FIELD

#DEFYHATENOW

Social Media Hate Speech Mitigation

CAMEROON
SECOND EDITION 2021

GUIDE

#defyhatenow
https://www.facebook.com/defyhatenow/
https://www.facebook.com/defyhatenowwca
https://twitter.com/defyhatenow
https://twitter.com/DefyhatenowWCA

Scan to download the Social Media Hate Speech Mitigation Field Guide Kit, Cameroon

FIELD

#DEFYHATENOW

Social Media
Hate Speech
Mitigation

CAMEROON
SECOND EDITION 2021

GUIDE

Cameroon FIELD GUIDE

#HATEFREECAMEROON

#defyhatenow Cameroon exists to strengthen the voices of the youth, communities, and media-focused organizations and to support them in their work.

Together, we aim to counter the hate speech, conflict rhetoric, and incitement to violence that is spread on social media platforms and online in response to **Cameroon's Anglophone conflict**.

The project provides capacity-building and media literacy training to enable community-based organizations and citizens, including people displaced by conflict, to become positive influencers. #defyhatenow teaches skills related to counteraction, fact-checking, early response monitoring, peacebuilding through art and tech, and trauma healing. Our goal is to spread positive, tolerant, peaceful messages online in order to be a counterforce to the agents of conflict posting online.

Since 2019 the **r0g_agency for open culture & critical transformation gGmbH**, in partnership with the **UNESCO Multisectoral Regional Office for Central Africa** in Yaoundé and with the support of the German Federal Foreign Office, has been mandated under the initiative of #defyhatenow to address the way the internet is fueling the country's crisis. Our in-country and international #defyhatenow teams and conflict specialists have been acting to address the civil conflicts and destabilization issues, including violent extremism and internal displacement, that have been fueled through hate speech on social media platforms. Our teams have also been working to mitigate the intentional disinformation spread regarding the **Covid-19 pandemic** that has exacerbated the conflict.

#defyhatenow engages people from all walks of life in order to create a peacebuilding framework. We are working to bridge gaps of knowledge and awareness about how social media works between those with easy access to technology and knowledge and those without.

Through the grant provided by the **German Federal Foreign Office**, the peacebuilding mission of **#defyhatenow** has been enabled to keep the many conflict mitigation conversations going, develop resources that are free and easy to access online, including this **Social Media Hate Speech Mitigation Field Guide**. Their grant has also allowed us to support local activists and peace-oriented media-based activities in local communities.

#DEFYHATENOW BACKGROUND

#defyhatenow was initiated in early 2014 by the **r0g_agency for open culture and critical transformation gGmbH (Berlin)** and initially focused its work on South Sudan, with support from the German Federal Foreign Office via the ifa/zivik program for civic conflict resolution. In 2016 the program expanded to include Kenya, Uganda, and Sudan and later, Ethiopia. Alongside social media literacy training, community activities and trainings on hate speech and conflict mitigation, **#defyhatenow** produced a comprehensive **Social Media Hate Speech Mitigation Field Guide**. In 2019 **#defyhatenow** began work in Cameroon in response to the Anglophone conflict.

This Field Guide is a toolkit to work together for compassion and tolerance - online and offline.

Historical Background

The current Anglophone Crisis in Cameroon and the longstanding tensions between the anglophone and francophone communities are rooted in Cameroon's colonial past. Before independence, Cameroon was administered as two separate entities by Britain and France, who were granted control of the former German colonial territory of Kamerun at the end of World War I.

Each territory thus came under a new foreign language, with English and French replacing German as official languages. The territories also had different political, administrative, educational, and legal systems. Different ways of life developed under French and British administration.

The British territories (there were two non-connected territories) in the western area made up about 20% of the total land area of today's Cameroon. The larger territory administered by France made up 80% of the land area.
The territory under French administration gained independence in 1960. The territories under British administration gained independence in 1961 and voted to join French Cameroon, thereby forming the Federal Republic of Cameroon.

Today, those who are from the territories formerly administered by Britain are referred to as Anglophones. Those who are from the territory formerly under the French administration are referred to as Francophones.

In the decades since independence and reunification, the government has been exclusively under francophone rule. Anglophones have been frustrated by the francophone majority in government and their self-interested allocation of socio-economic amenities.

Over the years, the government has worked towards erasing anglophone cultural heritage, trying to ban the use of English in schools and courts. This has been met with anglophone resistance. In 2016, this conflict reached a tipping point, with protests by anglophone teachers and lawyers escalating into the present-day separatist armed conflict.

The differences between anglophone and francophone cultures has become the basis for negative stereotyping.

These stereotypes, along with escalating language tensions, have given rise to violence and hate speech. Our goal at #defyhatenow is to teach people how to mitigate hate speech online and spread messages of peace, tolerance, and reconciliation.

For more detailed background information on the Anglophone Crisis, here are some suggested links:

https://theconversation.com/history-explains-why-cameroon-is-at-war-with-itself-over-language-and-culture-85401

https://africacenter.org/spotlight/understanding-cameroons-crisis-of-governance/

https://www.researchgate.net/publication/333613485_CAMEROON'S_UNFOLDING_CATASTROPHE_Evidence_of_Human_Rights_Violations_and_Crimes_against_Humanity

SOCIAL MEDIA HATE SPEECH MITIGATION FIELD GUIDE

The **#defyhatenow** Social Media Hate Speech Mitigation Field Guide offers tools and strategies that can be used by individuals and entire communities to engage in peacebuilding activities online, locally and around the world. This booklet is designed for use 'in the field' and includes training materials and workshop resources designed to raise awareness, develop counter-narratives, and mitigate violence related to online hate speech.

"Incitement to violence is very specific, in that it needs strategic action to counter and requires community leaders and citizens to become involved and engaged in direct actions to mitigate the threat of violence erupting. While hate speech can form a basis for incitement, one can still use personal strategies to engage with speakers, bring down the tone of rhetoric, and shift attitudes. Group strategies are needed to respond as a community to dangerous speech online and mitigate the factors contributing to violence offline."
Stephen Kovats, r0g_agency, Berlin

Contents

Social Media and Conflict

Contents

A meet up session with students of GBHS SODIKO_Bonaberi on how they can engage their communities against hate speech. Photo by Angela Okon

Social Media and Conflict

The #defyhatenow project works to mitigate hate speech online and spread positive, peaceful alternatives to hate.
Before we can talk about positive alternatives, we first have to talk about what hate speech is and why it is such a threat to peace.
Understanding this is the goal of this chapter.

► Hate Speech

A Definition

When we talk about any term, such as **hate speech**, the first step is always to create a working definition of what that term means to us.

Here is the definition adopted by the European Court of Human Rights:

> **Hate speech is all forms of expression which spread, incite, promote, or justify racial hatred, xenophobia, anti-Semitism, or other forms of hatred based on intolerance, including intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility towards minorities, migrants, and people of immigrant origin.**

A simpler way to state this is:

> **Hate speech is all communication that attacks people and describes them as less valuable, less good, or less deserving of full participation in society and hate speech makes these verbal attacks because the target is a member of a particular group.**

And an even simpler way:

> **Hate speech is speech marked by hatred of someone else because of their membership in a group.**

It is important to note that hate speech focuses on disdain for people based on a group they belong to. It is not about a conflict between two individual people, based on their personal relationship, but rather hatred towards an entire group of people. This is what makes hate speech so dangerous, because it easily escalates into larger conflicts.

Hate speech can be spoken, written, or conveyed through images.

In this Field Guide, along with talking about how to counter hate speech, we will also talk about how to counter communication that fosters a climate of prejudice and intolerance – what you might think of as hateful speech. While hateful speech does not meet the definition of being serious enough to be hate speech, this kind of communication may fuel discrimination, hostility, and violent attacks later on and always has a negative effect on communities.

► QUESTIONS TO CONSIDER ◄

- Which of these definitions makes the most sense to you? Can you describe what hate speech is, in your own words?

- What groups in your area are most often the target of hate speech?

- How do you think hate speech affects the people who are targeted?

Hate Speech and Violence

So now that we know what hate speech is, the question is, why does it matter?
The simple answer is that words have profound power to either bring communities and countries together or to tear them apart and fuel conflicts, wars, and genocides.

When we look back to violent acts that took place in the past, we see that they did not start with actions, they started with words. That is why we need to work together to end the spread of hate speech, so that violence does not erupt and people are not hurt.

But how exactly does hate speech lead to violent conflict?

To better understand this, here is an image that shows how the cycle of hate speech intensifies and how the more often it cycles through, the more it intensifies.
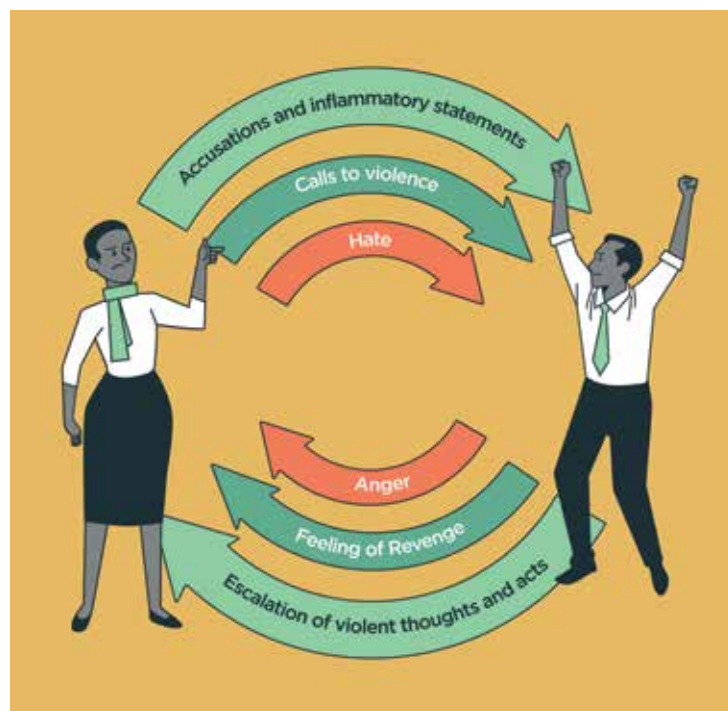
► QUESTIONS TO CONSIDER ◄

- When looking at the image above, you can see that these cycles happen because the person hearing hate speech responds with anger or revenge. But how could you instead stop the cycle, even if hate speech is directed at you? What reactions would have a different, more positive outcome?

- What effect do you think hearing hate speech has on your community? Does it increase or decrease your levels of trust in other people? Levels of anger? Feelings of revenge?

- How does listening to hate speech or dangerous speech on the radio, in person, or online make you feel?

- What could you do to stop the spread of hate speech in your own community, your family, or school?

Hate Speech vs. Dangerous Speech

Dangerous Speech is any form of expression (speech, text, or images) that can increase the risk that its audience will condone or commit violence against members of another group.

When does hate speech become dangerous speech?
So far, we have talked about hate speech and how it can fuel conflicts. When hate speech is combined with a call to action, it turns from hate speech to dangerous speech.
What is dangerous speech? Dangerous speech is any communication that calls for people to condone, or take part in, violence against another group or members of that group. Violence means causing bodily harm to someone else.

What is important about dangerous speech is that it includes committing (acting on) and condoning (approving of) violence. The reason dangerous speech includes both acting on and approving of violence: when there is large-scale violence between people, only a small proportion of people, usually young men, carry out violence. However, people close to them - such as family members, friends, and teachers - often approve of and encourage the violence.

When a society suffers major intergroup violence, a few commit the violence and many stand by, watching and approving of it.

This also means that when people stand by and witness violence, other people assume they are approving of it. So if you don't approve, speak up! The more vocal the non-approving voices become, the harder it is for dangerous speech to take hold.

If we look at past outbreaks of violence, we can see that dangerous speech rises steadily before outbreaks of mass violence. Often, a few influential speakers gradually incite a group to violence. This means that violence can often be prevented by interfering with this process:

- **We could try to inhibit the speech, not letting the speaker share the dangerous speech message.**
- **We could try to limit how widely the message is spread.**
- **We could work to undermine the credibility of the speaker.**
- **We could teach people how to recognize dangerous speech and why calls to violence should not be listened to, so that the speech has less impact on the listener.**

In general, it is easiest to identify dangerous speech that comes just before violence is about to break out. Its meaning tends to be clear and it often calls for and directly endorses violence.

Years or months earlier, dangerous speech is often expressed in much less clear language, so that both its meaning and its impact are less obvious. The calls to violence are not as obvious. But we should still take any speeches seriously that imply violence is the solution to problems.
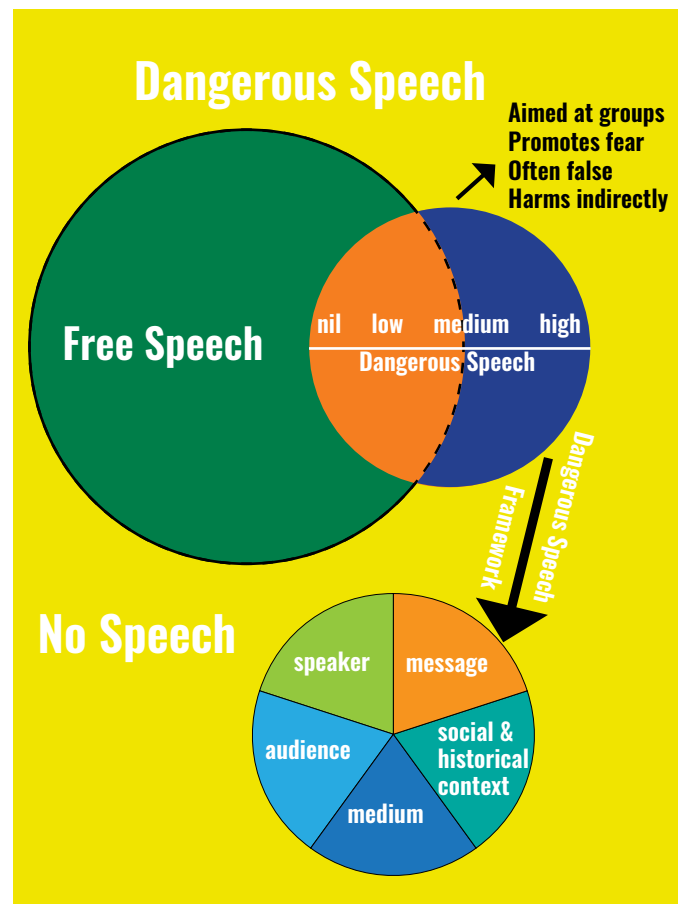
Source: https://dangerousspeech.org/guide/

► QUESTIONS TO CONSIDER ◄
- **Can you explain the difference between hate speech and dangerous speech?**
- **If you come across dangerous speech, how could you try to interfere to prevent violence?**
- **In most cultures, violence is committed by young men. Knowing this, what are some things we could do to prevent these young men from being infected by dangerous speech, if they come across it? What are some preventative steps we can take?**

Hate Speech Laws

Most people don't realize that many countries have laws prohibiting hate speech. Hate speech laws were created in Europe after World War II (1939-1945) when countries were confronting the fact that millions of Jews were murdered by their governments during the war, a genocide known as the Holocaust or Shoah. During the Holocaust, people who had ethnic or religious ties to the tribe of Israel were murdered solely because they belonged to that group. But the Holocaust did not start with genocide - it started with the dominant group fueling hatred through targeted hate speech. So laws banning hate speech were created after the war. Governments realized that by not allowing hate speech to be spread, you could prevent fighting, genocide, and wars.

Countries such as Germany and Austria passed hate speech laws decades ago. Many other countries have since followed. For example, under Kenyan law, a person commits an offense if they stir up "ethnic hatred". France goes further, forbidding any communication intended to incite discrimination, hatred, or harm with regards to ethnicity, nation, race, religion, sex, sexual orientation, or handicap.

Hate Speech Laws in Cameroon

Did you know that in 2019, Cameroon also passed a law prohibiting hate speech?
The law in its full form is listed below.

The hate speech law, passed on 24 December 2019, is an amendment of Law No. 2016/7 of 12 July 2016 relating to the Penal Code for hate speech.

SECTION 1
Law No. 2016/7 of 12 July 2016 relating to the Penal Code is hereby amended and supplemented as follows: Section 241: (new) Contempt of race or religion.

1 **Whoever commits a contempt, within the meaning of Section 152 of this Code, of the race or religion of one or many citizens or residents shall be punished with imprisonment for from 6 (six) months and with fine of from 5 000 (five thousand) francs to 500 000 (five hundred thousand) Francs.**

2 **Where the offence is committed by means of the press or wireless, radio, television or social media or any other means likely to reach the public, the maximum of the fine provided for in subsection 1**

above may extend to 20 000 000 (twenty million) Francs.

3 **Where the offence is committed with intent to arouse hatred or contempt between citizens, or residents, the penalties provided by the foregoing subsections shall be doubled.**

SECTION 241-1
(new) contempt of Tribe or ethnic group.
1 **Whoever, by any means, makes hate speech against people or incites them to violence due to their tribal or ethnic origin shall be punished with imprisonment of from 1 (one) to 2 (two) years and with fine of 300 000 (three hundred thousand) francs to 3 000 000 (three million) francs.**

2 **Where the benefit of mitigating circumstances is given, the punishment provided for in Subsection 1 above shall not be less than 3 (three) months imprisonment and the fine shall not be less than 200 000 (two hundred thousand) francs. Execution shall not be suspended except in case of diminished responsibility of infancy.**

3 **Where the author of the hate speech is a Public Servant as per the provisions of Section 131 of this Code, leader of a political party, of the media, of a Non-Governmental Organisation or a religious institution, the punishment provided for in subsection 1 above shall be doubled and the benefit of mitigating circumstances shall not be given.**

SECTION 77
Law no. 2010/12 relating to Cyber Security and Cyber Criminality.

This is a law on cyber-crime that punishes hate speech with imprisonment for up to 5 years and up to a fine of 5 Million francs for anyone disseminating hate speech online.

Source: https://www.prc.cm/en/multimedia/docu-ments/8027-law-2019-020-of-24-december-2019

► QUESTIONS TO CONSIDER ◄
- Did you know that there were laws forbidding hate speech?
- Do you feel these laws are enforced by the authorities?
- Could you ask your government officials to make enforcing hate speech laws a priority?

Hate Speech Online

While it is important to have laws forbidding hate speech, whether the laws are enforced or not varies widely between countries. Many countries respond to hate speech in printed form, but have a hard time with online hate speech. There is so much content online across all social media platforms and on the internet. This makes it hard for any one government to monitor everything.

What this means is that even though there are laws forbidding it, hate speech is still often shared and reposted on social media. It is therefore your responsibility as users of social media and websites to make sure that we do not post or share hate messages and to report messages of hate that we come across to the platform.

► Hate Speech Targets

Hate speech, both in person and online, attacks many people, but there are some groups that are more often the target of hate speech than others. Below, we will look at three of these groups and discuss what actions we can take to help minimize the impact of hate speech against these groups.

Tribal Groups

Tribalism is defined as the behaviors and attitudes that are a result of having an identity strongly defined by the tribe someone belongs to. Tribalism creates a way of seeing the world that leads people in one tribe to see themselves as separate, distinct, and often better than people from another tribe.

Members of a tribe traditionally have strong loyalties to their tribe, because tribes foster strong relationships among its members. These relationships are based on proximity, kinship, and the mutual survival of both the individual and the tribe itself.

In return, tribes demand loyalty. The feeling of belonging depends on knowing who is part of the tribe and who is not. Tribes foster the us-versus-them thinking that we will discuss in chapter 4 on Identity.

Cameroon is one of Africa's most diverse countries. There is the country's Anglophone – Francophone divide, but there are also over 250 other groups that create a complex patchwork of affiliations and tribal identities. An ethnic Anglophone, for example, is a person whose ethnic roots are in the former British Southern Cameroons, as opposed to a linguistic Anglophone, meaning an ethnic Francophone Cameroonian who through education or socialization speaks English.

Against this backdrop, hate speech against linguistic and tribal groups has often been used by politicians seeking their own gain. Most violence and atrocities in Cameroon today are based on tribal conflicts. These conflicts continue to destabilize the country and harm many people.

On social media, mis- and disinformation, along with hate and hateful speech, have become common. False allegations spread rapidly online and members of other groups are often dehumanized and people try to incite hatred across borders.

What can be done? First off, report any hate speech you come across online.

And keep in mind that while we all long for the safety and feeling of belonging, we don't have to resort to tribalism. Our desire to belong to a group can also lead us to other actions, such as wanting to be part of a group that works together, cooperates, embraces diversity, and helps other people who are different from us but united through our joint humanity.

Source: https://www.rightsforpeace.org/post/fanning-the-flames-hate-speech-and-elections-in-cameroon

► QUESTIONS TO CONSIDER ◄
- Have you or someone you know been attacked online because of your ethnicity or tribe?
- What actions could you take to make members of the targeted groups feel safer online?
- What could you do to promote cooperation between groups?

Women

Women are disproportionately the target of hate speech online and other gender-based violence, making the online and offline worlds far less safe for women than for men.
Online, women are far more likely to be verbally attacked, degraded, threatened, and harassed than their male counterparts are.

Tanya O'Carroll, the director of Amnesty Tech, calls online abuse a form of censorship. "Especially given that Facebook, Twitter, and others talk about the fact that they're mainly there to promote freedom of expression, they're very reluctant to take down content," she says. "These platforms are talking about the censorship consequences of removing content, but what about the censorship consequences of this kind of targeted harassment? What are the net consequences of silencing women in this very orchestrated way? Women turning off their accounts or turning away from sharing specific kinds of content, for example sharing political views - those are also censorship consequences," says O'Carroll.

Insults, public shaming, intimidation, hacking, and cyber-stalking are behaviors that women are often confronted with on the Internet. Women in leadership positions, politics, and journalism have been targeted by online violence and hate speech and have experienced the wrath of online abuse, threats, and bullying.

But not using these platforms is not the answer. As Gomolemo Rasesigo from Gender Links Botswana says, "we cannot run away from social media." Especially in politics, female candidates need social media to help reach out to potential voters and increase awareness of the issues at stake. So the solution is not to stop women from using platforms, but rather for all of us to work towards making it safer for women. How? By reporting all bullying and harassment we witness online.

Another very common violence against women online is harassment through stalking, sex chats and video calls, demands for nudes, hacking, threats for extortion, and the sharing of non-consensual pornography (commonly known as revenge porn).

Non-consensual pornography can happen because individuals, at one point in a relationship, share photos with each other. Yet when the relationship ends, these women are blackmailed into transferring money or having the photos posted online. At times, women are also targeted with the threat of photos being leaked, even though they are photoshopped images and not real. But the fear of these fake photos being online is nonetheless terrifying for these women. This trauma often forces women to deactivate profiles, effectively shutting them out of the online world.

► **QUESTIONS TO CONSIDER** ◄
- Have you or someone you know been the target of hate or hateful speech online based on being a woman?
- Why do you think women in positions of power are more likely than their male counterparts to be attacked online?
- What could you do to make the online world safer for women?

LGBTQ+ Community

Discrimination and harassment are constant problems for LGBTQ+ Cameroonians. They are at risk of imprisonment and extortion by law enforcement officials. Some gay men have even been entrapped by neighbors or acquaintances conspiring to report them, which creates a climate of distrust and fear. Many LGBTQ+ Cameroonians believe the only way to protect themselves is to hide their sexuality, especially since the 2005 arrests when the media began to portray being gay or lesbian as a menace to public safety.

Cameroon's laws criminalize consensual same-sex conduct under article 347(1) of the Penal Code, which punishes sexual relations with a person of the same sex with up to five years in prison. Laws pertaining to the online world also exist, including Article 83(1), which states that any person who makes sexual propositions to a person of their sex through electronic communications shall be punished with imprisonment of up to two years and a fine of up to 1,000,000 CFA francs or only one of these two penalties. The penalties are doubled if the proposals have been followed by sexual intercourse.

What this means is that members of the LGBTQ+ community are not only legally discriminated against, but this also makes them a particularly vulnerable target for hate speech online, which has in recent years also led to physical violence targeting LGBTQ+ individuals.

The same advice for countering hate speech against this community applies: Report all harassment and hate speech you see online.

Source:
https://www.refugeelegalaidinformation.org/cameroon-lgbti-resources

► **QUESTIONS TO CONSIDER** ◄
- What actions could you take to make women and members of the LGBTQ+ community feel safer online?

► Social Media Platforms

In this chapter, we have defined what hate speech, hateful speech, and dangerous speech is. We have also talked about hate speech laws and identified some common targets of hate speech. And we have said that when you come across hate speech online, you should report it. But how does that work? Below, we give you instructions.

In Cameroon, the most widely used social media platform is Facebook, accounting for about 85% of social media usage. Other platforms, such as Instagram, YouTube, Twitter, and TikTok are also gaining users, but Facebook is undoubtedly the most used platform in Cameroon. The messaging app WhatsApp is also very popular.

Since these are the platforms most used, this is also where you are most likely to encounter hate speech. Because these platforms are owned by private companies, what they consider hate speech and how you report hate speech varies from platform to platform.

Hate Speech Guidelines across Platforms

Below are the guidelines for each platform. The text is taken directly from the platform's website.

Facebook / Instagram says:
We define hate speech as a direct attack against people - rather than concepts or institutions - on the basis of what we call protected characteristics: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease. We define attacks as violent or dehumanizing speech, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, cursing and calls for exclusion or segregation. We also prohibit the use of harmful stereotypes, which we define as dehumanizing comparisons that have historically been used to attack, intimidate, or exclude specific groups, and that are often linked with offline violence. We consider age a protected characteristic when referenced along with another protected characteristic. We also protect refugees, migrants, immigrants and asylum seekers from the most severe attacks, though we do allow commentary and criticism of immigration policies. Similarly, we provide some protections for characteristics like occupation, when they're

referenced along with a protected characteristic. Sometimes, based on local nuance, we consider certain words or phrases as code words for PC groups.

We recognize that people sometimes share content that includes someone else's hate speech to condemn it or raise awareness. In other cases, speech that might otherwise violate our standards can be used self-referentially or in an empowering way. Our policies are designed to allow room for these types of speech, but we require people to clearly indicate their intent. If the intention is unclear, we may remove the content.

We aim to prevent potential offline harm that may be related to content on Facebook. While we understand that people commonly express disdain or disagreement by threatening or calling for violence in non-serious ways, we remove language that incites or facilitates serious violence. We remove content, disable accounts and work with law enforcement when we believe there is a genuine risk of physical harm or direct threats to public safety.

YouTube says:
At the heart of our approach are the four Rs: we Remove content that violates our policies as quickly as possible, Reduce the spread of harmful misinformation and content that brushes up against our policy lines, Raise up authoritative sources when people are looking for news and information, and Reward trusted, eligible Creators and artists.

Hate speech is not allowed on YouTube. We remove content promoting violence or hatred against individuals or groups based on any of the following attributes: Age, Caste, Disability, Ethnicity, Gender Identity and Expression, Nationality, Race, Immigration Status, Religion, Sex/Gender, Sexual Orientation, Victims of a major violent event and their kin, Veteran Status. If you find content that violates this policy, report it.

WhatsApp says:
You will not use (or assist others in using) WhatsApp in ways that [...] are illegal, obscene, defamatory, threatening, intimidating, harassing, hateful, racially or ethnically offensive, or instigate or encourage conduct that would be illegal or otherwise inappropriate, such as promoting violent crimes, endangering or exploiting children or others, or coordinating harm; or involve publishing falsehoods, misrepresentations, or misleading statements.

Twitter says:
Hateful conduct: You may not promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. We also do not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories.

Violent threats: We prohibit content that makes violent threats against an identifiable target. Violent threats are declarative statements of intent to inflict injuries that would result in serious and lasting bodily harm, where an individual could die or be significantly injured, e.g., "I will kill you."

Note: we have a zero-tolerance policy against violent threats. Those deemed to be sharing violent threats will face immediate and permanent suspension of their account.

TikTok says:
TikTok is a diverse and inclusive community that has no tolerance for discrimination. We do not permit content that contains hate speech or involves hateful behavior and we remove it from our platform. We suspend or ban accounts that engage in hate speech violations or which are associated with hate speech off the TikTok platform.

We define hate speech or behavior as content that attacks, threatens, incites violence against, or otherwise dehumanizes an individual or a group on the basis of the following protected attributes: Race, Ethnicity, National Origin, Religion, Caste, Sexual orientation, Sex, Gender, Gender identity, Serious disease, Disability, Immigration status.

Do not post, upload, stream, or share hateful content related to an individual or group, including:
- claiming that they are physically, mentally, or morally inferior
- calling for or justifying violence against them
- claiming that they are criminals
- referring to them as animals, inanimate objects, or other non-human entities
- promoting or justifying exclusion, segregation, or discrimination against them
- content that depicts harm inflicted upon an individual or a group on the basis of a protected attribute

Slurs
Slurs are defined as derogatory terms that are intended to disparage an ethnicity, race, or any other protected attributes listed above. To minimize the spread of egregiously offensive terms, we remove all slurs from our platform, unless the terms are reappropriated, used self-referentially (e.g., in a song), or do not disparage. Do not post, upload, stream, or share content that uses or includes slurs.

Hateful ideology
Hateful ideologies are those that demonstrate clear hostility toward people because of their protected attributes. Hateful ideologies are incompatible with the inclusive and supportive community that our platform provides and we remove content that promotes them.

Do not post, upload, stream, or share:
- Content that praises, promotes, glorifies, or supports any hateful ideology.
- Content that contains names, symbols, logos, flags, slogans, uniforms, gestures, salutes, illustrations, portraits, songs, music, lyrics, or other objects related to a hateful ideology.
- Content that denies well-documented and violent events have taken place affecting groups with protected attributes.
- Claims of supremacy over a group of people with reference to other protected attributes.
- Conspiracy theories used to justify hateful ideologies.

How to Report Hate Speech

Social media platforms continue to find new ways to monitor and counter hate speech on their platforms. It is therefore always a good idea to flag and report any hate speech you come across. Social media platforms keep updating their user interface, so if the instructions below don't seem to work, updated information can always be found online.
Most often, there is some icon next to the post (dots, arrow, etc) that you can click and then select **Report** from a drop-down menu.

Below are links to the Help Pages for the various platforms. You can find help and answers to questions there.

Facebook Help Page
https://www.facebook.com/help

Instagram Help Page
https://www.facebook.com/help/instagram/192435014247952

YouTube Help Page
https://support.google.com/youtube/answer/2802027

WhatsApp Help Page
https://faq.whatsapp.com/?lang=en

Twitter Help Page
https://help.twitter.com/en

Report hate speech on TikTok
https://support.tiktok.com/en/safety-hc/report-a-problem

► QUESTIONS TO CONSIDER ◄

- Which social media platforms do you use?
- Do you usually report posts or accounts when you see them posting something that goes against guidelines? Why or why not?
- How often do you see hate speech messages on social media? How does that make you feel?

Additional Materials

Additional information on dangerous speech an be found at:
https://dangerousspeech.org/what-is-ds/

An article on how being a part of a group can lead to prosocial action:
https://www.theatlantic.com/ideas/archive/2022/01/tribalism-myth-group-solidarity-prejudice-conflict/621008/