagine better futures.</p>

## B. Countering Disinformation through counterspeech

Note: Countering disinformation has been the subject of much international research: a substantial amount of strategising has been done in recent years. Counter-disinformation has now evolved into a specialised practice area. What is listed below are only a few strategies that may be considered by civil society groups – some may not have the specialised expertise or skills needed for counter-disinformation operations at the scale and speed required.

Intervention	Comment/caution
<p>B1. <i>Watch in silence (do nothing)</i>: Monitor and document disinformation items to understand their trends and patterns, but not react to them or engage the producers in any manner.</p>	<p>Some disinformation efforts have an inherently limited reach; their visibility can be inadvertently increased if a high profile response is made.</p>
<p>B2. <i>Reach out and reason</i>: Identify the key originators of disinformation if and where possible, and engage in public or private conversations to persuade them to cease their efforts.</p>	<p>This can be potentially hazardous and should only be attempted by well-organised counter-speakers</p>

<sup>1</sup> An authentic reference for this is an official book titled “People of Sri Lanka” published in 2017 by the then Ministry of National Coexistence, Dialogue and Official Languages. <https://bit.ly/3ykLkj9>

## B. Countering Disinformation through counterspeech

Note: Countering disinformation has been the subject of much international research: a substantial amount of strategising has been done in recent years. Counter-disinformation has now evolved into a specialised practice area. What is listed below are only a few strategies that may be considered by civil society groups – some may not have the specialised expertise or skills needed for counter-disinformation operations at the scale and speed required.

Intervention	Comment/caution
<p>B3. <i>Direct refuting</i>: This is done by fact-checking services where statements reported in the media or uttered by key public figures are verified against authentic information sources – and outcome publicised. Fact-checks can reveal a given statement as true, misleading, false or blatantly false. By mid 2021, Sri Lanka had half a dozen fact-checking services operated by thinktanks, media companies or civil society.</p>	<p>Fact-checking is a specialised area of work requiring expertise, time and resources. When publishing findings of any fact-check, ensure total accuracy or fact-checkers' credibility can get damaged. It is much easier to share fact-checks done by dedicated fact-checking services to increase reach.</p>
<p>B4. <i>Indirect countering</i>: Without directly contradicting or questioning any false/misleading content, generate public discussion around the same topic, bringing out authentic information and expert viewpoints to clarify issues involved. [Example: Anti-vaccine or vaccine hesitancy sentiments are being promoted by some who hold deeply entrenched views. Instead of debunking their claims, undertake non-technical public communication about the benefits of vaccination in response to COVID-19 and other communicable diseases.]</p>	<p>It is vital to get credible, articulate subject experts for effective public communication. Encourage questions and discussion without pushing any single dominant viewpoint. Do not ridicule people's misconceptions; instead gently clarify and persuade. Also avoid getting into prolonged public arguments with anyone with entrenched or dogmatic views.</p>

## B. Countering Disinformation through counterspeech

Note: Countering disinformation has been the subject of much international research: a substantial amount of strategising has been done in recent years. Counter-disinformation has now evolved into a specialised practice area. What is listed below are only a few strategies that may be considered by civil society groups – some may not have the specialised expertise or skills needed for counter-disinformation operations at the scale and speed required.

Intervention	Comment/caution
B5. <i>Alternative narratives:</i> This involves promoting greater use of, and respect for, facts in our personal and professional lives – demanding evidence for claims and assertions. Also, encourage more people to be sceptical and question what is said by politicians, journalists, academics and others. <sup>2</sup>	There is often a gap between perceptions and reality, which is where latest data, analyses and research can help. Be careful not to use outdated data as our world keeps changing!

---

<sup>2</sup> For an exploration of fact-checking and the demand for factual evidence in the Sri Lankan context, see: <https://www.slideshare.net/NalakaG/your-facts-or-my-facts-factchecking-in-a-posttruth-world>







## 07. Strategies used by other counter-speakers

In a paper written in 2016, American journalist and counterspeech promoter Susan Benesch has shared a few counterspeech strategies to hate speech that have worked favourably on the social media platform of Twitter. The paper cautions, however, that these approaches may not work for all cultures or all online platforms.<sup>1</sup>

---

<sup>1</sup> Considerations for Successful Counterspeech, by Susan Benesch, Derek Ruths, Kelly P Dillon, Haji Mohammad Saleem and Lucas Wright. Guide written for the Kanishka Project of Public Safety Canada, 2016. <https://dangerousspeech.org/considerations-for-successful-counterspeech/>

- **Warning of Consequences:** Counter-speakers often warn of the possible consequences of speaking hatefully on a public platform like Twitter, and in many cases, this seems to have been effective at getting the [hate] speaker to delete the original posts.
- **Shaming and Labelling:** Some counterspeech labels original content as hateful, racist, bigoted, misogynist, etc. Since these labels carry such a shameful stigma for many people, hate speakers who do not consider themselves as racists, for example, quickly delete or amend their posts. [This may not work with everybody.]
- **Empathy and Affiliation:** Changing the tone of a hateful conversation is an effective way of ending the exchange. It may prevent the escalation of the hateful rhetoric being used in the present moment, even though its long term benefits are not clear.
- **Using Humour and Satire:** Sometimes (but not always) humorous counterspeech can shift the dynamics of communication, defuse tensions and draw more attention to a counter message than it would otherwise. Humour comes in many forms, including caricatures/cartoons, satire, spoofs and sarcasm. Humour should be used carefully, ideally to soften messages that could otherwise seem too harsh or aggressive.
- **Using Images:** Images are often more persuasive than just text. Knowing this, many counter-speakers include them in the form of memes, graphics, photographs, animated gifs and videos when responding to hateful or dangerous speech. Images can often overcome cultural and linguistic boundaries, which can allow counterspeech to spread virally. Visuals can also send people along emotive pathways in their minds while textual or verbal material leaves them in more rational, logical and linear pathways.

In the same paper, the author listed some strategies that are often ineffective. In some cases, these may even be counterproductive or harmful, they cautioned:

- **Hostile or Aggressive Tone, Insults:** These can backfire and lead an escalation of hate speech. Naming and shaming are to be done only using polite language.

- **Harassment and Silencing:** Some strategies for responding to hateful speech cross the line between counterspeech and harassment. Some respond to speech they disapprove of with threats. This should never be done.
- **Exposing falsehoods or false logic:** Corrections that demean, insult or threaten an original speaker's worldview can lead him/her/them to become even more entrenched. As such, this should only be done very carefully. General clarifications are fine.





## 08. Counterspeech and social media platforms

Social media is increasingly used by those engaged in harmful speech, but **the same social media platforms can also be used for counterspeech.**

How social media companies deal with harmful speech, and how they encourage or support counterspeech efforts, is an important factor in this regard.

### 8.1 PLATFORM LEVEL SELF-REGULATION FRAMEWORKS

Platform level self-regulation is one approach to tackling problems arising from harmful online content and behaviour. In this, all users are required to abide by a set of rules that they agree to at the time of opening an account on that platform. There is content monitoring by the platform, and also a mechanism for users to report to the platform about any content violating the rules. Violators initially receive warnings, followed by penalties (usually a suspension of their account for several hours or days); repeat violations could lead to accounts being blocked or terminated.

Here are links to rules of some key social media platforms:

Platform	Self-regulatory framework and link
Facebook	Community Standards (available in Sinhala and Tamil also) ( <a href="#">Link</a> )
Instagram	Community Guidelines ( <a href="#">Link</a> )
YouTube	Community Guidelines ( <a href="#">Link</a> )
Twitter	Rules and Policies ( <a href="#">Link</a> )
TikTok	Community Guidelines ( <a href="#">Link</a> )

Given the massive scales involved – with billions of new content being generated every day and night – content monitoring and enforcing these rules is a huge challenge. Some decisions by platforms to regulate specific items of content have triggered heated arguments on the limits of free speech online.

## 8.2 FACEBOOK'S CONTENT MODERATION AND COUNTERSPEECH PROMOTION

As the world's largest social media platform, Facebook has recognised counterspeech as "a more proactive approach to countering extremism and hate speech than simply deleting extremist posts". Monika Bickert, Facebook's head of global policy management, noted in March 2017, social media networks need counterspeech, a crowd-sourced response to extremism. "The best remedy is good speech that gets people thinking and challenging ideologies. We focus on trying to amplify some of the voices to counter violent narratives."<sup>1</sup>

Facebook's view of counterspeech is captured in a 2016 study: "Counter-speech is a common, crowd-sourced response to extremism or hateful content. Extreme posts are often met with disagreement, derision, and counter-campaigns. Compared with simply deleting controversial content, combating extremism in this way has some advantages: it is faster, more flexible and responsive, capable of dealing with extremism from anywhere and in any language and retains the principle of free and open public spaces for debate."<sup>2</sup>

---

1 The Guardian, 12 March 2017. <https://www.theguardian.com/culture/2017/mar/12/facebook-policy-chief-social-media-must-step-up-fight-against-extremism>

2 Counter-speech on Facebook. <https://demos.co.uk/project/counter-speech-on-facebook-phase-2/>







## 09. Safeguarding the counter-speakers

Those engaging in counterspeech work are called counter-speakers. Throughout history, counter-speakers have faced many challenges including being ridiculed or threatened.

Counterspeech work – even when done with much courtesy and restraint – can anger and antagonise some who are spreading hatred or disinformation.

“I’ve interviewed lots of counter-speakers, and most of them talk about how lonely and emotionally difficult the work is – not to mention the fact that they often become the targets of online attacks themselves” says counterspeech promoter Susan Benesch.

“An online anti-hatred effort is successful if it can reach its goal, whether that goal is to reach the larger reading audience or to change the mind or behaviour of person posting hateful comments.”

- Susan Benesch, Founder of Dangerous Speech Project, in 2019 interview

Full text at: <https://cyber.harvard.edu/story/2019-08/combating-hate-speech-through-counterspeech>

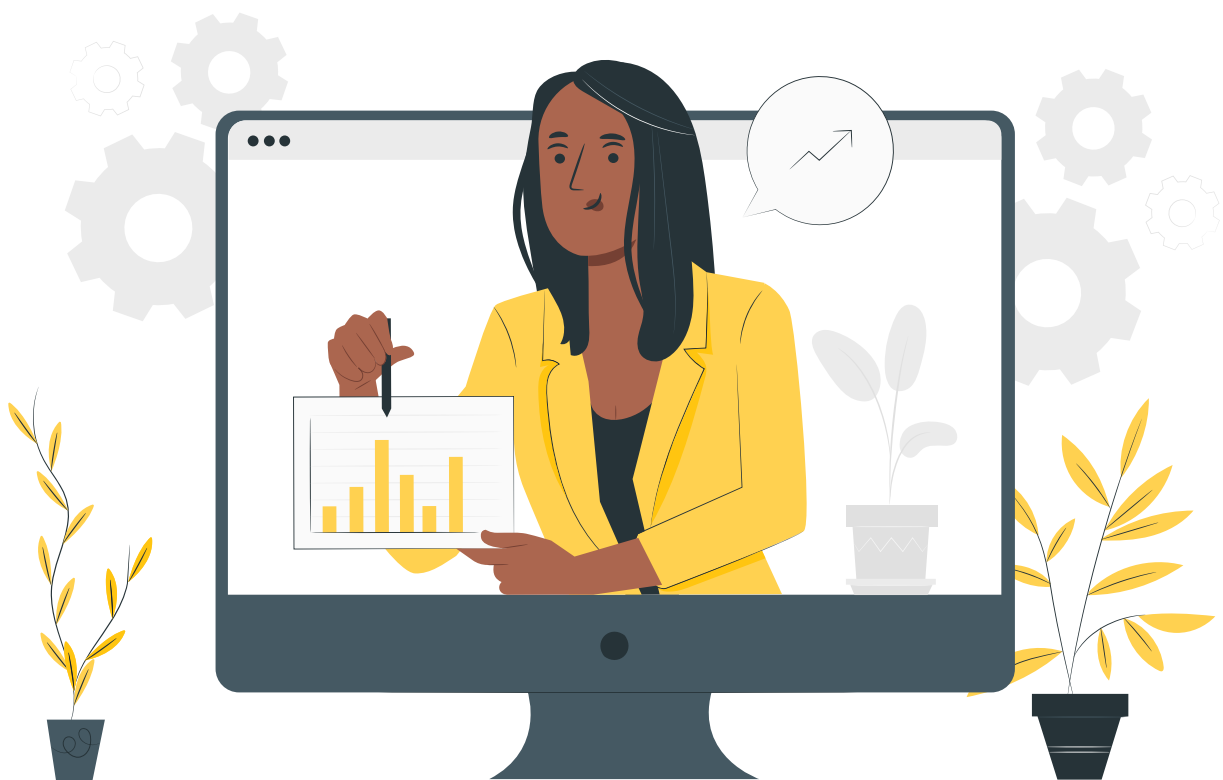
After interviewing dozens of counter-speakers from different countries and societies, she says counterspeech work can take a heavy toll on a person’s personal and emotional life.

As such, counter-speakers need to be mindful of their mental health. They should not take on too much in ways that can be overwhelming.

Also important is to carefully choose which online arguments or debates to get involved in, and decide how far to remain engaged in any argument. Instead of working alone, it is better to work as a team of trusted colleagues so that mental stress and strain can be discussed in private.







## 10. Implementing Counterspeech: Some practical advice

The overarching advice for everyone planning or engaging in counterspeech is this: **please don't make things worse!**

Also remember: **Counterspeech cannot solve** everything that is wrong with problematic speech online or offline. Each situation is different and needs a customised response within that context.

Some individuals spreading hatred or lies (or both) may be firmly convinced that they alone are right: such persons are unlikely to change their minds or stop their actions easily. But counterspeech still tries to engage them in a discussion based on reason and facts.

For those who want to try out counterspeech, here are some tips for achieving better results -- as suggested by the Net Safety Collaborative in the United States.<sup>1</sup>

## BEFORE YOU START

- **Protect yourself:** take steps to protect yourself from possible retaliation. Think about how your online identity or profiles could be used against you.
- **Remind yourself** that behind each comment – no matter how hateful – is a human being. Treat them with courtesy and respect, even if you totally disagree with that they say.
- **Think about what you want to accomplish.** Do you want to change the person’s mind, or how they post or tweet? Stop them from attacking someone?

## COUNTERSPEECH DO’S

Things you can try when 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.”
- **Label the comment, not the person,** like “That word comes from a racist stereotype.”

---

1 <https://socialmediahelpline.com/counterspeech-dos-donts-text-version/>

- **Show empathy and connection with the target** (“I’m a LGBT person too, and...”) or with the speaker (“I’m angry about this too, but...”).
- **Try using humour.** If you’re not mocking the person, humour can help soften the exchange and attract others to show their support.
- **Counter with images** that are silly, clever, or funny – but not hurtful – to de-escalate a situation.

## COUNTERSPEECH DON'TS

Here are a few actions that counter-speakers should not be doing:

- **Don’t label people negatively** – for example, calling them a bigot or racist or ignorant.
- **Don’t assume** the person or group you are countering always has bad intentions (maybe they are simply misled or uninformed)
- **Don’t be hostile,** insulting or aggressive – it can escalate the situation instead of easing
- **Don’t talk down to the other side** – it can shut down all communication.
- **Don’t nitpick, or correct spelling and grammar.** Use a courteous tone and, if countering dis/misinformation, offer a link to an authentic source.
- **Don’t try to silence the person** with threats, social exclusion or any other ‘punishment’.

More practical advice is found at:

<https://www.jagarhar.se/kolumnen/best-practices-counter-speakers/>

<https://onlineharassmentfieldmanual.pen.org/guidelines-for-safely-practicing-counterspeech/>

# 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.

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

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



## Before You Start

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

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



**Think about** what you want to accomplish. Do you want to change the person's mind, or how they post or tweet? Stop them from attacking someone else? Change other people's minds 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 when 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 your intent is kind and you're not mocking the person, humor can soften the exchange and attract others to show their support. Counter 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 intentions.  
**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.



Adapted from "Considerations for Successful Counterspeech" by Susan Benesch, Derek Ruths, Kelly F. Dillon, Haji Mohammed Saleem, and Lucas Wright (The Dangerous Speech Project, October 2016) and Megan Phelps-Roper's TED Talk of February, 2017. Edited by Anne Collier of CaniHelpLine.org, in cooperation with the Dangerous Speech Project, iHeartMob, #CANHELP, and Project HEAR. Design by Kendall Simpson. Special thanks to student adviser Chet Ellis in Conn. and #CANHELP's student focus group in California.

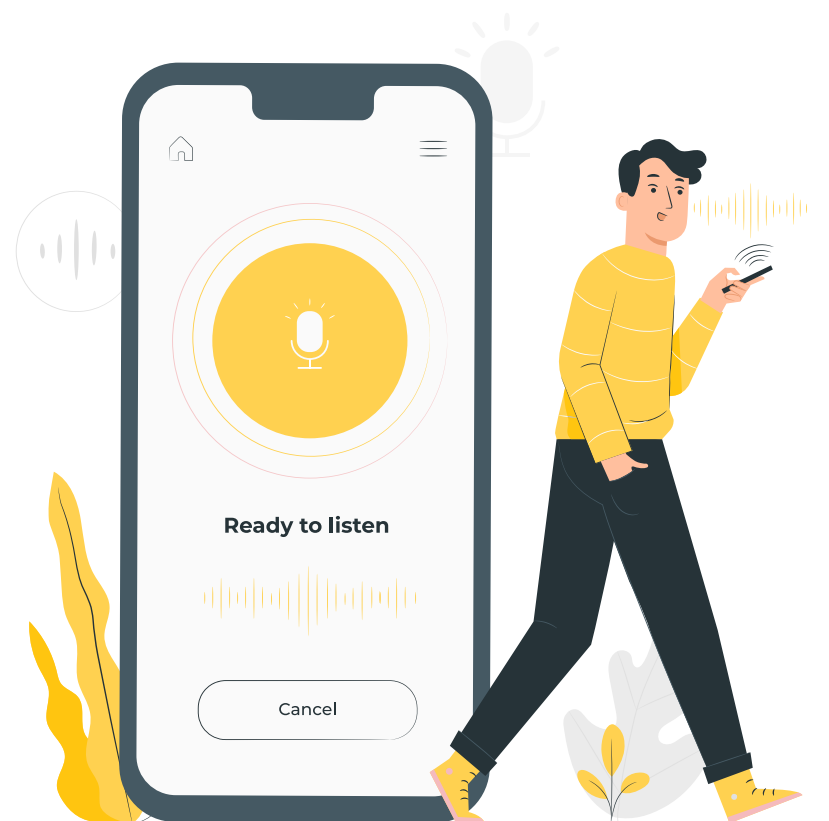
Find our full resource at [CounterspeechTips.org](http://CounterspeechTips.org). For more tips on countering on- and offline harassment, visit [ihollaback.org/resources/bystander\\_resources](http://ihollaback.org/resources/bystander_resources).

Image source: <https://socialmediahelpline.com/counterspeech-dos-and-donts-for-students/>









## II. After Strategising, Action!

After you have strategised well and carefully decided on the best course of action, you can start implementing.

Some counter-speakers get into action like material production and dissemination without going through a strategic thinking process. That can make counterspeech efforts ineffective, or in worst cases, poorly planned counterspeech activity can even cause harm!

Guiding you through material production and their campaign-style dissemination is beyond the scope of this strategy. An excellent resource for this has been produced by the Hashtag Generation, a movement led and run by a group of young tech-savvy, socially conscious Sri Lankans. In 2021, published a toolkit on producing and releasing counterspeech. It is an ideal companion to this Strategy which focuses only on strategic considerations.

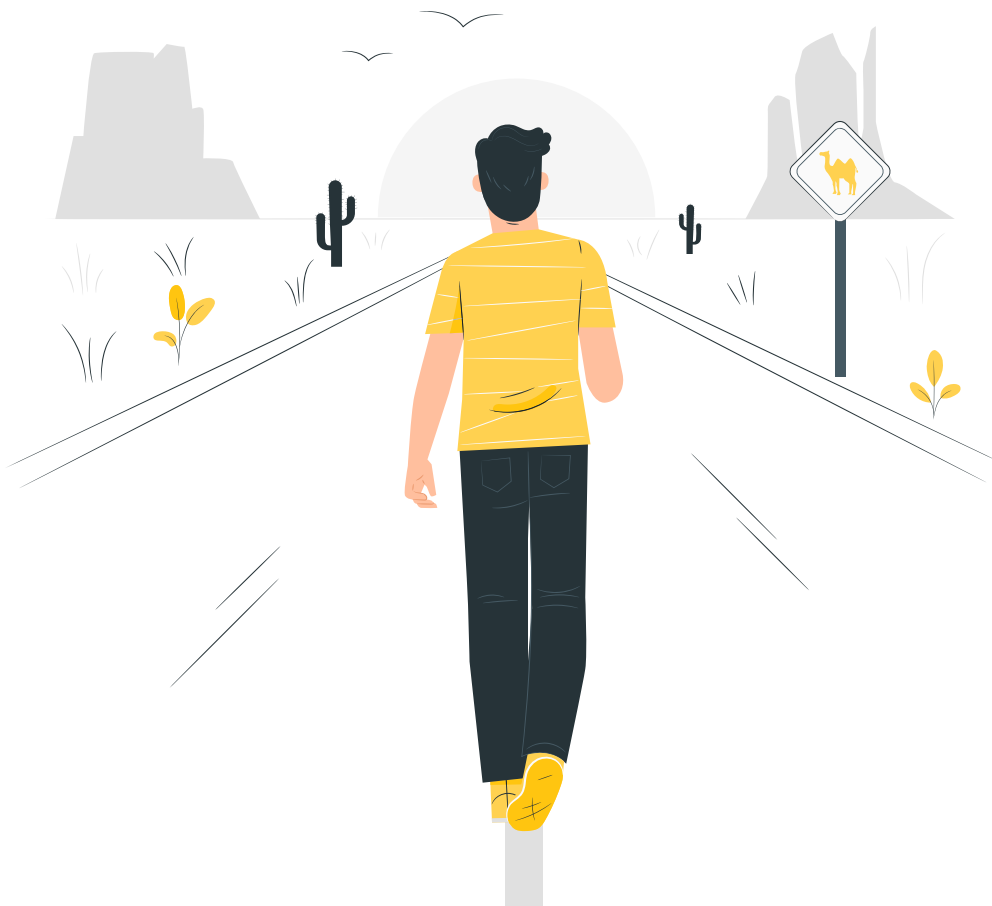
**Counterspeech in Sri Lanka: Toolkit published by Hashtag Generation can be accessed at these locations:**

Report in [English](#), [Sinhala](#), [Tamil](#)









## 12. Conclusion: The way forward

At the heart of all counterspeech lies a recognition that all human beings have equal rights, and nobody should face discrimination, harassment or violence for being who they are.

Counterspeech is only one of several possible responses to problematic speech. It may not work in every situation: some judgement is needed before trying out a counterspeech approach.

The bottom-line should always be: **do not make matters worse!**

When it works, counterspeech can help defuse tension and de-escalate heated arguments or debates so everyone involved can pause to think.

In the medium to long terms, developing media and information literacy can enable individuals to be more discerning. Reforms in mass media and formal education sectors are needed for a more tolerant, inclusive society but are likely to be contentious and

In the long term, counterspeech elements and approaches need to be embedded into formal education, mass media culture, political culture and the civic consciousness of all citizens.





# ANNEX 1

## FACT-CHECKING SERVICES IN SRI LANKA

By the end of 2021, there were several fact-checking services in Sri Lanka - some operated by journalists, others by thinktanks or civil society groups. Their scope, emphasis and methodologies vary, as does their outreach. Such action in fact-checking is indicative of rising levels of dis/misinformation and society’s growing concerns about how falsehoods, distortions and conspiracy theories are affecting politics, economics, public health and even electoral integrity. Here is a listing of the key fact-checking services operating in Sinhala, Tamil or English. It is not an exhaustive list since new efforts keep emerging from time to time. Please note that this is a listing offered without any assessment of their rigour, impartiality or societal impact.

Fact-checking service	Coverage/scope	Online presence
FactCheck.lk  Started in mid-2018	Monitors a select sample of Lankan print media (Sinhala, Tamil, and English) to identify and fact-check statements attributed to high-level decision makers in public office.	Website: <a href="http://factcheck.lk">http://factcheck.lk</a>  Social Media extensions: <a href="https://twitter.com/factchecklka">https://twitter.com/factchecklka</a> <a href="https://www.facebook.com/factchecklka/">https://www.facebook.com/factchecklka/</a>
Fact Crescendo	An independent digital journalism initiative and a part of Crescendo Transcription Private Limited in India. It monitors content shared and viralling on Facebook, and fact-checks in several languages including in Sinhala.	Website: <a href="https://www.factcrescendo.com">https://www.factcrescendo.com</a>  Social Media extensions: <a href="https://www.facebook.com/factcrescendosl/">https://www.facebook.com/factcrescendosl/</a> <a href="https://twitter.com/factcrescendosl">https://twitter.com/factcrescendosl</a>

Fact-checking service	Coverage/scope	Online presence
<p>AFP Fact Check Sri Lanka</p> <p>Started in 2017</p>	<p>The global news agency AFP employs digital verification specialists around the world to monitor online content in local languages. They take into account local contexts including culture and politics. AFP debunks misleading content online and has a dedicated fact check reporter in Sri Lanka.</p>	<p>Website: <a href="https://factcheck.afp.com/afp-sri-lanka">https://factcheck.afp.com/afp-sri-lanka</a></p> <p>Social Media extensions:  <a href="https://www.facebook.com/AFPFactCheck/">https://www.facebook.com/AFPFactCheck/</a>  <a href="https://twitter.com/AFPFactCheck">https://twitter.com/AFPFactCheck</a></p>
<p>FactSeeker.lk</p>	<p>FactSeeker is established to debunk fake news. It verifies news items of public interest and those that affect fundamental rights of every citizen. It is a project of Sri Lanka Press Institute (SLPI) and “aligns with SLPI’s vision of nurturing an informed public committed to democratic ideals”.</p>	<p><a href="https://factseeker.lk/">https://factseeker.lk/</a></p> <p>Social Media extensions:  <a href="https://www.facebook.com/FactSeekerSL/">https://www.facebook.com/FactSeekerSL/</a>  <a href="https://twitter.com/factseeker15">https://twitter.com/factseeker15</a></p>

Fact-checking service	Coverage/scope	Online presence
Citizen.lk  Started in 2017	Fact checking service started by Citizen.lk news and current affairs website, operating in Sinhala and Tamil. This is an initiative of the Citizen Media Network, which also runs Citizen.lk, Esana, Your News and Citizen Wrap.	Website: <a href="https://citizen.lk/factcheck/">https://citizen.lk/factcheck/</a>  Social Media extensions: <a href="https://www.facebook.com/citizenlk/">https://www.facebook.com/citizenlk/</a>  <a href="https://www.youtube.com/channel/UCNu dB J T q - 9sEGCgtL2F0AZA">https://www.youtube.com/channel/UCNu dB J T q - 9sEGCgtL2F0AZA</a>

**Note:** In early 2021, the government also launched a fact-checking service. Known as Fact Research.lk, it is operated by the Media Centre for National Development (MCND) of the Department of Government Information, under the Ministry of Mass Media. <http://www.factresearch.lk/>

## ANNEX 2:

### EXAMPLES OF COUNTERSPEECH EFFORTS IN SRI LANKA

There are two types of counterspeech:

- spontaneous responses coming from many individuals, usually reacting to an incident
- organized counter-messaging campaigns, usually by an organisation or social movement

Here are a few examples from Sri Lanka each of the above categories.

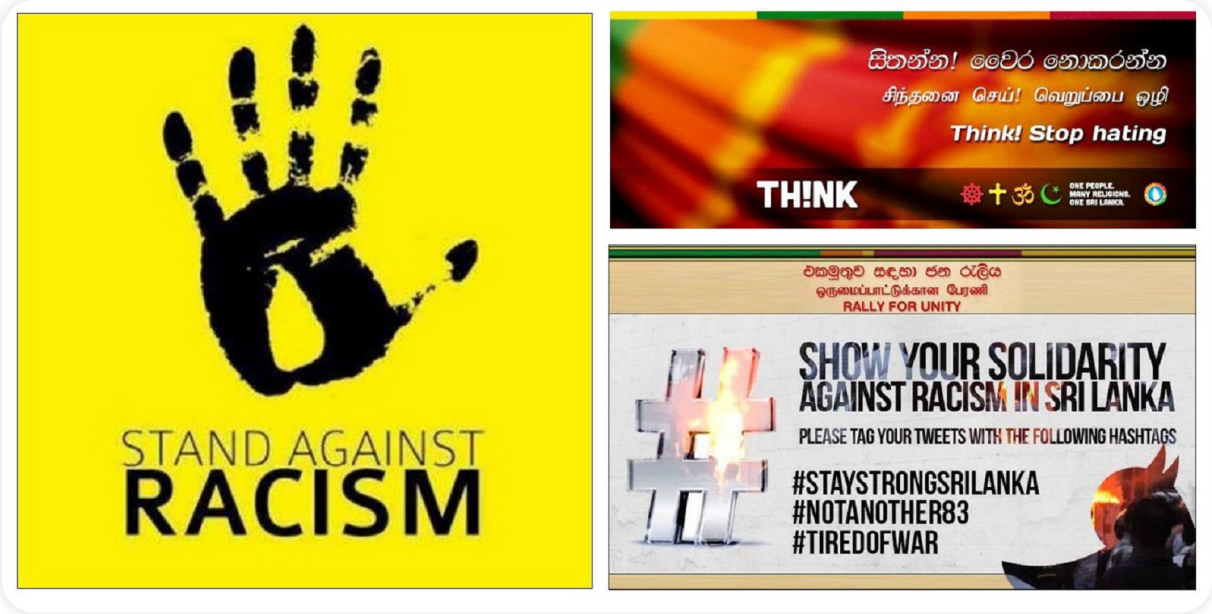
#### SPONTANEOUS RESPONSES COMING FROM MANY INDIVIDUALS

##### *Example 1: Aftermath of Aluthgama violence, June 2014*

Anti-Muslim violence erupted in Dharga Town and Aluthgama, close to Beruwala in the Kalutara district on 14 to 16 June 2014. It was ignited by a private dispute escalating into a communal clash. The violence left at least four persons dead, several dozen injured and Hundreds made homeless after attacks on homes, shops, factories, mosques and a nursery.

Coming just five years after the civil war ended in mid 2009, the racist violence shocked many Lankans. The mainstream media (both state owned and privately owned) largely ignored the incident but thousands of concerned citizens took to social media to express their concerns. Community-generated hashtags (common keywords) included: #TiredOfWar #NotAnother83 #StandAgainstRacism #Aluthgama and #StayStrongSriLanka.

Memes – such as the palm sign with ‘Stand Against Racism’ – were widely circulated, rallying citizens around on calls for racial harmony, compassion for the affected and restraint all around.



For more details: <https://collidecolumn.wordpress.com/2014/06/20/when-worlds-collide-112-social-media-candles-for-media-blackouts/>

*Example 2: Ekama-Lei response*

Around 2015-16, images started appearing on three-wheelers and T-shirts promoting Sinha-Lei which means ‘pure Sinhalese blood’. It soon became a rallying call for Sinhala ultra-nationalism. Those promoting a multi-ethnic and multi-cultural Sri Lanka came up with a counterspeech response Ekama-Lei, implying that blood is not categorised according to ethnicity. [Note: Those promoting this imagery on Facebook and elsewhere came under severe verbal attacks, which highlights the risks that counter-speakers sometimes face.]





*Example 3:* On 21 September 2021, a political cartoon by Dayan Kottachchi was published on the Daiiya Cartoon Facebook page showing a monk splitting the national flag of Sri Lanka. It was indicative of the monk's propagation of hatred against ethnic and religious minorities. Within hours of this cartoon being published, a variation of it was published on the 'Voice of Beruwala' Facebook page: it showed Sri Lankans of all ethnic groups pushing back the divided national flag, strengthening a united Sri Lanka.

This is a good example of user-generated counterspeech that started with a professional drawn cartoon and changed it to convey a positive message.

Original cartoon first appeared at: <https://www.facebook.com/DaiiyaCartoon>

## **ORGANIZED COUNTER-MESSAGING CAMPAIGNS, USUALLY BY AN ORGANISATION OR SOCIAL MOVEMENT**

### *Example 4: The Na Tree Project*

This is a social media based campaign to enhance public understanding about the different ethnic groups, religious faiths and cultural practices found in Sri Lanka. The name comes from the Na Tree (Ironwood, botanically known as *Mesua nagassarium*) which was declared the National Tree of Sri Lanka in 1986.

The Na Tree Project's messaging was done in English, Sinhala and Tamil and ran from July to December 2019 on Facebook, Instagram and Twitter.

Along with visual memes, it also used these hashtags: #OneSkyOneEarthOneNation #KnowYourNeighbour #Unity #Diversity

Shared here are only a few of the many memes produced by the Na Tree Project.



More at:

<https://www.facebook.com/thenatreeproject/>

<https://twitter.com/natreeproject>

<https://www.instagram.com/natreeproject/>

### ANNEX 3:

## EXAMPLES OF COUNTERSPEECH EFFORTS IN OTHER DEVELOPING COUNTRIES

*Example 1: Panzagar or "flower speech" campaign in Myanmar*

Myanmar (also known as Burma) is an ethnically diverse nation with 135 distinct ethnic groups officially recognised. According to 2016 official statistics, 90% of the country's total population were Buddhists while there were 6.3% Christians and 2.3% Muslims. Anti minority sentiments have been around for years and, as

internet and social media use increased after 2010, these also spread online.

In 2014, a group of Myanmar activists launched a campaign to tackle online hate speech against Muslims. Panzagar, which literally means “flower speech”, was set up by Nay Phone Latt, a blogger and executive director of Myanmar ICT For Development Organization (MIDO), a civil society organisation. The campaign’s slogan, when translated to English, means “Let’s watch what we say so that hate between humans does not proliferate”. Its symbol is a person holding a flower in his or her mouth, which means spreading peace through positive speech. It used several hashtags including: #Panzagar #NoHateSpeech #Myanmar #AntiHateSpeechCampaign and #WhiteRose4Peace

Campaign’s Facebook page: <https://www.facebook.com/panzagar>

Originator’s reflections: <https://www.boell.de/en/2018/02/07/halfway-there-myanmars-difficult-transition-toward-freedom-expression>

### *Example 2: Nipe Ukweli or “give me truth” campaign, Kenya*

Kenya’s election held on 27 December 2007 triggered widespread violence after the results were announced and nearly pushed the multi-ethnic country into a civil war. It led to over 1,000 Kenyans being killed, half a million others being driven from their homes. Later, analysts said that inflammatory statements and songs shared on local radio stations and at political rallies, as well as through text messages, emails, posters and leaflets had all contributed to the violence.

As the next national election approached in 2013, Kenyan civil society groups took various steps to avoid post-electoral violence. One initiative was Nipe Ukweli (Swahili for “give me truth”) that sought to tackle the known catalysts of violence. The idea came after disinformation spread in August 2012 that some churches were being attacked and burned. A Kenyan Twitter user posted a photo of one unharmed churches saying “stop the lying”.

Running up to the March 2013 elections, organizers of the Nipe Ukweli campaign worked to educate Kenyans on the impact of hateful speech, and to encourage people to resist and refute false and damaging rumours by employing the same social media channels. As a web page and hashtag, Nipe Ukweli served as a counterweight to dangerous speech during the 2013 election. Large scale violence was averted at this election thanks to the efforts of many officials, activists and



other citizens.

### *Example 3: Peace Provocateurs, in Ambon, Indonesia*

Peace Provocateurs is the opposite idea of conflict provocateurs. This movement was formed in September 2011 by the inter-religious youth group in Ambon City, Indonesia, based on the belief that every person has “seeds” of peacefulness in their hearts.

Ambon is the capital and largest city of the Indonesian province of Maluku where in 2010 the population mix was 58% Protestant Christians, 39% Muslims and 2% Catholic. In 1999-2000 there was inter-communal violence between Christians and Muslims that left several thousand dead. Even after an uneasy peace was restored, there have been occasional, small scale flare-ups. One such incident was in September 2011, when a Muslim motorcycle rider died in a traffic accident in a Christian neighbourhood, triggering false rumours on mobile phone text messages and online that Christians had killed him. It led to Muslim attacks on two Christian areas and a retaliatory attack on a Muslim neighbourhood that destroyed 750 homes and displaced more than 3,000.

An interfaith network of religious leaders, students, activists, teachers and journalists decided to join hands to prevent further outbreaks of violence. Whenever harmful claims spread through rumours or online, they rush volunteers to specific locations to verify facts which are double-checked and sent out as text messages setting the record straight. They called themselves The Peace Provocateurs. They are still active, and when there are no rumours to dispel, they work to diffuse religious tensions by texting and tweeting examples of the two communities working and living together peacefully.

Facebookpage:<https://www.facebook.com/Peace-Provocateurs-102530499863186/>  
Moreinfo: <https://www.independent.co.uk/news/world/asia/how-peace-provocateurs-are-defusing-religious-tensions-indonesia-7562725.html>

### *Example 4: The Resiliency Initiative – across Asia*

The Resiliency Initiative empowers local communities in the Asia Pacific with digital tools to combat hate, violence, and conflict within and beyond their networks. Launched in April 2021, this resource portal provides free access to tools and community networks to navigate the online space and use social media responsibly and effectively.

This project, which is a partnership between the Facebook company and the Asia Foundation, has started working in 10 countries in Asia (Indonesia, Thailand, Philippines, Malaysia, Pakistan, Sri Lanka, Maldives, Bangladesh, Nepal, India) through selected civil society groups.

The project has already released a Step-by-Step Guide to CSO Social Media, covering:

- How to Connect Social Media with The Real World
- How to Communicate with Your Audience
- How to Create Effective Social Media
- How to Measure Your Impact
- Social Media Do's and Don'ts

More at: <https://resiliencyinitiative.org/>

<https://counterspeech.fb.com/en/initiatives/resiliency-initiative/>



