CENTRE for INFORMATION RESILIENCE

NO SAFE SCROLL

Investigating TFGBV on TikTok and YouTube in Ethiopia

NO SAFE SCROLL

Investigating Gendered Hate Speech on TikTok and YouTube in Ethiopia

April 2025

TABLE OF CONTENTS

WARNING: This report contains hate speech, rude language, and harmful narratives. These narratives are not the views of CIR, they were posted on social media by social media users and analysed as part of this investigation into gendered hate speech in Ethiopia. While efforts have been made to display the findings sensitively, the report still includes information which some readers may find distressing. For resources on gender-based violence, see NO MORE's Global directory of support services.

1. EXECUTIVE SUMMARY

The Centre for Information Resilience's (CIR) investigations into technology-facilitated gender-based violence (TFGBV) in Ethiopia reveal that gendered abuse is harmful and pervasive, fostering an environment where women are routinely dismissed or ridiculed. TFGBV is deeply intertwined with entrenched gender norms and misogynistic attitudes. Online abuse is shaping who gets to speak, lead, and exist safely in digital spaces. Ensuring that women and girls can participate safely and freely in digital spaces is not only a matter of online safety – it is essential for their full inclusion in public life.

CIR has expanded its research on TFGBV and gendered hate speech on Facebook, Telegram and X (formerly Twitter) to two more platforms: TikTok and YouTube. By expanding the evidence base, this report equips stakeholders with the knowledge to drive meaningful policy change and strengthen advocacy efforts.

The project aims to:

- strengthen the evidence base on TFGBV in Ethiopia;
- inform government institutions, civil society organisations (CSOs), social media companies, and the public about TFGBV in Ethiopia; and
- empower Ethiopian stakeholders with practical recommendations for addressing TFGBV.

This research uncovers new insights through data-driven analysis and expert workshops. Using its academically peer-reviewed methodology, CIR analysed gendered hate speech on YouTube and TikTok, and compared the findings with CIR's study on Facebook, Telegram and X. CIR used Natural Language Processing (NLP) techniques to aid the sampling of posts, and researchers manually annotated 17,082 comments across four languages (Amharic, Afaan Oromo, Tigrigna, and English) analysing: hate speech targets, hate types, and sentiments of speech. Finally, CIR investigated whether there were gendered trends by video genre.

KEY FINDINGS

In Ethiopia, women and girls face a barrage of insults, stereotypes, mockery, and degrading rhetoric; this diminishes their voices and restricts their participation in public life. Women who challenge traditional gender roles, such as those in leadership positions, sports or those advocating for feminism and women's rights, face particularly severe abuse, including sexualisation, discrediting, insults, and stigma. For example, feminism was framed as a threat to traditional values, with feminists discredited through accusations of financial fraud or a conflation with lesbianism. These narratives, along with the dismissal of digital rights, have contributed to a digital environment that silences women.

Men and boys are also restricted by rigid gender roles, with hate speech being used to target perceived weakness, lack of masculinity, or for supporting gender equality. While both male and female political figures face online abuse, men are criticised for their policies or political ties, and women are undermined based on their appearance or traditional societal roles. Additionally, abuse targeting men often weaponises or objectifies the female gender as a means of shaming and emasculating them. Resultantly, women not only experience gender-based violence firsthand, but are also symbolised in hate targeting men, further stigmatising women and girls.

However, hate speech does not exist in isolation but intersects with ethnicity, religion, and political affiliations, creating complex and multi-layered abuse that reinforces and intensifies existing tensions. Colourism and anti-Blackness were identified as underexplored but pervasive issues. Political events and internal conflicts fuel 'ethnic and gendered' hate speech, while religious influences and narratives are used to discredit and demonise women, and justify misogyny. These findings extend beyond women and girls, indicating that inflammatory political discourse, historical references and terms like "banda", "traitor", "infidel" and "devil" are weaponised to incite real-world violence, as seen during the height of hostilities between the Tigrayan People's Liberation Front (TPLF) and Government forces.

Female and male content creators face distinct types of genre-specific abuse, often linked to industry norms and societal biases. Notably, women and girls discussing political issues encounter higher levels of aggression compared to other genres. These patterns highlight the need for a nuanced approach that considers both gender dynamics and the cultural context of each field.

There are both platform-specific and overarching trends in how women and girls are targeted online, signalling the need for both technological solutions and societal shifts to challenge and dismantle the norms that sustain gender-based violence online. This study highlights a crisis in content moderation and regulatory enforcement, with a widespread lack of trust in platforms' ability to tackle online abuse effectively. An improvement in content moderation, stronger enforcement of regulations, and greater accountability from social media platforms, especially in countries with a diversity of languages, is necessary. It requires both broad and platform-specific interventions, as different platforms facilitate and amplify hate speech in distinct ways. For example, TikTok's algorithm-driven virality and YouTube's capacity for long-form political discourse create distinct challenges in combating hate speech.

Finally, the persistent and misguided view that the online space is separate from the 'real world' must be overcome. Online and offline spaces are intrinsically linked; it's clear that offline discourse impacts online activity, and online harms do not stay

online, their impacts are both significant and far-reaching, psychological and physical.

2. INTRODUCTION

Social media and digital technologies are shaping how people connect, communicate, and engage with the world. As more personal and professional interactions take place online, the rise of TFGBV has become increasingly apparent. In Ethiopia, TFGBV takes many forms, including hate speech, harassment, and revenge pornography, reinforcing harmful gender norms and restricting women's voices. The impacts extend beyond individual harm, creating a chilling effect that discourages women's participation in public life.

Online abuse is shaping who gets to speak, lead, and exist safely in digital spaces. Ensuring that women and girls can participate safely and freely in digital spaces is not only a matter of online safety – it is essential for their full inclusion in public life. CIR has been at the forefront of efforts to understand and address these issues through research, workshops, and a conference in Addis Ababa.

CIR's 2024 report – a comprehensive analysis of gendered abuse on Facebook, Telegram, and X in Ethiopia – brought critical issues to light and underscored the need for further investigation. Building on these findings, this study expands the scope of analysis to gendered hate speech on YouTube and TikTok, offering fresh insights into the evolving challenges of TFGBV in Ethiopia's digital landscape. By shedding light on the patterns, narratives, and intersectional dimensions of online abuse, this report aims to inform meaningful interventions, strengthen advocacy, and drive policy change to create safer online spaces for all.

2.1 WHY CONTINUE RESEARCHING TFGBV?

TFGBV is the latest iteration of an age-old problem: gender-based violence (GBV). It reflects and reinforces deep-seated societal attitudes, such as misogyny and cultural biases, and often escalates from subtle microaggressions to physical harm. It creates an environment of fear and exclusion, deterring women and girls from participating in public life, both online and offline. This exclusion can have devastating impacts; silencing and ostracising women and girls from public spaces hinders informed decision-making, leading to less representative public spaces and democratic processes, amplifying cycles of marginalisation.

> It's terrifying that no matter how many 'advancements' we make GBV finds new ways to follow

WORKSHOP PARTICIPANT, 2025

CIR has laid a strong foundation by exploring the multifaceted nature of TFGBV in Ethiopia and carrying out a deep dive into gendered hate speech as one form of TFGBV. CIR's research provided valuable insights through interviews with survivors, analyses of social media content, and the creation of a publicly available lexicon of inflammatory keywords in multiple languages. CIR also published academic work, including a paper on its innovative methodology for annotating hate speech in indigenous African languages. These efforts, informed by workshops and roundtables in Addis Ababa, formed recommendations that are already influencing advocacy and training. For instance, several Non-Government Organisations (NGOs) are using CIR's findings and recommendations in their programming and advocacy work.

There were calls for continued research on additional platforms to broaden understanding. For example, discussions with Ethiopian stakeholders revealed a need for further investigation into TFGBV on platforms like TikTok and YouTube - areas yet to be explored despite their societal significance. This project seeks to fill this data gap and empower CSOs and policymakers in Ethiopia. The findings point to the need to improve content moderation policies, particularly for underrepresented languages. By continuing its research and fostering collaboration among Ethiopian stakeholders, CIR seeks to raise awareness and drive actionable change against TFGBV, ensuring safer digital spaces for women and girls.

> Technology is the present and the future. We have to create a safe space for people to use it

WORKSHOP PARTICIPANT, 2025

2.2 PROJECT AIMS

- Strengthen the evidence base on TFGBV in Ethiopia.
- Better inform government institutions, NGOs, social media companies, and the public about TFGBV in Ethiopia.
- Empower CSOs and government institutions in Ethiopia with practical recommendations on addressing TFGBV.

2.3 RESEARCH QUESTIONS

Following discussions with Ethiopian stakeholders, CIR has expanded its quantitative study on Facebook, Telegram and X to two additional platforms: YouTube and TikTok. This research will look at the following questions in relation to those platforms:

1. Forms (type and sentiment): What 'type' and 'sentiment' of gendered hate speech are prevalent?
2. Intersectionality: Does hate speech vary when multiple protected characteristics are targeted?
3. Platform variation: Does gendered hate speech vary by social media platform?
4. Relationship with content: Does the hate speech relate to the video content?

2.4 RESEARCH SCOPE

To expand on its existing work, CIR used the same quantitative methodology and lexicon from its previous study and applied it to content from YouTube and TikTok, and conducted cross-platform comparisons. Due to the different platform formats (namely, video and comment interactions), CIR included an additional research question that explores the relationship between the video content and the comments.

This research uses the Ethiopian Government's definition of hate speech, as set out within the 'Hate Speech and Disinformation Prevention and Suppression Proclamation', to ensure relevance to the Ethiopian context as much as possible. Due to resource and time constraints, four languages were selected for analysis due to their prevalence on Ethiopian social media: Amharic, Afaan Oromo, Tigrigna and English. Future studies could expand to additional languages.

Gendered hate speech encompasses many rhetorical forms of TFGBV. Future research could investigate other forms of TFGBV, including the use of imagery to spread hate, revenge pornography and case studies on prominent individuals.

3. METHODOLOGY

CIR expanded its 2024 study into online gendered hate speech in Ethiopia (on Facebook, Telegram and X) to two additional platforms: TikTok and YouTube. The same, academically peer-reviewed methodology was used to maximise the comparability of the findings across the platforms.

CIR used keyword matching and hate speech detection models to sample posts from YouTube and TikTok. Data collection took place between September and December 2024, although posts were collected since the channel or account's creation. A team of researchers annotated the comments according to three dimensions: hate speech target, type, and sentiment (see: appendix 7.3, annotation protocol; or section 3.1, conceptual framework).

Natural Language Processing (NLP) techniques formed a key component of the methodology for analysing and interpreting textual data from social media. This involved creating a lexicon, gathering data from social media platforms, and preprocessing the data, which included tasks like eliminating duplicates and anonymising posts. CIR also applied NLP methods to partially automate the detection of hate speech in the English dataset, refining the sample for manual annotation. Lastly, NLP techniques were used during the manual data annotation process through the development and utilisation of a labelling schema. The following sections detail each of these steps.

3.1 CONCEPTUAL FRAMEWORK

CIR adopted its existing conceptual framework to assess the key components of hate speech: target, type, and sentiment (see figure 1 below). Having a clear **'target'** and **'hate type'** is essential for a piece of content to classify as hate speech. CIR annotators only classified content as 'hate speech' if the target of the hate speech (a protected identity group, such as gender) and the type of hate speech could be clearly identified. Additionally, CIR included an assessment of **'sentiment'** of the hate speech, for a richer analysis.

The **hate targets** included are those found within the Ethiopian Government's Hate Speech and Disinformation Proclamation. Although the research focussed on gendered abuse, other protected characteristics are included to enable an assessment of intersectional abuse (hate speech with multiple targets, such as gender and ethnicity).

In the framework, the **'hate type'** refers to the method of abuse (such as **'threats'**), while **'sentiment'** refers to the emotive or linguistic qualities of the abuse (such as **'mockery'** or **'stereotyping'**). For example, hate may be conveyed using nuances in

language, such as sarcasm, irony, or satire. Each category is defined in figure 1 below.

This conceptual framework underpins the research methodology, including the data annotation protocol, and is essential to interpreting the research findings. To understand more about the conceptual framework, read CIR's 2024 report. The conceptual framework has been academically peer-reviewed and published in the Resources for African Indigenous Languages Journal.

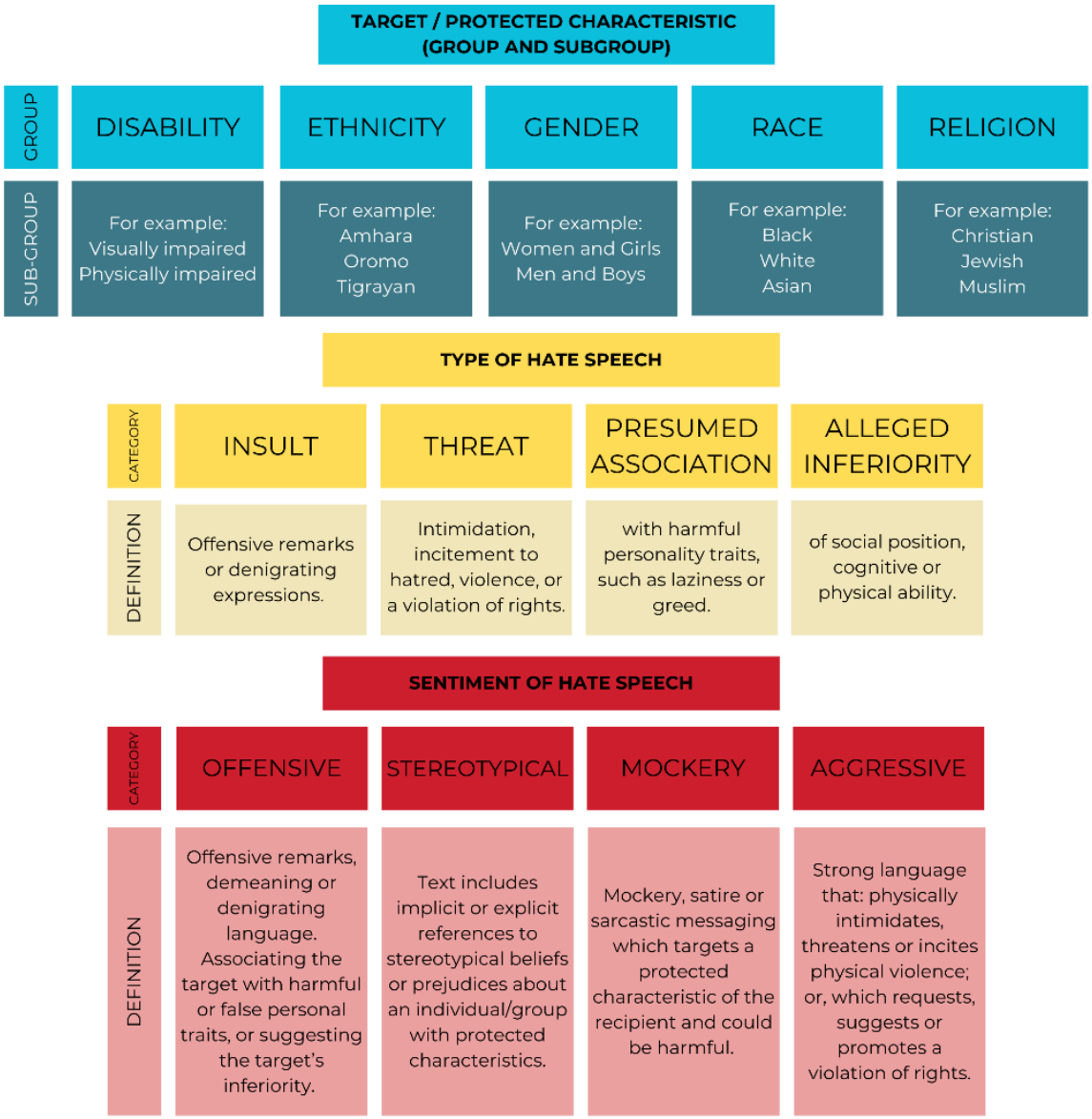


Figure 1: CIR's Conceptual Framework. This underpins the research design and annotation schema, as found in the Annotation Protocol in section 7.3.

3.2 LEXICON DEVELOPMENT

CIR developed a lexicon of 2,058 inflammatory keywords across four languages (Amharic, Afaan Oromo, Tigrigna, and English), which may be indicative of hate speech along gendered, ethnic, and religious lines. CIR believes that this is the most comprehensive lexicon at present for the Ethiopian context.

The lexicon was developed through desk-based research (identification and refinement of existing hate speech lexicons), the identification of keywords and narratives during CIR's semi-structured interviews, and a roundtable of experts in Addis Ababa in July 2023. A full list of resources can be found in the bibliography. CIR used the same lexicon in the investigation into TFGBV on Facebook, Telegram, and X. Note that the presence of one of these keywords alone does not signal that the text is hate speech. Instead, it is used to flag potential posts, which are then manually labelled.

3.3 ACCOUNT SELECTION

CIR engaged Ethiopian social media experts to curate a list of widely popular and influential YouTube channels and TikTok accounts in Ethiopia. The comments on these channels' and accounts' videos were extracted for analysis.

ACCOUNT CATEGORISATION

CIR's social media experts categorised the channels and accounts by genre. This enabled an exploration of the relationship between the video genre and the comments. The channel categories are shown in figure 2 below. Due to the hundreds of hours of content posted online, our team was unable to code each individual video. Instead, an overarching assessment was made on the content of each channel or account from a sample of videos. Multiple content categories could be selected.

Additionally, the team identified whether the channel or account is a known sender or receiver of abuse and the gender presented by the account or channel holder. This was done to establish whether accounts that are known to engage in hateful discourses receive unique patterns of hate themselves. CIR has not undertaken further analysis to determine if the presented gender is correct.

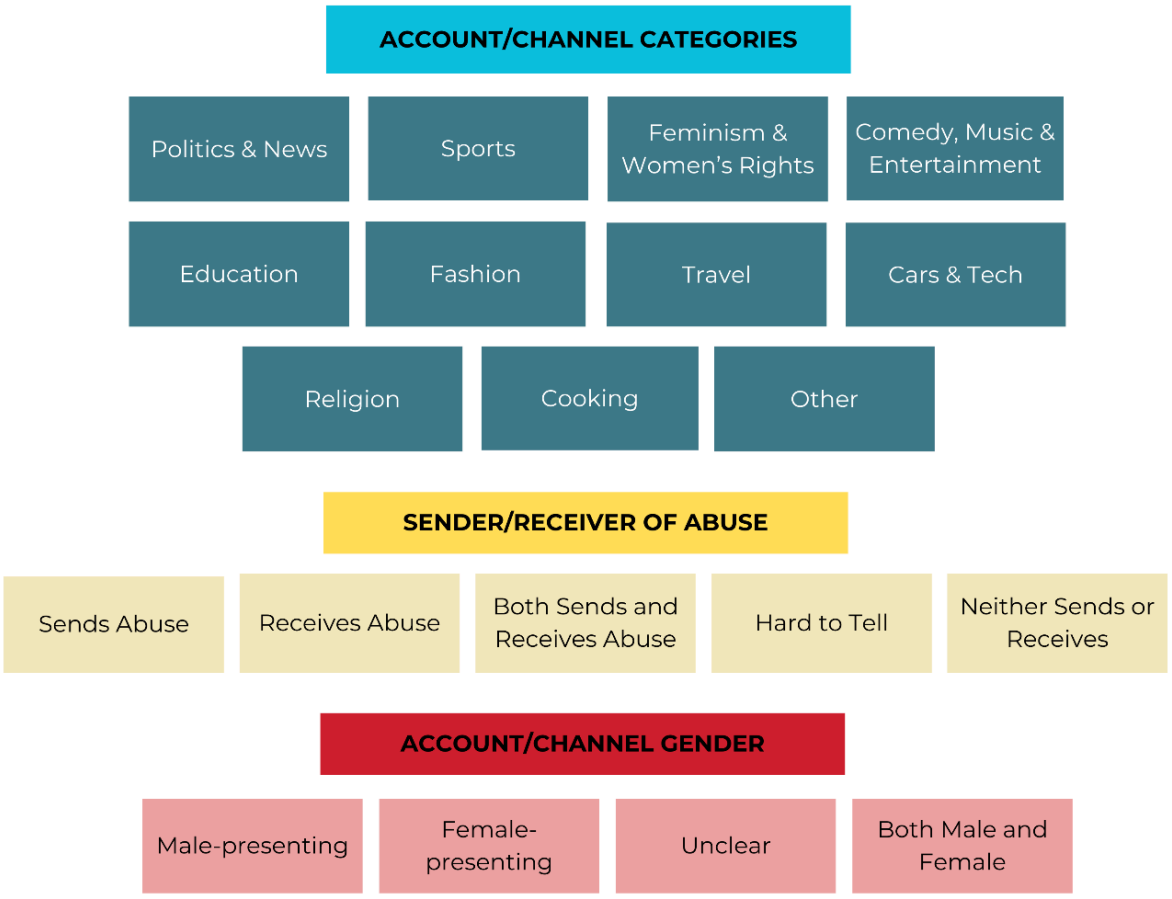


Figure 2: The categorisation of accounts and channels by CIR.

3.4 DATA COLLECTION AND PROCESSING

CIR used keyword matching and hate speech detection models to sample posts from YouTube and TikTok. Data collection took place between September and December 2024, although posts were collected from the date of the channel or accounts creation.

3.4.1 COLLECTION

YOUTUBE

CIR used an open-source Python library to collect comments from 160 YouTube channels. The script collects approximately 60-80% of all comments on videos in the candidates' YouTube channels. Using this approach, CIR collected 7,810,908 YouTube comments.

TIKTOK

CIR used the open-source platform 'Open Measures' to acquire comments on 364 TikTok accounts. Using this approach, CIR collected 1,515,933 TikTok comments. Open Measures collects comments on the most recent 1,000 videos and has a limit of 10,000 comments per video. CIR believes this limitation has not restricted analysis as CIR deems that no video within the dataset had more than 10,000 comments.

As the comments were retrospectively collected for both platforms, posts that breached the platform's policies may have already been taken down. This means that some of the most harmful online rhetoric could be missing from the sample.

3.4.2 DATA PRE-PROCESSING

CIR took several steps to pre-process the data, including cleaning, transforming, and organising raw text data to make it suitable for analysis and modelling. This was essential to improving the quality and reliability of language models and to extracting meaningful insights from the textual data. The key tasks are outlined below:

TEXT CLEANING:

Text data collected from social media often contains irrelevant or erroneous information that complicates analysis or interpretation. Data cleaning tasks included:

- Removing Hypertext Markup Language (HTML) tags and special characters. HTML tags are labels that tell browsers how to display information, and special characters are extra symbols in text. HTML tags and special characters are useful for formatting webpages or changing the visual representation of text but are not needed when processing text.
- Converting texts to lowercase to ensure case insensitivity. Lowercasing text ensures uniform treatment of words regardless of case. For instance, "apple", "APPLE", and "APPle" become equivalent to "apple" after conversion.
- Removing or replacing punctuation.
- *As emojis can be used to add emphasis or emotion to messages, they were not removed during these steps.*

The following cleaning tasks were done for the English and Amharic language datasets only. CIR carried out these steps to prepare the textual comments for automated hate speech classification (see section 3.5):

- Handling or removing numerical values, dates, and other non-textual information.
- Removing stop words (common words like "and," "the," and "in" that do not carry significant meaning).
- Handling and normalising abbreviations and acronyms. Abbreviations and acronyms are often used to conceal or disguise harmful content. Thus,

converting all variations of popular words to a single form helps improve the performance of hate speech detection models.

3.4.3 SAMPLE FOR ANALYSIS

CIR randomly sampled comments from the full dataset, ensuring the sample accurately represented the dataset. This was achieved by selecting a balanced number of comments across the languages and the lexicon.

3.4.4 DATA ANONYMISATION

CIR anonymised the usernames in posts in line with ethical requirements to protect the privacy and confidentiality of individuals whose data was used for research and analysis. This was done by replacing all usernames (i.e. any word appearing after the "@" symbol) with the word "USERNAME".

3.4.5 PROPORTIONAL DATA

Due to the sampling methods for analysis, after annotation, the data was turned into proportional data: relative amounts of hate speech for each category, shown in percentages. This allows a comparison of the types and sentiments of hate speech, despite the different total numbers of hate within the dataset.

3.5 CLASSIFICATION AS HATE

CIR streamlined the identification of hate speech by adding an automated classification step before manual annotation. CIR fine-tuned and deployed a machine learning model (based on transformers) to classify whether the post contained hate speech or not. CIR filtered out non-offensive content from the analysis.

CIR only applied this step to English and Amharic due to the dearth in pre-trained models for Tigrigna and Afaan Oromo. Resultantly, the identification of hate may have been easier in English and Amharic.

3.5.1 ANNOTATION

CIR's team of experienced annotators manually annotated 17,082 comments, following the annotation protocol (see appendix 7.3). The team's linguistic expertise included all four languages in the study: Afaan Oromo, Amharic, Tigrigna and English. Ensuring consistency across the multi-lingual team was key, with CIR's conceptual framework providing clear definitions of hate speech, its targets, types, and sentiment (see section 3.1). The protocol includes examples of how to interpret the conceptual framework.

Using Doccano, an open-source tool, annotators tagged the comments, labelling hate speech details such as targets (gender, ethnicity, religion, race, disability), type, and sentiment (see section 3.1). Categories were not mutually exclusive, allowing for multiple selections (e.g., aggressive and mocking hate speech).

3.6 WORKSHOPS

CIR held a series of workshops and meetings with stakeholders in August and November 2024 and January 2025, both virtually and in person, in Addis Ababa. The sessions provided a forum to share and discuss preliminary findings and narratives and explore possible recommendations.

The workshops were 60% women and 40% men, and individuals represented multiple human rights related fields, including women's and girls' rights, gender-based violence (both online and offline), and online safety. CIR wanted to engage with subject matter experts and hear different perspectives about how to research and respond to these issues in Ethiopia. Discussions during workshops have been anonymised to protect the identities of the individuals involved.

3.7 LIMITATIONS

Where possible, CIR took measures to reduce the limitations of the methodology and the findings. However, several resource, context, or result limitations must be considered, which have been outlined below.

LANGUAGES

Over 80 languages are spoken in Ethiopia. While not all these languages are represented on social media, it was only possible to select four languages for this study due to resource and time constraints. As a result, four languages that are prevalent on social media were selected: Amharic, Afaan Oromo, Tigrigna and English. The findings, therefore, represent a sample of speakers of each of the four languages on social media. Despite this, carrying out the study in four languages was resource intensive, requiring skilled researchers who were familiar with the languages, context, and social media analysis. Future studies could adopt CIR's methodologies to conduct similar studies on different languages.

DATA COLLECTION

Tailored data collection methods were made for each platform to ensure that CIR complied with each platform's terms and conditions. While necessary, this resulted in the collection of different amounts of data from each platform. As a result, CIR could not compare how prevalent gendered hate speech is across YouTube and

TikTok. Instead, proportional data was used to assess the type and sentiment of hate speech on each platform.

Initial data collection returned far fewer comments in Tigrigna from TikTok than the other three languages. To see if CIR had underrepresented accounts with Tigrigna commentators, efforts were made to actively look for additional accounts. However, this did not drastically increase the amount of Tigrigna in the dataset. As CIR took even samples from each language dataset, this did not impact the representation of Tigrigna in the study. The lack of TikTok comments in Tigrigna could perhaps reflect the current content on TikTok, as this was not the case in our YouTube dataset. Workshop participants suggested this could also be the result of limited internet connectivity and infrastructure in the Tigray region, coupled with the larger data requirements of TikTok.

To compare the findings with the previous study, textual analysis of the comments on TikTok and YouTube videos was conducted. CIR also assessed each channel or account to determine the types of content the user shared, but full video analysis was not possible due to the time it would take to analyse each video. Future research could investigate the content within the videos themselves.

DATA PROCESSING

Prior to sampling the dataset, CIR used a pre-trained machine learning model to filter out posts that the model classified as 'not hate', thus increasing the chances that CIR would find hate speech to analyse in the dataset. It is important to note that due to the lack of pre-trained models in Tigrigna and Afaan Oromo, this pre-classification step was applied exclusively to posts in English and Amharic.

The absence of pre-trained models is partly due to the scarcity of annotated datasets available for model training in these languages. However, CIR's annotated datasets in these languages could be leveraged to train machine learning models for future studies. In addition, pre-trained models have the potential to automate the identification of hate speech on social media platforms. This could be used for platform content moderation and may lead to a reduction in hate speech across languages.

MEETING THE THRESHOLD FOR HATE SPEECH

The rigorous application of our annotation protocol, while ensuring the integrity of the dataset, meant that hateful content was excluded due to insufficient context or ambiguity. For example, many posts contained abusive or insulting language but lacked a clearly identifiable targeted protected characteristic, thus excluding them from the dataset. Future studies could explore online discourses that exist on the

cusp of hate speech to provide deeper insights into the broader dynamics of harmful language online.

Additionally, researchers observed a significant volume of political hate speech in Ethiopia's social media landscape, which was excluded because it fell outside the Ethiopian government's definition of hate speech. Similarly, while posts targeting media organisations, such as "ESAT is #1 enemy of the Oromo people," were prevalent, they did not meet the criteria for inclusion unless they specifically targeted individuals, such as journalists, in a way that satisfied the hate speech definition.

DUALITY OF LANGUAGE

A notable research challenge was the duality of certain words, which complicated efforts to determine whether comments constituted hate speech. For instance, the term "bitch" is commonly recognised as a gendered slur, but was sometimes used positively, such as in phrases like "Go bitch!" or "Queen of bitches." These nuances underscored the critical importance of context in interpretation, highlighting the value of qualitative research and the expertise of data annotators in making informed assessments.

IDENTIFYING RELEVANT DATA & THE IMPACT OF THE DIASPORA

The TikTok data set comes from exclusively Ethiopian accounts; however, while many of these accounts identified as 'Ethiopian', it is possible that they were Ethiopian Diaspora. CIR analysed the data and identified diaspora accounts to omit from the sample. This was particularly apparent in the English language content. The data indicated that many of the influencers were located in the US, Canada and France since many of them were commenting on the political affairs within these countries. For example, many comments referred to an incident where American comedian, Kathy Griffin, had posed in a photo with a fake head of President-elect Donald Trump. Similarly, many of the comments referenced a popular American politics influencer, Harry Sisson, demonstrating the links between Ethiopian and American influencer spaces. CIR refined the dataset, removing words with the following terms: Trump, Walz, Harry, JD, Vance, Vivek, Democrats, and Demonrats. These posts were omitted from the dataset to improve the dataset's relevance.

4. RESULTS AND DISCUSSION

This investigation aimed to deepen the understanding of TFGBV in Ethiopia by broadening CIR's previous analysis on gendered hate speech with the inclusion of two additional popular platforms: YouTube and TikTok. Building on CIR's earlier

research, this investigation focuses on the **types of hate** speech women and girls face (section 4.2), focusing on the written methods employed by abusers and the **sentiment** behind the hate speech – whether **offensive, stereotypical, aggressive, or mocking**. Additionally, CIR compared the abuse faced by women and girls with that targeting men and boys within the dataset (section 4.3).

CIR also investigated trends in **intersectional hate speech** (section 4.4) to understand how abuse changes when women and girls are targeted not only for their gender but also for other protected characteristics, such as ethnicity or religion. By expanding the dataset to include other **hate targets**, CIR was also able to compare how abuse manifests when directed at women and girls or when directed at individuals for their ethnic, religious, or racial identities (section 4.5).

Furthermore, CIR analysed hate speech across different content **genres** (section 4.6), revealing clear gender-based patterns, with female and male content creators facing distinct types of abuse.

During workshops, participants shed light on their experiences on TikTok and YouTube and the platform specific dynamics that they felt contributed to hate speech on those platforms. To set the scene, the analysis below will set out these findings (section 4.1), before diving into the dataset. Insights from CIR's earlier study, as well as findings from interviews, roundtables, and workshops, were also integrated to add depth to the analysis. The following analysis uses proportional data, to focus on the nature of the hate speech and narratives, rather than purely the scale.

4.1 THE PLATFORMS

Workshop discussions shed light on the distinctive ways abuse manifests on TikTok and YouTube and suggested reasons for these variations. The views presented in this section are the workshop participants' and have not been independently verified by CIR.

The participants reported that TikTok's design fuels polarisation, and YouTube's design supports politically motivated, ethnic and gendered abuse. A lack of trust in content moderation, recent content moderation policy changes, and the ease of cross-platform dissemination of content signals the need for both broad and platform-specific strategies to tackle online abuse and create safer digital spaces for women and marginalised communities. Despite existing guidelines against hate speech, enforcement remains weak – stronger action from platforms is essential.

4.1.1 TIKTOK

Workshop participants expressed significant concerns about the prevalence of hate speech and harmful content on TikTok. They perceived TikTok as the most harmful platform, citing high levels of violence, hate speech, sexualised content, and threats. Given TikTok's popularity among women and human rights defenders, participants noted that abuse on the platform is often more personal and direct compared to other social media sites.

TIKTOK'S DESIGN FUELS POLARISATION

TikTok is a highly polarised platform in Ethiopia. During manual labelling, CIR identified many positive interactions and signs of camaraderie, such as "Yes, queen" and "You go, girl". At the same time, there was significant hate speech and objectification. Workshop participants said that TikTok's design incentivises polarising content to attract followers. Many TikTok creators focus on attention-grabbing, often controversial, content to grow their audience, exacerbating the platform's negative dynamics. As a result, the participants felt that TikTok content creators are becoming increasingly polarising and offensive, and content and comments are becoming more widely accepted. For example, content often frames discussions as men versus women, reinforcing and amplifying harmful stereotypes. Additionally, even humorous content on TikTok frequently incorporates hateful rhetoric, normalising discriminatory attitudes. Overall, participants viewed TikTok's content ecosystem as one that simultaneously promotes division and cohesion. This encourages personal attacks and exacerbates online hate speech and harassment.

4.1.2 YOUTUBE

Workshop participants anticipated that YouTube would host more misinformation, disinformation, and smear campaigns, particularly against public figures. They noted that YouTube's content often involved more political discourse, which could fuel more aggression and abuse targeting well known figures.

YOUTUBE HOSTS MORE POLITICISED, ETHNIC AND GENDERED HATE

Overall, CIR found more politicised, ethnic and gendered hate on YouTube. The hate was slightly more aggressive or threatening in nature than on other platforms. Workshop participants believed that YouTube's longer-form content allows for deeper, more extensive discussions, which can result in harsher and more aggressive abuse. The often political nature of discussions allows for deeper, more entrenched ethnic or gendered narratives to surface. For example, CIR found politically motivated ethnic or gendered abuse against notable figures, such as Abiy Ahmed, Shimelis Abdisa, as well as Betty, from 'the Betty Show' and Adenech Abebe.

When asked why there might be higher levels of aggression, workshop participants suggested that YouTube might be more aggressive because of the sensationalism and misinformation that is on offer on the platform, as it can be triggering.

4.1.3 PLATFORM USERSHIP

While not all areas have equal access to social media, workshop participants claimed that this divide between rural and urban areas is even more pronounced for TikTok and YouTube due to the higher data requirements for video streaming platforms. Additionally, certain areas of the country – particularly those recovering from or within ongoing conflict – suffer from internet blackouts and internet infrastructure damage, limiting further access.

4.1.4 CONTENT MODERATION

Although 93% of TikTok users and 87% of YouTube users during workshops reported seeing gendered hate speech on the platforms, only 47% and 40%, respectively, reported it. The rest either 'did nothing' or 'told their friends'. Discussions on the reporting mechanisms and wider content moderation resulted in the following themes being raised: ineffective moderation, lack of trust, the shrinking civic space in Ethiopia, and the impact of cross-platform dissemination.

INEFFECTIVE MODERATION, RESULTING IN LACK OF TRUST

A recurring concern in the workshops was the ineffectiveness of content moderation on TikTok and YouTube. Many participants felt that reported content often wasn't removed, and some participants had stopped reporting altogether due to frustration with the platforms' lack of action. Even those that continue to report harmful content reported a growing sense of resignation. One participant told CIR that their TikTok account was cloned by an impersonator. They reported it to TikTok but received a message saying that the account didn't breach their guidelines. The impersonator account was still live during the workshops in January 2025.

Multiple participants said that they'd reported really violent or sexual content on TikTok and received responses stating, "the content reviewed does not appear to violate our Community Guidelines". There was a shared belief that this was not good enough. It signals that there is a significant problem with either their policies or methods for assessing the content. Workshop participants expressed a belief that these platforms do not have the appropriate tools or personnel to moderate in the many different Ethiopian languages represented online. This has resulted in a lack of trust in content moderation.

THE IMPACT OF CROSS-PLATFORM DISSEMINATION

While it is important to think about each platform in isolation due to the unique ways they function, workshop participants expressed their concerns about cross-platform pollination of hate. Harmful content originating on TikTok and YouTube can easily be shared across platforms like X, Instagram, Facebook, and Telegram and vice versa. This limits the effectiveness of content moderation on individual platforms, highlighting the need for a more comprehensive approach to regulating harmful content across the entire online ecosystem.

Despite platform guidelines against hate speech, enforcement remains inadequate, and platforms must take stronger measures to protect users, especially women and marginalised communities.

CHANGES TO CONTENT MODERATION

Changes to content moderation policies by X and Meta have left workshop participants fearful for the future. For example, X's reduction in content moderation and focus on free speech following its acquisition (October 2022), and Meta's removal of third-party fact-checking (in January 2025). This means that the sites are more reliant on automated content moderation efforts, which are less effective in low resource languages, such as Amharic, Afaan Oromo, and Tigrigna. Workshop participants are concerned that YouTube and TikTok will follow the trend. Regardless, the levels of cross-platform dissemination of content means even if they do not, there may be knock on effects on YouTube and TikTok.

While this change has allegedly been in the interest of free speech, the workshop participants said they believed this would lead to increasingly polarised online discourse, and the spread of more mis/disinformation in already vulnerable information environments. These concerns are not restricted to the workshop participants, with many commentators reflecting the same fears, internationally. This study alone reveals that existing moderation practices are not sufficiently keeping women and girls, as well as other identity groups in Ethiopia, safe.

4.2 GENDERED HATE SPEECH TARGETING WOMEN AND GIRLS

This study uncovers both platform-specific trends and broader patterns in how women and girls are targeted online. They faced attacks on their dignity and worth (**insults**), messages implying cognitive, social, or physical inferiority, and degrading stereotypes. Mockery was also prevalent, where abusers used sarcasm and satire to undermine women. While threats and aggressive speech are less frequent, they pose a serious risk to women's safety on and offline.

While the **'hate types'** were fairly consistent across YouTube and TikTok, there were notable contrasts in the **'sentiments'** of hate speech. Additionally, a comparison with CIR's earlier study reveals variations across Facebook, Telegram, X, TikTok, and

YouTube, highlighting the need for both broad strategies and platform-specific interventions to create safer spaces for women and girls.

CIR identified key narratives within the abuse, including gendered stereotypes and criticisms related to women's role in society. Women who challenge social norms, including those in leadership roles or women's sports, were sexualised and discredited and faced stigmatisation and insults. Feminism was frequently discussed and was framed as a threat to traditional values, and feminists were discredited.

This section first deep dives into the data findings, exploring the key trends in **'hate types'** and **'sentiments'** before unpacking the narratives that reinforce this harmful content.

4.2.1 TYPE OF HATE SPEECH TARGETING WOMEN AND GIRLS

This investigation delved into the distinct **types of hate** speech targeting women and girls within the dataset, revealing a spectrum of abuse: from **insults** and **threats** to **accusations of inferiority** or **associating** the female gender with harmful characteristics (see figures 3 and 4 below). These categories highlight the multifaceted nature of gendered hate speech online.

	TYPE OF HATE SPEECH			
CATEGORY	INSULT	THREAT	PRESUMED ASSOCIATION	ALLEGED INFERIORITY
DEFINITION	Offensive remarks or denigrating expressions.	Intimidation, incitement to hatred, violence, or a violation of rights.	with harmful personality traits, such as laziness or greed.	of social position, cognitive or physical ability.

Figure 3: An excerpt from the contextual framework and annotation protocol, defining the different types of hate speech.

Insults – encompassing a range of degrading and denigrating language – were the most prevalent type of hate speech, making up 44.9% of the analysed content. This category captured the verbal attacks aimed at undermining women's dignity and worth, often framed in harsh, dismissive tones that seek to silence, shame, or degrade.

The second most prominent type, **alleged inferiority,** accounted for 25.8% of the dataset. This category included rhetoric asserting the supposed inferiority of women and girls, targeting their social roles, cognitive abilities, or physical capabilities. These remarks not only perpetuate harmful stereotypes but also reinforce existing inequalities, portraying women as lesser in virtually every domain of life.

Presumed association – encompassing claims linking women to negative traits such as greed or laziness – formed 19.8% of the hate speech identified. These comments rely on stereotypes to discredit women by associating them with undesirable behaviours. This can overtly or subtly erode women and girls' reputation and credibility (collectively or individually).

Though proportionally smaller, **threats** represented 9.6% of the hate speech. This category included intimidation, incitement to violence, and threats to violate individuals' rights. While less frequent, the presence of such content signals a significant risk with the potential to escalate into real-world harm. For several reasons, CIR believe this is an under-representation of the sheer scale of threatening content on TikTok, including analysis of comments only (not hate within the video content), content moderation, and the retrospective data collection – meaning the most overtly threatening comments may have been taken down.

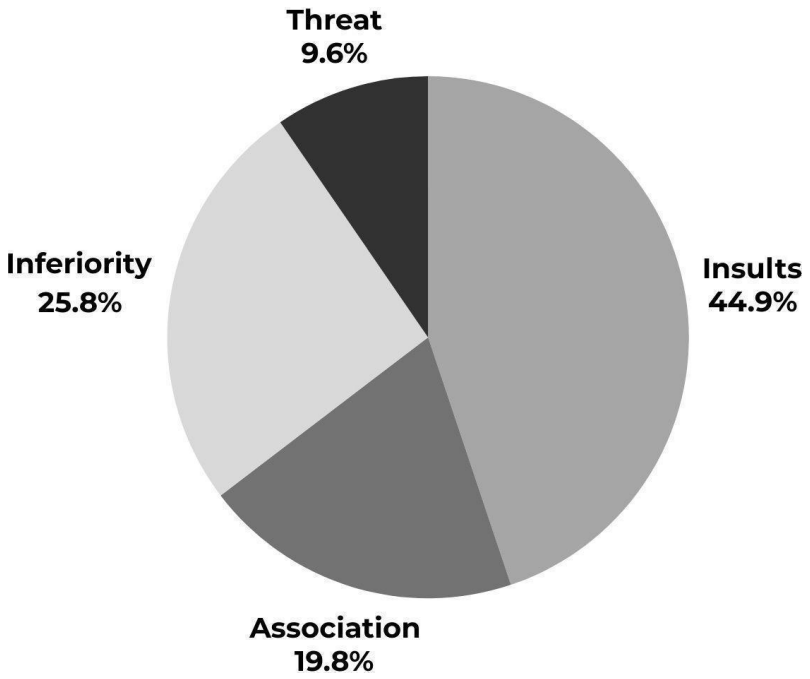


Figure 4: Pie chart showing the different types of hate speech targeting women and girls in CIR's dataset.

4.2.2 SENTIMENT OF HATE

CIR investigated the different **'sentiments'** of hate speech targeting women and girls. These findings illustrate the diverse and layered forms of hate speech women face online, offering a clearer understanding of how abuse manifests and the risks it poses. Researchers coded whether the social media content contained **aggressive** language, **offensive** language, **mockery** (at the expense of the target), or **stereotypes** (see figures 5 and 6 below).

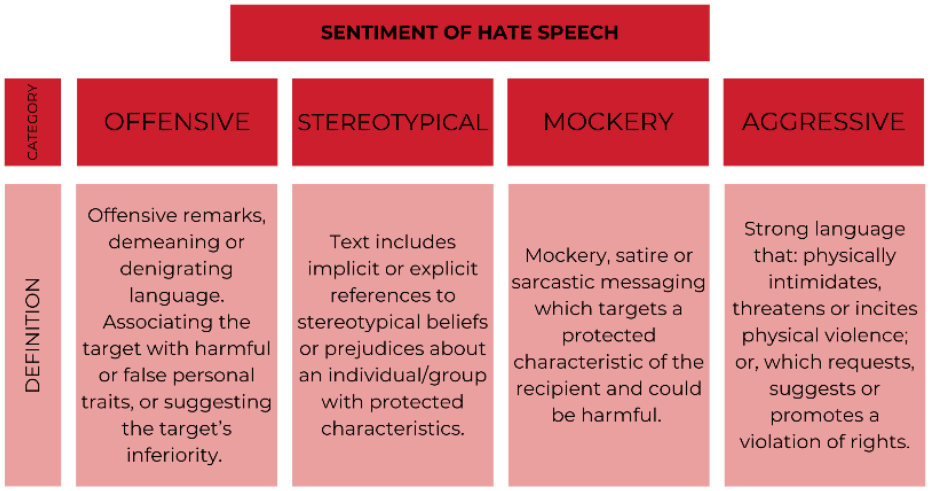


Figure 5: An excerpt from the contextual framework and annotation protocol, defining the different hate speech sentiments.

The most prevalent sentiments in the dataset were **offensive** language and **stereotypical** language, accounting for 35.7% and 32.6%, respectively. **Offensive** Language captures a broad spectrum of harmful speech, from outright insults and demeaning comments to more insidious associations that question the target's worth or reinforces their perceived inferiority. To classify as '**stereotypical**', the text contained implicit or explicit gender-based prejudices that perpetuate damaging clichés about women and girls or reinforce societal biases.

The third most prevalent sentiment, **mockery**, made up 20% of the hate speech. This often took the form of sarcastic remarks, jokes, or satirical comments, using humour as a weapon to ridicule or undermine women.

While comprising only 11.7% of the dataset, **aggressive** language represents the most explicitly harmful sentiment. These comments included threats, intimidation, and incitement to violence, aiming to silence and endanger women in digital spaces.

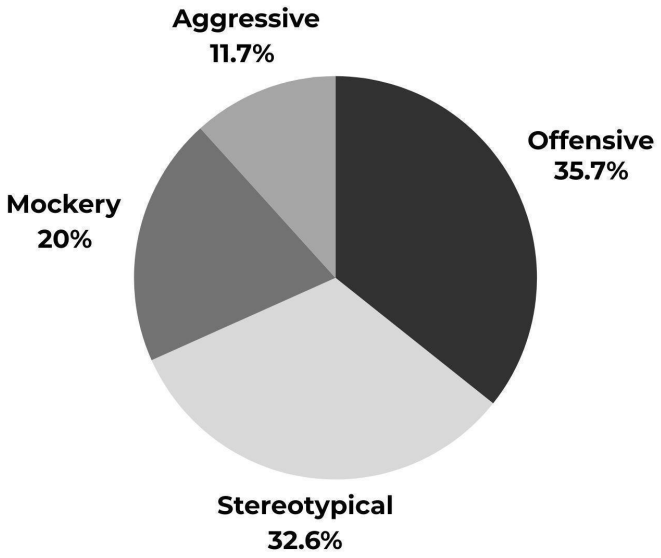


Figure 6: Pie chart showing the different sentiments of hate speech targeting women and girls in CIR's dataset.

4.2.3 PLATFORM SPECIFIC PATTERNS (HATE TYPE & SENTIMENT)

YOUTUBE AND TIKTOK

CIR's analysis uncovered subtle differences in how hate speech targeting women and girls manifests across TikTok and YouTube. While certain **'hate types'** and **'sentiments'** were distributed similarly, others showed notable platform-specific trends. Workshop participants felt these differences reflected the platforms' differing cultures and user behaviours.

TikTok had a higher prevalence of **'insults'** (47.4%) compared to YouTube (39.5%). It also had a significantly higher proportion of hate speech rooted in **'stereotypes'** that leverage gender-based clichés and prejudices (36.1% compared to 24.2%). A workshop participant suggested that TikTok's interactive, often punchy, comment-driven culture may perpetuate or amplify casual, derogatory remarks, while its fast-paced nature encourages impulsive comments and simplistic and stereotypical messaging.

YouTube had more hate speech centred on **'alleged inferiority'** (31.6%) compared to TikTok (23.1%). It also had more **'aggressive'** hate speech containing threats, intimidation, or incitement to violence (16.4% compared to just 9.8% on TikTok). Workshop participants suggested that the longer-form, opinion-heavy content on YouTube may provide more opportunities for users to engage in this type of discriminatory discourse, encouraging a more combative environment, where hate speech could escalate.

Interestingly, the proportions of **'threat'**, **'presumed association'**, **'mockery'**, and **'offensive'** speech were relatively consistent across both platforms, suggesting there is a shared underlying trend of stereotyping, ridicule, belittling, and intimidation across both social media spaces. These shared trends indicate a baseline level of harmful discourse that transcends platform-specific dynamics. Moreover, these findings highlight the need for tailored interventions, addressing the distinctive dynamics of each platform to create safer spaces for women and girls.

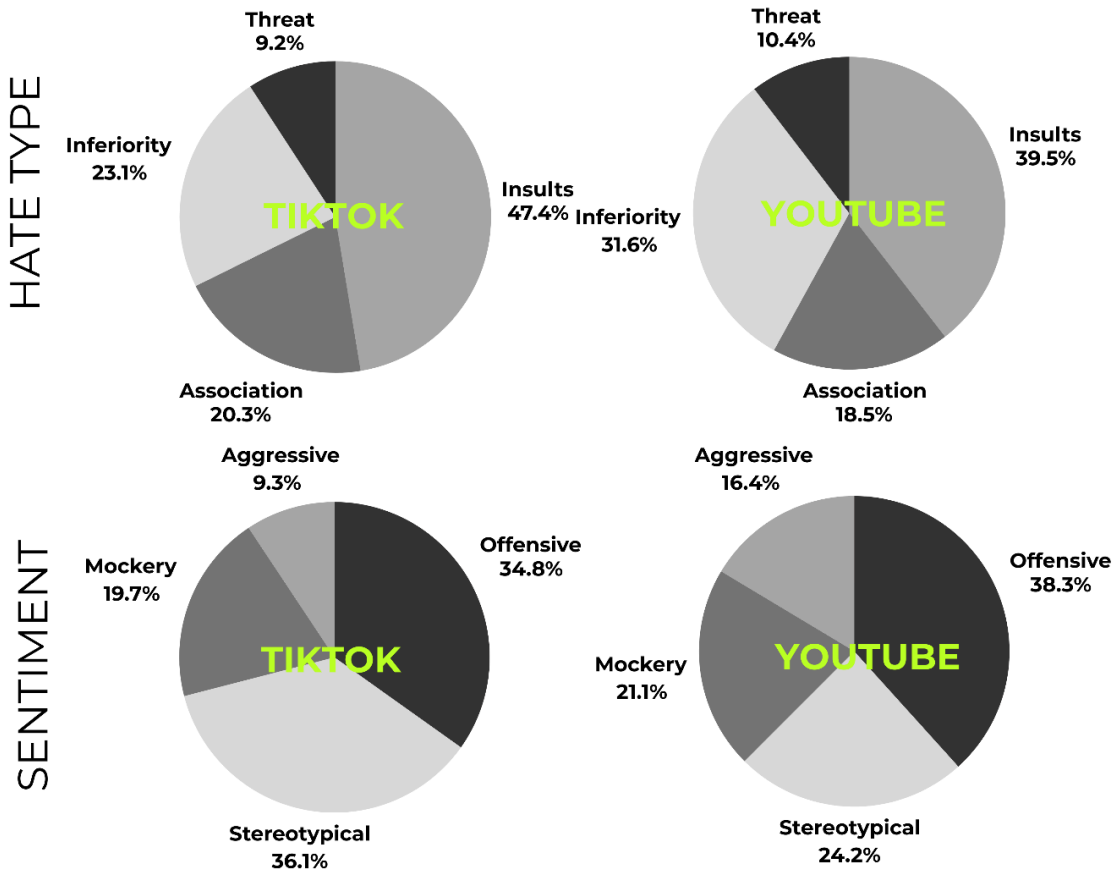


Figure 7: Pie charts showing the different types and sentiments of hate speech targeting women and girls on TikTok and YouTube in CIR's dataset.

FACEBOOK, TELEGRAM, TIKTOK, YOUTUBE AND X

As CIR used the same methodology in this investigation as its earlier study, it is possible to compare the findings across five major platforms: Facebook, Telegram, TikTok, YouTube and X. While certain trends were consistent, each platform demonstrated distinct patterns in how abuse manifested when targeted at women and girls.

Platform variations:

- Across all platforms studied, **offensive language** was the most prevalent sentiment of hate speech, with Facebook receiving proportionally more (55.7%) and TikTok the least (34.7%).
- Telegram recorded the highest proportion of **threatening** hate speech (15.6%), almost double that of TikTok (9.2%). This difference may be influenced by the platform's closed structure, where channel owners manage content moderation.
- Telegram also had the highest proportion of **mockery** (29.1%); X had the least (9.3%).
- YouTube had the largest proportion of **aggressive** speech (16.4%), almost three times more than Telegram (5.9%).
- YouTube also stood out for hate speech rooted in **alleged inferiority** (31.6%).
- X was the leading platform for **insult**-based hate speech (48%), narrowly surpassing TikTok (47.4%), while Facebook had the least (25%).
- Facebook exhibited the highest proportion of hate speech **associating** women with **unfavourable characteristics** (31.7%).
- TikTok had the highest proportion of **stereotypical** hate (36%), while Facebook had a comparatively lower share at 12.5%.

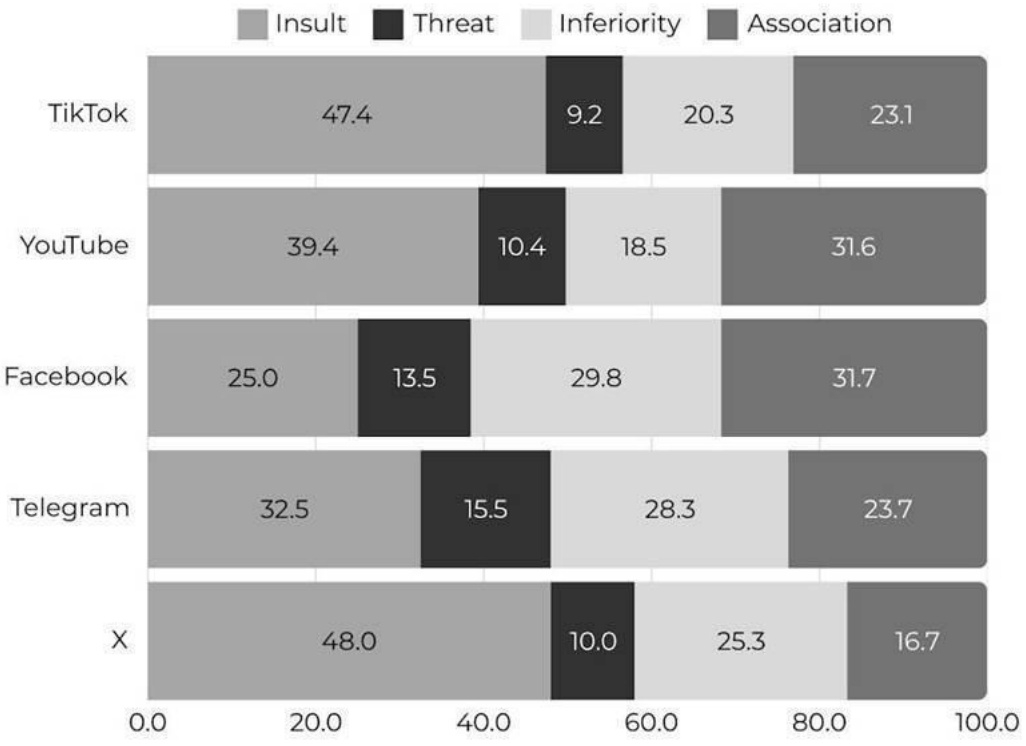


Figure 8: Bar chart showing the different types of hate speech targeting women and girls across all platforms in CIR's two studies.

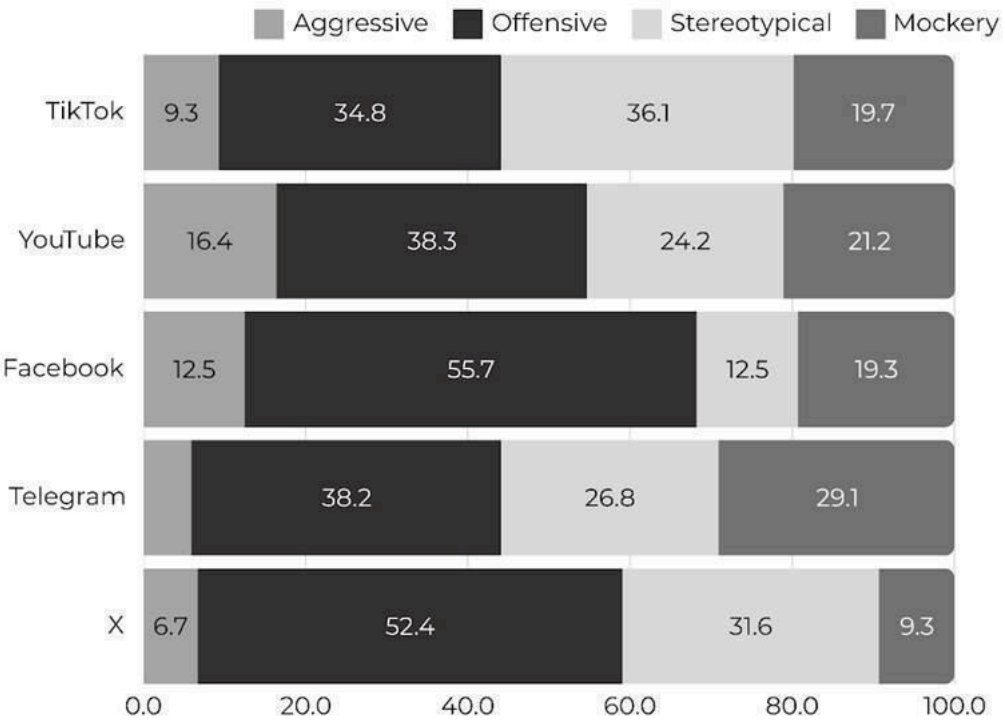


Figure 9: Bar chart showing the sentiments of hate speech targeting women and girls across all platforms in CIR's two studies.

4.2.4 NARRATIVES

As part of the annotation process, key narratives within gendered hate speech were identified and documented. These narratives were later examined in workshops, where discussions provided deeper insight into their meaning and impact (for more information on the workshops, see section 3.6). The workshop discussions revealed that many of these narratives are deeply rooted within socio-cultural and religious norms. Women often reproduce these narratives unknowingly, or knowingly. They can even remain invisible to those that are harmed by them. This section delves into the most dominant themes used to target women and girls online, revealing how harmful narratives are constructed and perpetuated.

Across the platforms, CIR observed recurring themes, including gendered abuse that reflects societal norms and expectations. Women's roles and appearances were frequently questioned, often couched in sexualised or dismissive language. Misogyny was prominent, with successful women, including political figures and sports persons discredited, sexualised and stigmatised. Similarly, notions of 'independence' seem at odds with traditional gender roles. Workshop participants felt that cultural and religious values often underpinned the abuse and confined women into 'women's roles' and stereotypes.

SEXUALISING AND SHAMING WOMEN IN POWER

Women were repeatedly sexualised and portrayed as weak or unqualified to do jobs considered 'male', like politics or other leadership roles. This was evident in comments critiquing women in positions of power. For example, CIR identified repeated suggestions that Adanech Abebe (a prominent political figure and Mayor of Addis Ababa) achieved her role due to personal relationships rather than merit, exemplifying the intersection of sexism and misogyny.

> Why is Adanech given so many different positions, within short time? Is she Abiy mistress?
>
> HATE SPEECH DATASET

This aligns with broader trends identified in our earlier studies, where women's achievements are routinely undermined through gendered discreditation. When this was raised during workshops, participants noted that Adanach is considered a "polarising individual" and has been politicised since her role as the Attorney General. In her role as Mayor of Addis Ababa, she is often seen at Prime Minister (PM) Abiy Ahmed's side at events and opening ceremonies, which people use as evidence that she's close to the PM.

Workshop participants suggested that social media users often shame and sexualise women in powerful roles, as it is an easy form of abuse – it does not require any clever narratives or real evidence. Participants reported that other women in leadership roles are targeted with similar abuse like the Ethio Telecom CEO, Frehiwot Tamiru, who also received sexualised abuse and criticism for her position. Similarly, CIR's data revealed targeting of the YouTube host Betty, from 'The Betty Show'. There was a shared belief among workshop participants that this type of abuse is currently 'part of the package' of being a woman in a powerful role, with a digital presence, and reiterated the desire for change.

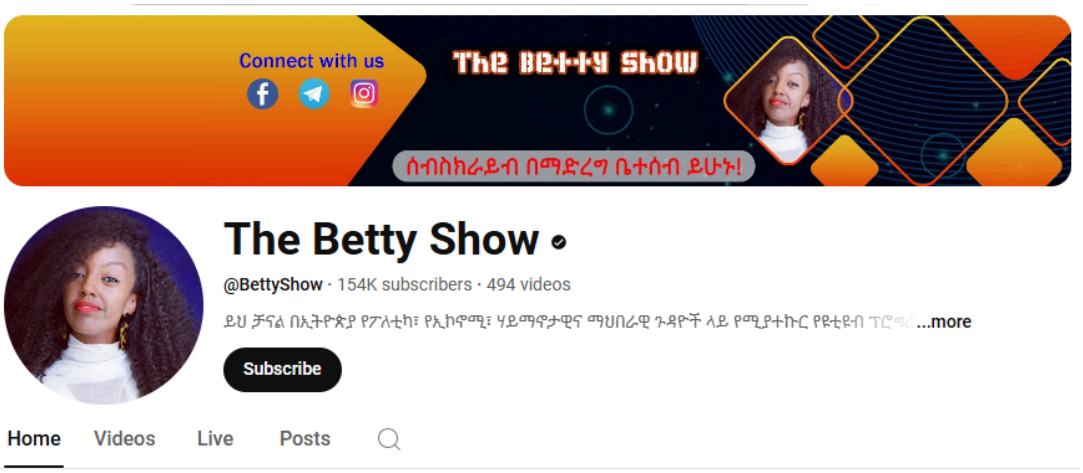


Figure 10: A screenshot of 'The Betty Show' YouTube channel, taken by CIR.

INDEPENDENT WOMEN AND SINGLE MOTHERS

CIR's research has found that women in leadership roles or who are active in public life – like politicians, political figures, journalists, civil society leaders, social media influencers, and TV presenters – are frequently targeted with online abuse. Workshop participants highlighted one reason for this trend – their independence. This aligns with CIR's earlier findings, reinforcing the pattern of hostility faced by women who challenge traditional gender expectations in public life. Workshop participants also suggested that unmarried women are often also targeted for the same reasons.

Workshop discussions on independent women sparked debate about the phrase "raised by a single mother," which came up repeatedly within the dataset. While it implies independence, it is also couched in disdain. Workshop participants unanimously said that this is a common slur in Ethiopia, used both on and offline. It is also used to shame and mock.

> They should disown you,
> and on top of that you are a single
> mother🤣🤣🤣
>
> HATE SPEECH DATASET

This insult is used in Ethiopia to denigrate individuals by implying moral or social inferiority. It highlights the cultural stigmatisation of single motherhood and its weaponisation in online hate. It also suggests the societal importance of having

male role models. Participants suggested that young single mothers are the most stigmatised, with single motherhood increasingly pitted as a failure of family duty, independence, and feminism – narratives that clash with societal and religious norms.

Interestingly, participants noted that if a child is raised by a single mother, there is the belief that the child wasn't raised properly, whereas if a child is raised by a single father, they are often praised as the "father of the year" for their efforts. This sparked an interesting debate over the role of men within the household. One participant criticised the insult by saying that being a single mother is often actually the result of a "failure of a man" and not the women, suggesting that the insult itself is flawed.

FEMINISM AS A TARGET OF ONLINE ABUSE

Similarly to CIR's earlier report, feminism emerged as a frequent and contentious topic in online conversations, often used as a tool and seen as a threat to discredit women and those seeking to protect or enhance women's rights.

When asked why feminism is so contentious in Ethiopia, workshop participants explained that patriarchal norms – reinforced by societal and religious values – underpin much of Ethiopia's social fabric, and this was reflected in the data. For example, comments often framed feminism as a destructive force, with some commenters asserting that it undermines traditional roles in Ethiopian society.

> Feminism is considered **radical** because it challenges deep rooted cultural and religious norms
>
> WORKSHOP PARTICIPANT, 2025

Debates around feminism also revealed competing ideologies, signalling that there is not a uniform view of what feminism is. For example, one commenter mocked another user by stating:

> Feminism ain't this, babe. Women should get an absolute choice to be either working mums or stay-at-home mums. You're imposing your idea of feminism

HATE SPEECH DATASET

Feminists were consistently targeted, shamed and discredited online. Debates frequently portrayed them as dangerous or misguided. For example, social media users depicted feminism as a tool for financial gain. Such remarks imply that feminists are motivated by self-enrichment rather than genuine advocacy for women's rights. This narrative diminishes the legitimacy of the feminist movement and perpetuates the stereotype that women cannot act altruistically or independently of personal gain.

> Ethiopian feminism is only about money laundering! Please stop using innocent women to fill your own pockets!

HATE SPEECH DATASET

During workshops for this report, participants reported that feminists are often accused of being paid by NGOs and foreign countries, labelling them as 'sellouts'. This aligns with a finding from the previous study that feminism is often considered a Western concept, at odds with, or even degrading, Ethiopian values. Thus, abusers may criticise, abuse, or ridicule women and girls who appear supportive of women's rights or feminism online for being influenced by Western values.

A concerning finding was the conflation of feminism with lesbianism, which is illegal in Ethiopia. Comments frequently linked feminist ideals to socially stigmatised identities. This framing seeks to delegitimise feminism by associating it with behaviours considered taboo in Ethiopian society, thereby reinforcing a culture of exclusion and hostility toward women advocating for gender equality. Such accusations may not only be used as insults but as a means to threaten the individuals' safety, restrict their freedom, and silence their voices in public forums

because successful prosecution of homosexuality claims carries a prison sentence. This association highlights the extent to which online abuse can spill into offline consequences, with far-reaching impacts on personal and professional lives.

When CIR put the narratives of 'financial incentive' and the conflation of 'feminism with lesbianism' to the workshops, many participants noting that these narratives were not only "unsurprising" but "common" and that they represent easy methods to delegitimise the desires of the feminist movement.

HATE SPEECH IN WOMEN'S SPORTS

A specific but significant form of hate speech included the targeting of women's sports. Common narratives included claims that women's sports were inferior and that they relied on men's financial support for survival. This rhetoric not only diminishes women's contributions but also reinforces traditional gender hierarchies, discouraging broader acceptance of women's athletic achievements.

Workshop participants were not surprised by this finding and said that women's sports are commonly ridiculed. Given that the Olympics took place during the timeframe of analysis, participants expected women's sport to be mentioned on social media. This sparked a discussion about the trailblazing sportswomen that compete for Ethiopia and the abuse they face (both on and offline) due to their gender.

Participants also recalled that, during the data collection timeframe, two women were elected into major positions within the Addis Ababa Football Federation. Mastawel Wendwesen and Haregeweyn Assef were appointed Vice President and Financial Officer, respectively. They were mocked by the public, and it was said that they were not qualified to talk about, or work on, the issue of Football (sources redacted due to privacy concerns).

DESENSITISATION AND NORMALISATION OF HATE SPEECH

Workshop participants were unanimously surprised by the finding that overt threats or aggression were less common than other forms of hate speech, especially on TikTok, where they saw far more 'aggressive', 'violent' and 'threatening' hate speech content. This finding aligned with CIR's earlier quantitative study. This could be because platforms may have already removed some of the most egregious examples or because the most shocking content is more notable and thus, easier to remember, like overt threats.

> TikTok is the most toxic, violent and scary platform for women - Influencers say outrageous things to take advantage of the virality it gives

WORKSHOP PARTICIPANT, 2025

Workshop participants believe that overt threats were far more common than the data suggested, particularly on TikTok, which one participant described as "an abuse-driven site." The disparity could be due to the removal of extreme content by platforms before data collection. Alternatively, as CIR's previous research highlighted, other forms of gendered hate speech have been normalised, resulting in them going under the radar. CIR's earlier work found that stereotypes and messages reinforcing women's inferiority often go unnoticed, while insults, mockery, and gendered abuse are so commonplace that they are frequently overlooked as forms of hate speech. It could be that there are lots of threats and violent content, but the quantity isn't as high as these other less visible forms of hate. For example, CIR's previous study concluded that the widespread normalisation of gendered hate speech on social media has led to societal desensitisation to the issue.

Moreover, workshop participants echoed this concern, observing that gendered abuse has become so endemic that it is no longer recognised as harmful, making it effectively 'invisible'. This normalisation, combined with a lack of understanding about what constitutes hate speech, means that less overtly aggressive or threatening forms—such as harmful stereotyping and mockery—are often dismissed, underreported, or misunderstood as less severe.

Policymakers and educators must not ignore these subtler but equally damaging forms of hate speech, as failing to address them only reinforces a culture where gender-based abuse is tolerated and continues unchecked.

DIGITAL RIGHTS

Digital rights are also often dismissed as unimportant, with critiques suggesting that the online world is a separate and ungovernable space. Advocacy for digital rights is frequently undermined, and according to workshop participants, women who speak up for their rights online are often met with abuse and told they should simply be grateful for internet access. Furthermore, participants highlighted a common view among social media users: having access to the internet equals empowerment, as those with internet access often represent the most affluent members of society or urban dwellers. However, workshop participants reject this claim, stating that access doesn't equal empowerment.

> TFGBV limits women's very limited access to the digital space

WORKSHOP PARTICIPANT, 2025

Workshop participants reported that the lack of governance in digital spaces and ineffectiveness of content moderation practises on social media platforms limits women's participation in online debate. Participants shared a belief that digital platforms should not operate with impunity and harms perpetrated online should have consequences.

A key barrier to talking about and researching TFGBV is the artificial distinction between the 'online world' and the so-called, 'real world'. It is clear that offline discourse impacts online activity, and online harms do not stay online, their impacts are both significant and far-reaching in peoples' everyday lives. As one participant noted, harms in online spaces "imitates violence in the physical space," reinforcing real-world inequalities and harms. Online and offline spaces are intrinsically linked.

> As we live in a digital era, it is essential that women have spaces where they can engage safely, share ideas, and have their voices heard

WORKSHOP PARTICIPANT, 2025

4.3 WOMEN AND GIRLS COMPARED TO MEN AND BOYS

This research, while centred on TFGBV against women and girls, also uncovered and analysed instances of hate speech targeting men and boys. This provides an opportunity to explore and compare the nature and sentiment of hate speech across different genders.

Although the forms of hate speech showed some variation between genders, one constant stood out: **insults** dominated as the most common type of hate speech overall. However, when examining sentiment, a far richer and more varied picture emerged, revealing distinct dynamics in how hate manifests across gender lines.

The study highlights how gendered abuse in Ethiopia reflects and reinforces societal norms, restricting both men and women within rigid gender roles. Women are objectified, excluded, and ridiculed, while men are attacked for perceived weakness, lack of masculinity, or for supporting gender equality. Political figures of both genders face online hate, but women in politics are more likely to be discredited based on their appearance or societal roles, whereas men are attacked for their policies or affiliations. The intersection of gender, politics, and ethnicity further complicates online abuse, creating a digital space where misogyny, rigid masculinity, and political divisions fuel hostility.

4.3.1 COMPARING HATE TYPE

The pie charts for men and boys in comparison to women and girls reveal slight differences in the **'types'** of hate speech targeting each gender subgroup (see figure 10). For both men and boys, as well as women and girls, **insults** emerged as the most common form of hate speech, but the proportions varied. Men and boys experienced the highest share of **insults**, which made up over half (56%) of the total hate speech directed at them. They also faced slightly more threats than women (10.2% compared to 9.6%).

In contrast, women and girls received a higher proportion of hate speech that conveyed their cognitive, physical or social inferiority (25.8%) compared to men (18.9%). Women were also more likely to encounter hate speech that involved **'presumed association'** (19.8%), where their gender was linked to certain negative traits, such as laziness or greed. Interestingly, the differences in **'presumed association'** were less pronounced than the disparities seen in the other categories.

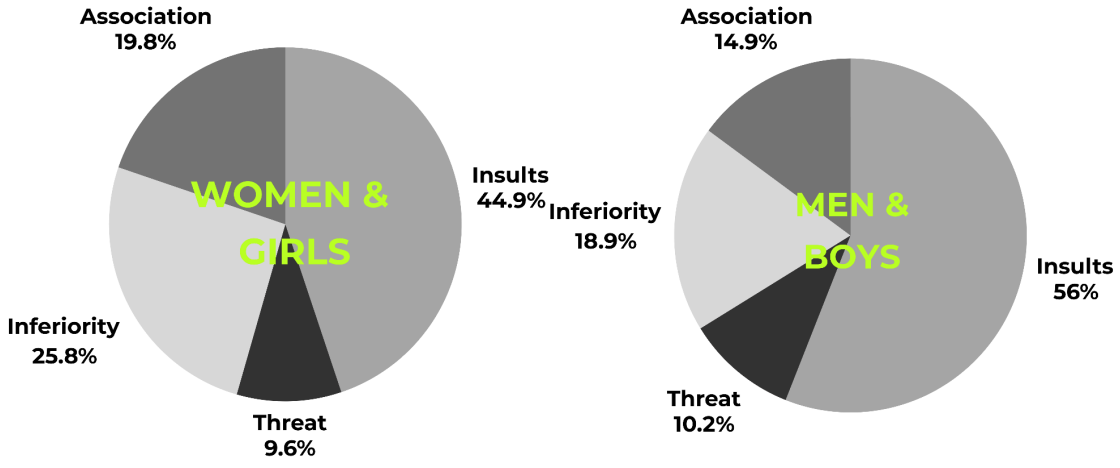


Figure 11: Pie chart showing the types of hate speech for each gender subgroup in CIR's dataset.

4.3.2 COMPARING HATE SENTIMENT

CIR's analysis of the **sentiment** of hate speech (classified into: **aggressive, offensive, mockery,** and **stereotypical**) revealed more notable differences than hate type above. For example, men and boys encounter a significantly higher proportion of **offensive** hate speech (57.2%) than women and girls (37.5%).

Women and girls face a greater proportion of **stereotypical** hate speech, with 32.6% of the abuse directed at them falling into this category, compared to 16.9% for men and boys. Although, **stereotypical** hate is still the second most prevalent sentiment for men and boys.

Furthermore, 20% of the hate directed at women and girls involves **mockery**, compared to just 10.3% for men and boys. Additionally, while men and boys face slightly more **aggressive** hate speech (15.5%) than women and girls (11.7%), the difference is not as pronounced.

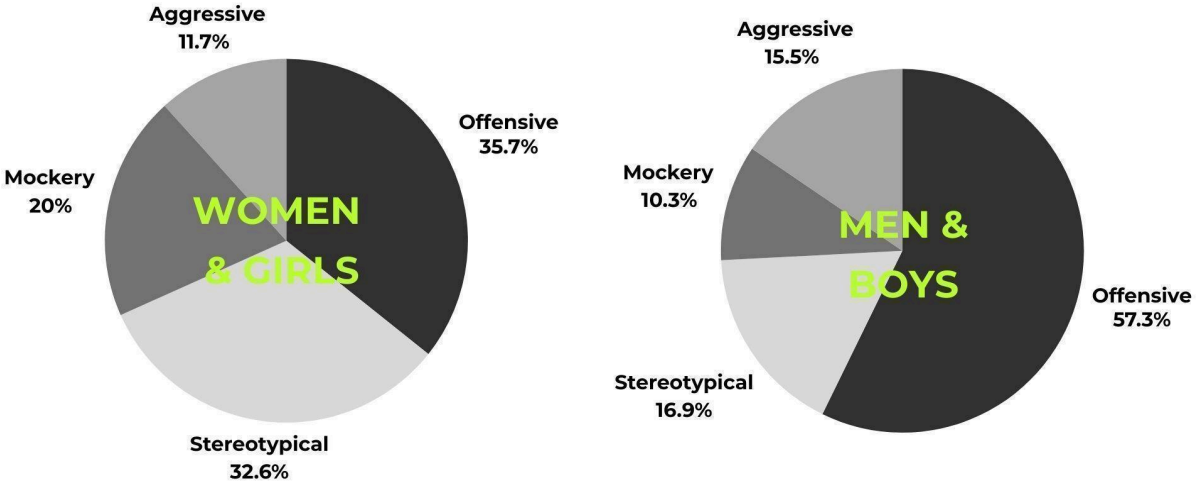


Figure 12: Pie chart showing the sentiments of hate for each gender subgroup in CIR'S dataset.

4.3.3 NARRATIVES

Exploring the gendered dynamics of online abuse in Ethiopia reveals distinct patterns in how men and women are targeted. While men's masculinity is frequently challenged, women experience misogyny, objectification, and exclusion from public spaces. Some men also face abuse for supporting gender equality, while political and ethnic divisions intersect with gendered hate speech, shaping online hostility in different ways.

Similarly to the women and girls only data (section 4.2), all participants were shocked that threats and aggressive speech formed only a small portion of the hate speech for both genders. There was a shared belief by participants that the

results are ‘toned down’. They felt threats, and aggressive speech was by far the most common, and it is what they felt they saw most of on the platforms.

MASCULINITY & THE OBJECTIFICATION OF WOMEN

Men are frequently attacked for not conforming to traditional expectations of strength and dominance. Common insults include, "you should have been a woman" and "grow some balls", which are used to attack their masculinity, highlighting a broader societal expectation of strength and power.

> This guy should be wearing a dress,
> so weak
>
> HATE SPEECH DATASET

Women and femininity are frequently used as tools for both gendered and ethnic insults. For example, the term "diqala" (used in hate speech targeting Tigrayans) explicitly insults men but translates to “sons of women”. The term carries the same contextual meaning as the phrase in English “son of a bitch”. Similarly to the phrase “raised by a single mother”, it implicitly weaponises the female gender as a means of shaming and emasculating men, here blending ethnic and gender-based violence in a single insult.

CIR’s data annotators observed that men frequently insult other men and women with language that objectifies women's bodies, reinforcing harmful gender dynamics. As a result, women not only experience gender-based violence firsthand but also become central to its perpetuation, used as both targets and symbols within abusive discourse. For example, workshop participants recorded a rise in the weaponisation of menstruation and women’s hygiene in hate directed against men. These insults suggest that being associated with femininity is inherently shameful, further entrenching rigid gender roles.

> Abuse targeting men often says things like: *"where did you get your pad?", "do you have enough tampons?", or "you're acting like you're on your period!"*
>
> WORKSHOP PARTICIPANT, 2025

One workshop participant mentioned that the Ethiopian gaming community coined a phrase, "emumu". This term, which represents a woman's genitalia, has now morphed into an abusive word towards women. This trend is not distinct to Ethiopia, for example, there are many English terms for a woman's genitalia that are commonly used as insults, that often represent the harshest and most derogatory swear words. In the context of Ethiopia, it can also be used to describe a man's lust or sexual desires. The term has morphed into a general insult used against both men and women. These types of slurs reinforce a culture of misogyny by associating femininity with insult and degradation.

Workshop participants highlighted that these gendered insults contribute to a feeling of male dominance in online spaces, leading many women to withdraw from digital discussions. The pressure to stay silent online is often greater for women than men, limiting their engagement in public discourse.

MISOGYNISTIC LANGUAGE

CIR observed the mainstreaming of misogynistic language and terminology, including the word 'simp' which began in the fringes of internet use. While the word has evolved in meaning over time, it is used derogatively to insult men perceived as too attentive to women or unmanly. Workshop participants were not surprised by these findings, linking this to the growing influence of figures like Andrew Tate, who have gained notoriety in Ethiopia and have become increasingly popular among young Ethiopian men and boys. Tate's messages reinforce patriarchal norms and condone violence towards women. While social media has created spaces for empowerment, activism, and connection, this finding underscores how harmful views can spread easily, transcending borders.

> If u respect her or treat her nice u are a simp and any woman hates simp she wants a guy that treat her like shit because bzo setoche asdedagachew nw *[for many women, this is how they were raised]*
>
> HATE SPEECH DATASET

(Translation by CIR).

Additionally, Ethiopian influencers like US–based Mota Keraniyo and content creators on podcasts and YouTube channels have been noted for spreading misogynistic content and offensive gendered slurs, often under the guise of comedy (podcast and channel names available on request). These narratives normalise the degradation of women, making misogynistic rhetoric more acceptable in everyday discourse. One participant recalled a video that explicitly described a woman by saying: "While others have dugout fuels, you have your emumu dugout. You have no mind, you think by your emumu" (source redacted to avoid amplifying the influencer).

POLITICAL FIGURES

Political figures of both genders experience online hate and abuse, but women in politics are often criticised for their appearance or societal role, while men face abuse based on their views, achievements, or affiliations. Workshop participants observed that gender-based attacks were far more prevalent against women, a trend consistent with the findings of an earlier CIR study. conflict

For example, in contrast to Adanach Abebe (see section 4.2.5), PM Abiy Ahmed also received a notable volume of anti-Abiy sentiment online. However, most of this fell outside of the Ethiopian government's official definition of hate speech as it targeted his policies and political views. In the few occasions where this strayed into hate speech, such as where his Oromo identity was cited, it was included in the study. The comparison of these two notable figures alone is indicative of the different types of abuse the genders face online, reinforcing the experiences of the social media users CIR spoke to.

FEMINIST MEN

Like women who advocate for gender equality, men who identify as feminists also face violent backlash. Workshop participants noted that being a male feminist in

Ethiopia is as dangerous as being a female feminist, with some reporting seeing death threats directed at men who support women's rights.

4.4 INTERSECTIONAL HATE

CIR's investigation into intersectional hate speech revealed a complex and shifting landscape of online abuse, highlighting how the risks change when other identities are targeted alongside gender. To classify as 'intersectional hate speech', the comment targeted more than one protected characteristic: **disability, ethnicity, gender, race, or religion.**

The findings are telling – hate speech doesn't simply target women and girls in a vacuum; it shifts and adapts depending on the additional identity markers present. When women and girls are targeted for both their gender and another form of identity (for example, ethnicity or religion) simultaneously, the nature of the hate speech shifts and is multi-layered.

By analysing the narratives, CIR reveals how gender, ethnicity, religion, and skin tone intersect to shape the nature of online abuse in Ethiopia. Political events and internal conflicts fuel ethnic and gendered hate speech, while religious influences underpin much of the rhetoric used to discredit and demonise women. For example, Muslim women face unique forms of abuse, particularly in relation to the hijab. Deeply ingrained religious narratives – such as the association of women with Eve's 'original sin' – are frequently used to justify misogyny. Additionally, colourism and anti-Blackness were identified as underexplored but pervasive issues. The findings reveal the complex and multi-layered nature of hate speech, showing how women's identities are weaponised against them in distinct and deeply entrenched ways.

In the analysis below, the 'Women and Other' category encompasses identities beyond those listed as distinct groups by CIR. For example, it was often used by researchers to note when women were being targeted alongside religion, but the exact religious target was unclear.

Compared to hate speech targeting women alone, the analysis uncovered some interesting differences (see figure 13 below):

- **Less Insulting and Offensive Speech:** When multiple identities are targeted, CIR found a surprising decrease in **'insulting'** and **'offensive'** hate speech. This suggests that the abuse moves away from degrading and direct insults, focusing more on other forms of hate speech when multiple facets of identity are combined.
- **More Inferiority:** Women in specific intersectional categories, such as "Women and Oromo," "Women and Muslim," and "Women and Other," were

more likely to be subjected to hate speech that emphasised their perceived inferiority than women targeted for just their gender. This points to a more insidious form of discrimination that targets both gender and other aspects of identity.

- **Increased Presumed Association:** The combination of gender and ethnicity, particularly for "Women and Tigrayan" or "Women and Other," led to more hate speech based on presumed associations. This included stereotypes and false beliefs about the target group, such as assumptions of greed, demonisation, or betrayal.
- **More Aggressive Speech:** Intersectional hate targeting groups like "Women and Oromo", “Women and Amhara”, and "Women and Tigrayan" saw slightly more aggressive speech, which included threats and intimidation, suggesting that these identities are often associated with higher levels of hostility. This is perhaps unsurprising given the ongoing and historical conflicts in these regions.
- **More Stereotypical Abuse:** Across all intersectional categories, women faced more stereotypical hate speech than when gender was the sole target. This indicates that the abuse is often more generalised and rooted in harmful cultural or religious beliefs about the roles and characteristics of women from specific backgrounds.
- **Mockery and Satanic References:** Of particular concern was the increased use of mockery in the "Women and Other" category. Many women in this group were mocked with harmful language, including likening them to Satan. This finding was more noticeable in this study compared to the previous study.

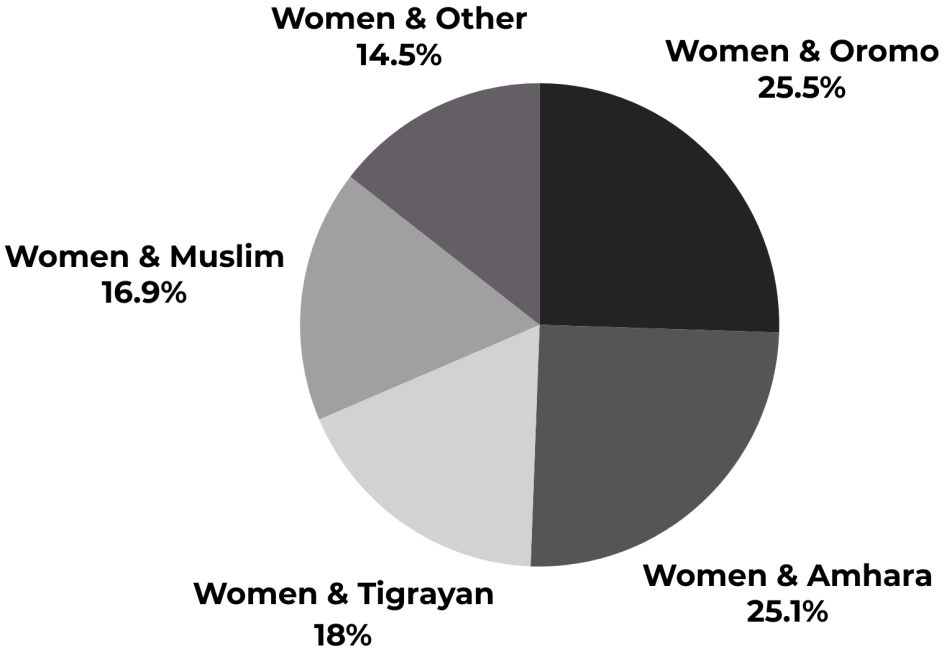


Figure 13: Pie chart showing the targets of intersectional hate speech within CIR’s dataset.

4.4.1 NARRATIVES

The nature of hate speech changes, often becoming more complex and nuanced, based on factors like ethnicity or religion. Workshop participants recalled that Ethiopian society is deeply religious and conservative, a characteristic played out both online and offline and amplified via social media platforms. This, they felt, explained many of the trends in the intersectional gendered abuse seen in this study. Additionally, they reported that the challenges that come with intersectional identities are part of everyday life.

> “No one is just a women. They also belong to ethnic group or religious groups. Intersectional abuse is what we all experience in our daily lives!”
>
> WORKSHOP PARTICIPANT, 2025

INTERNAL CONFLICTS & POLITICS

This study signals that current events offline impact online debate and hate speech. Political events have a considerable impact on the topics of discussion online and the rhetoric used. This can be seen through the narratives within the intersectional abuse targeting women and girls of Oromo, Amhara, and Tigrayan ethnicities in the context of ongoing conflict in these areas of Ethiopia and internal politics.

The more aggressive nature of intersectional hate targeting Amhara, Oromo and Tigrayan women came at no surprise to the majority of workshop participants. Political hate often strays into ethnic slurs and highly charged language towards certain groups. Workshop participants said that hate speech that is reactive to political events and uses inflammatory rhetoric is quite noticeable online.

Amhara women received more mockery than Oromo women in CIR’s earlier findings, but this time, Oromo women faced more mockery and inferiority. When asked why this might be, participants suggested that this is the result of the changing political climate and different events dominating the national agenda.

HIJAB BANS IN SCHOOLS

Abuse levelled at Muslim women was prominent within the dataset. Workshop discussions revealed that there had been rising levels of anti-Islam rhetoric, particularly in the city of Axum, Tigray, where the hijab was reportedly banned in four schools. This led to fierce debates online and in the community over whether schools should be religious spaces or secular. From November 2024, tensions were high, leading to protests, demonstrations and calls for the hijab to be allowed. According to workshop participants, this led to counter-demonstrations and online campaigns targeting Muslim women.

Muslim women have also faced abuse related to their appearance. They have been attacked online for not being "Muslim enough", for example, if they do not wear their hijab or if they wear jeans.

SATANIC REFERENCES

CIR noticed mocking speech and comments thar presumed the association of women with evil tendencies. The language was deeply religious and often directly referenced religious texts, such as passages related to Eve. Her biblical association with sin was repeatedly used to demonise women.

> Jesus no get wife, God no get wife, Satan says they run from women. There must be something they are not telling us about this gender😂😂😂
>
> HATE SPEECH DATASET

When this finding was shared with workshop participants, there was often a 'knowing nod'. They were not surprised that religious influences are visible within gendered abuse. Participants said that comparing women to Satan and Eve's 'original sin' was a common way to demonise women offline, and that it goes back generations; it is now being played out and amplified online.

COLOURISM AND ANTI-BLACKNESS

One participant suggested that there was a key and pervasive narrative missing from the study: colourism or anti-blackness; a form of discrimination based on skin tone. It is rooted in historical, social and cultural norms and results in microaggressions and bias in everyday life.

Although Ethiopia avoided colonisation, commentators have suggested that Ethiopia is still impacted by Western notions of beauty and that the media

reinforces these ideals and beauty standards. The workshop participant explained that light skin is sometimes considered more beautiful or advantageous and is associated with wealth. They continued to say that darker-skinned individuals often receive increased abuse and stigma.

A YouTube influencer, Weyni Tesfai, addressed her views of the root cause of colourism in Ethiopia, suggesting that the legacy of slavery contributed to 'anti-black' sentiments. Interestingly, she reported that many Ethiopians do not identify themselves as 'black'. However, the topic of slavery is not only often unaddressed but also deeply contentious, leading to the YouTuber receiving hate speech for her views on the topic on Reddit.

The workshop participants added that it is not always clear cut. For example, if individuals are considered 'too white', it can be a reason to target women, suggesting they do not belong in Ethiopia or they are foreign. They also added that depending on where you are from in the country, rather than being negative, dark skin can embody being Ethiopian and firmly from the African continent.

During workshops and roundtables in 2023, CIR debated keywords that are used to discriminate against individuals in Ethiopian society. This formed the lexicon developed for this study and it included terms that are used to discriminate due to skin colour, for example: bariya (ባሪያ), which means slave and is normally used to refer to darker skinned Ethiopians; and Shankila (ሻንቅላ), a derogatory term, often associated with 'dark skinned'.

The findings highlight the unique forms of abuse faced by women at these various intersections, where both **gender and identity** are used as weapons against them and reveal the multifaceted ways women and girls experience discrimination online.

4.5 COMPARISON WITH OTHER FORMS OF IDENTITY-BASED HATE SPEECH

While the primary focus of this research was on TFGBV, CIR also identified hate speech targeting various other protected identity groups, in line with the Ethiopian Government's definition of hate speech. This section analyses **gendered** hate speech in comparison with **ethnic, religious, racial, and disability**-based hate, revealing the intricate landscape of online hate speech in Ethiopia (see figure 14 below). The results show clear differences in the types and sentiments of hate speech faced by different groups.

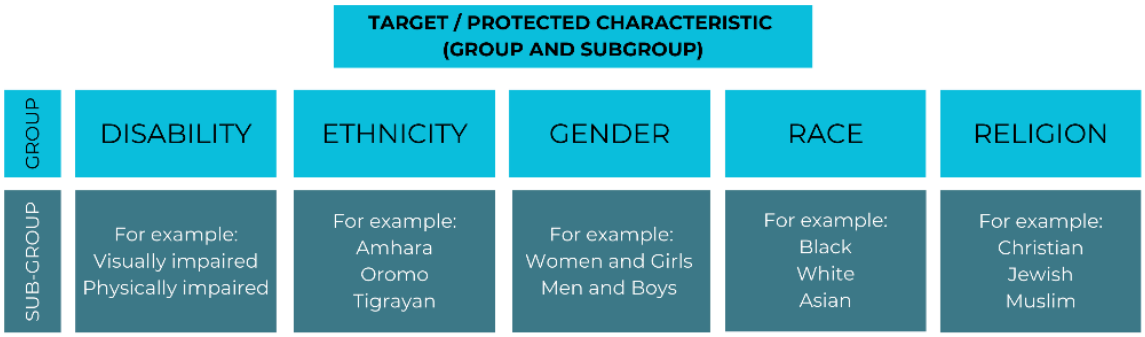


Figure 14: An excerpt from the contextual framework and annotation protocol, defining the different targets, as per the Ethiopian Government's 'protected characteristics' in the Hate Speech Proclamation.

Narrative analysis reveals the vastly different types of hate speech levelled at ethnic groups compared to women and girls. Women and girls are less likely to face **threats** or **aggressive** speech than ethnic groups, yet they are disproportionately subjected to hate involving **stereotypes** and **mockery**. For example, the intersection of ethnicity and politics fuels inflammatory discourse, with figures from different backgrounds accused of betraying their country or aligning with foreign powers. Historical references and terms like "banda" and "traitor" are weaponised to incite real-world violence, as seen during the height of conflict between the TPLF and Government forces (commenced in November 2020). Additionally, religious rhetoric is frequently used to justify hate, with terms like "devil" and "infidel" reinforcing ethnic and political divisions. These findings highlight the deeply interconnected nature of hate speech, reflecting and amplifying Ethiopia's ongoing political, ethnic, and religious tensions.

4.5.1 COMPARING HATE TYPE WITH OTHER IDENTITY GROUPS

To understand how hate speech manifests across different identity groups, CIR compared the **types** of hate speech experienced by the five most prevalent identity groups in our dataset: women and girls, men and boys, Tigrayan, Amhara, and Oromo (see figure 15).

Insulting language was the dominant form of hate across all groups. Women and girls narrowly received more (44.9%) than Tigrayan (44.1%), Amhara (33.7%), and Oromo (31.7%). However, men and boys received an even greater proportion of insulting hate speech, with 56% of the hate directed at them falling into this category.

The ethnic groups – Tigrayans, Amharas, and Oromos – faced slightly more **threats** than the gendered groups. Tigrayans were particularly targeted with threats, with 20.4% of the hate speech falling into this category. In contrast, women and girls experienced fewer threats (9.6%) than the other four groups.

The gender subgroups – women and girls, men and boys – received less hate speech involving the **presumed association** of their gender with certain negative traits compared to the ethnic groups. Women and girls were subjected to more messaging suggesting their cognitive, social, or physical **inferiority** (25.8%) compared to men (18.9%), Amharas (23.7%), and Tigrayans (15.6%). However, Oromos received the most of this type of hate speech (26.1%).

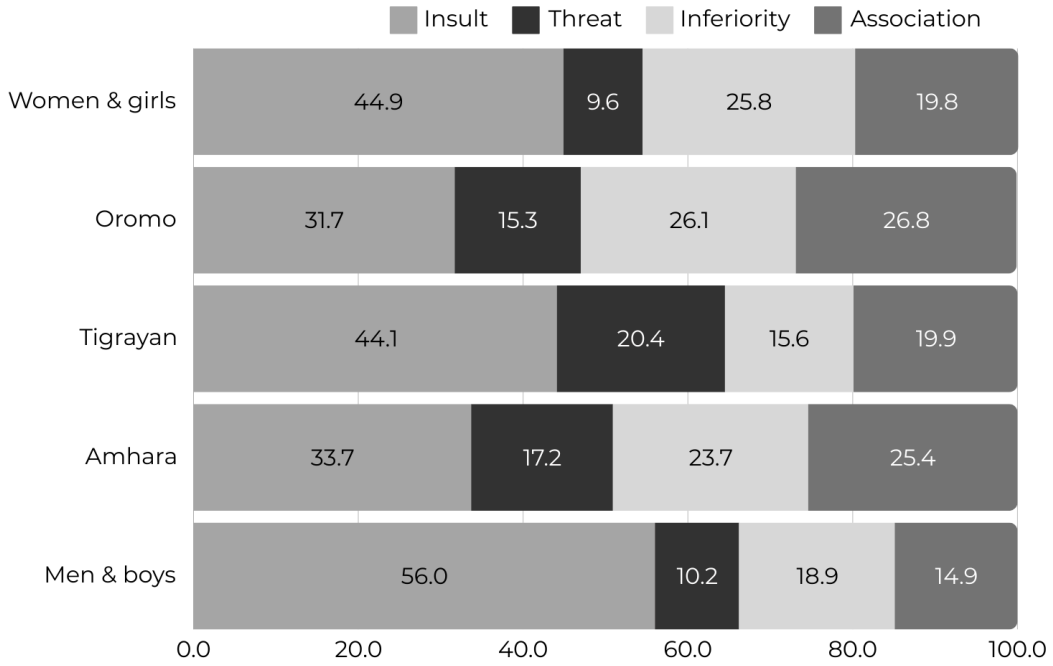


Figure 15: Bar chart showing the types of hate speech by identity group in CIR's dataset.

4.5.2 COMPARING SENTIMENT WITH OTHER IDENTITY GROUPS

While the **types** of hate speech varied across identity groups, the **sentiment** – the tone behind the speech – revealed more nuanced trends that went beyond the simple gender versus ethnicity divide (see figure 16).

Offensive hate speech dominated across all subgroups, but interestingly, women and girls received just slightly more offensive hate (35.7%) than Amharas (35.2%) and Oromos (35.3%). Men and boys, however, were the primary targets of offensive hate speech (57.2%), as were Tigrayans (44.7%).

Women and girls were subjected to a larger share of **stereotypical** hate speech (32.6%) than men and boys (16.9%) and Tigrayans (18.0%). However, Amharas (32.6%) and Oromos (34.8%) also faced substantial stereotypical abuse, highlighting how ethnic identity is often entangled with harmful stereotypes.

Women and girls endured more **mockery** (20%) than any other group, with the abuse frequently mocking their gender or appearance, underscoring the uniquely gendered nature of this type of abuse.

Tigrayans faced the highest proportion of **aggressive** hate speech (30%), significantly more than any of the other subgroups. This suggests that ethnic identity, particularly in politically charged contexts, can attract more violent forms of hate.

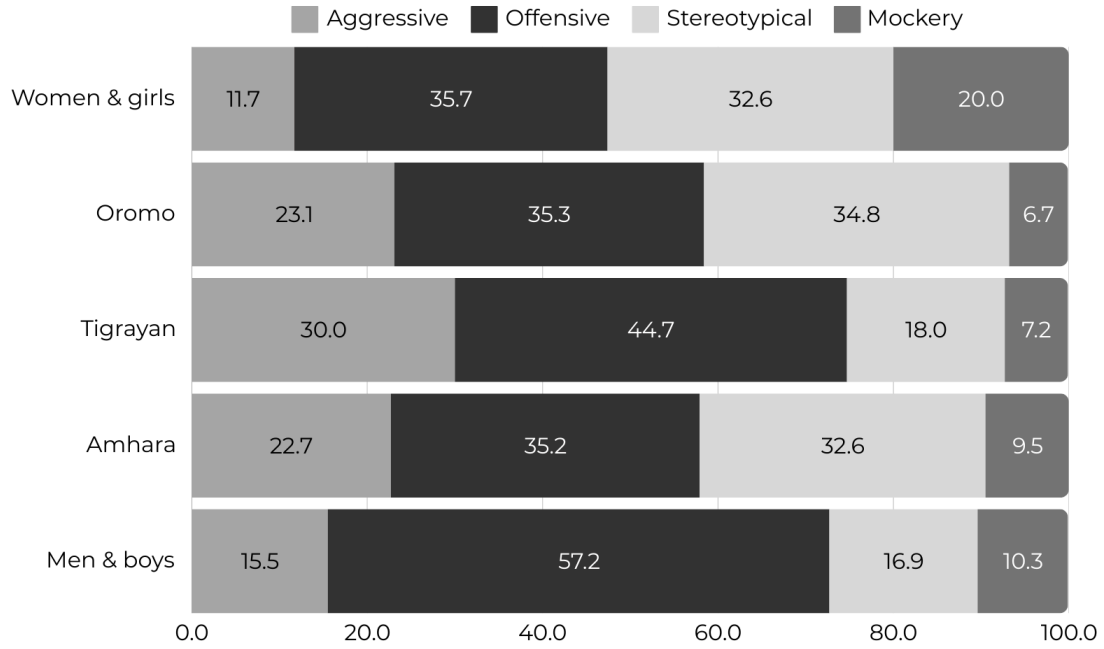


Figure 16: Bar chart showing the different sentiments of hate speech by identity groups in CIR's dataset.

4.5.3 NARRATIVES

Women and girls, as well as various ethnic groups, experience hate speech in distinctive ways that reflect both societal attitudes and the complexities of Ethiopia's diverse cultural and political landscape.

ETHNIC DIVISION AND HOSTILITY

Gendered hate speech, which, as the data suggests, overwhelmingly targets women and girls, manifests through insults, mockery, and stereotypical portrayals. The experience of hate among ethnic groups – such as Tigrayans – tends to involve more threats, aggression, and narratives related to political events. This aligns with the findings from CIR's last investigation, which concluded that current events influenced online debate, with events like the conflict in Amhara driving online dialogue and inflammatory language. It is clear, however, that online events also impact abuse targeting women and girls, as seen through abuse related to the hijab ban.

The role of ethnicity in fuelling online abuse is particularly evident in the use of language that pits one ethnic group against another and dehumanising language.

Terms like "donkeys" or references to groups as "traitors" or “banda” reflect the ethnic animosities that often dominate online discourse in Ethiopia.

Prominently, during the height of hostilities between the TPLF and Government forces, the term “banda” merged into a movement to identify Tigrayans, who would be labelled as “bandas” and their home addresses would be posted on social media. People would then hunt them down and, in some cases, kill them, as was the case of the Tigrayan University Lecturer who was killed by a mob in Amhara following calls for his death on Facebook. According to a workshop participant, there is a YouTuber who identifies “banda” and doxes them online. This individual, who claims to be an investigative journalist, releases videos of their homes and provokes viewers to cause them harm. CIR staff identified this individual and corroborated the workshop participant’s account. The YouTube links to these videos are still live (source redacted due to privacy and safety concerns).

The analysis also revealed that ethnic hate speech was far more prevalent on platforms like YouTube compared to TikTok. This could be due to the distinct nature of each platform: YouTube's longer-form content and larger political discussions seem to foster more hate rooted in ethnic identity, while TikTok's short-form, more entertainment-focused content may reduce the scope for such hate.

WHEN POLITICS & ETHNICITY MEET

Inflammatory and harmful speech tied to political beliefs was prominent within the dataset, particularly in the context of government supporters or critics. The Ethiopian Government’s definition of hate speech doesn’t include ‘political identity or views’ as an identity characteristic; however, when it strays into ethnic or religious abuse, CIR analysed it. Words like “donkeys”, “traitors”, “banda,” and “Minilik supporters” were used in this context.

Some political figures were vilified not just for their political stances but also for their ethnic backgrounds, and such accusations included that certain leaders were “selling out” or betraying their country for foreign powers. Abusive comments directed at political figures often crossed into generalised ethnic hate. For example, the quote below highlights how political hate can quickly spiral into ethnic accusations, portraying opponents as enemies of the state or traitors.

> Screw Getachawu Rada and Debretsion, they are Egyptian Trojan horses
>
> HATE SPEECH DATASET

Shimelis Abdisa, president of the Oromia region, also received a lot of political and ethnic hate, claiming he was against Amharas and suggesting ways to harm him or remove him from office. Such expressions of hate expose the complexities of how ethnicity and politics are intertwined in Ethiopia's ongoing political discourse.

History is often used within ethnic abuse, for example, by talking of certain battles or leaders. Social media users frequently labelled others as "Minilik supporters" in derogatory messaging. During a workshop, one participant noted that the term may be used to describe someone who is anti-federalism rather than an ethnic slur.

THE ROLE OF RELIGION

Religious undertones are normalised within hate speech in Ethiopia. This is a product of its rich religious history and current religious society. While no religious group made it into the top five most targeted groups, listed above, religion played a significant role in the targeting of certain ethnic and political groups, with religious labels being used to fuel divisive rhetoric. The term "devil", as already mentioned, was frequently employed in reference to ethnic groups, reinforcing a mix of ethnic and religious hate. Additionally, calling people "infidels", "non-believers", or "Satanists" was common within the annotated hate speech, indicating how religious language can be used within the text itself rather than merely as the target.

> Religious words and references are common in everyday conversations - this is normalised both on and offline
>
> WORKSHOP PARTICIPANT, 2025

4.6 GENDERED PATTERNS IN ONLINE HATE SPEECH ACROSS CONTENT GENRES

CIR's analysis of hate speech across different content genres reveals clear gender-based patterns, with female and male content creators facing distinct types of abuse. Women receive more hate in social and lifestyle topics, where insults and derogatory remarks dominate, while men face greater hostility in technology, religion, and travel-related content, often tied to ideological and political disagreements. Comedy, music & entertainment and politics and news showed a more balanced distribution. This reveals that the video content and the gender of the content creator can impact the hate speech experienced.

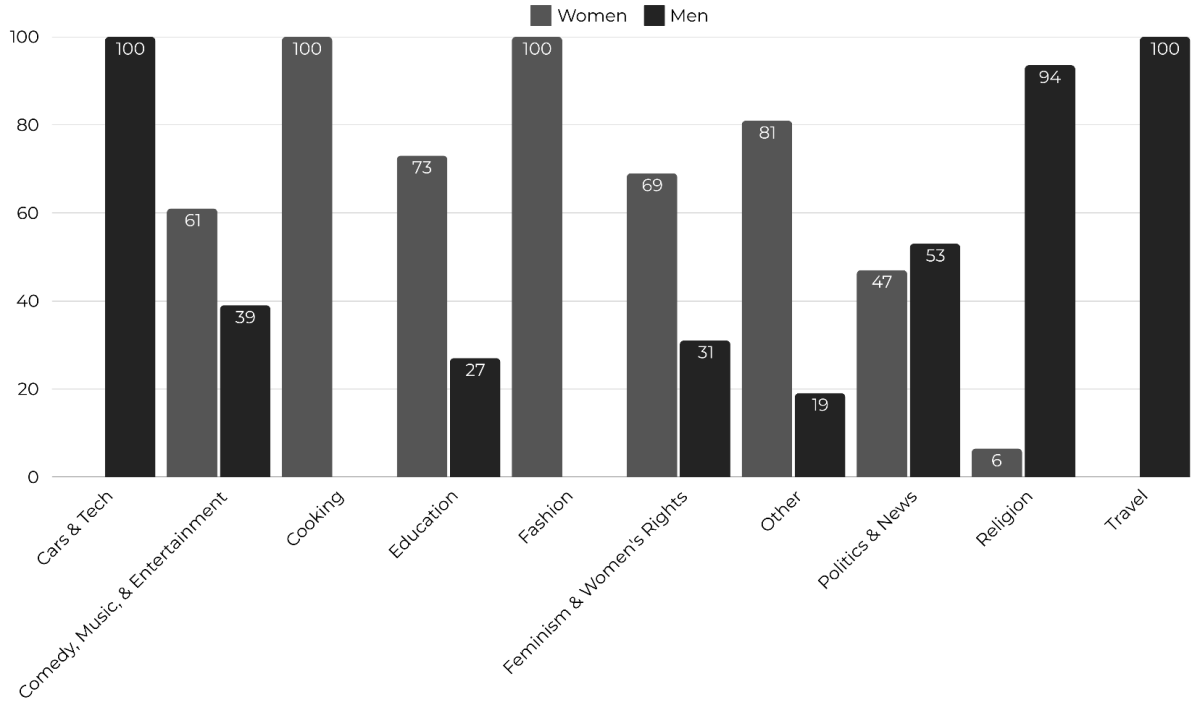


Figure 17: Bar chart showing the different proportions of hate speech received by male and female content creators by video content genre in CIR's dataset.

4.6.1 FEMALE AND MALE CONTENT CREATORS, ACROSS GENRES

To explore these differences further, CIR examined the **types** of hate speech and **sentiments** of hate directed at female and male content creators across the most prominent video content genres. The findings reinforce the study's earlier findings and spotlight a few interesting trends. Addressing online abuse requires a nuanced approach that considers both gender dynamics and the cultural landscape of each genre.

Across all genres, the findings reinforce earlier conclusions: Female content creators are systematically subjected to dismissal and ridicule, with the nature of

abuse varying depending on the topic they engage with (see figures 18-21 below). Hate speech targeting male content creators varies by genre, reflecting industry norms and societal biases.

Cars & Tech

- Female content creators - no data.
- Male content creators views on car and tech were belittled or dismissed.

Comedy, Music & Entertainment

- Female content creators were often not taken seriously, facing ridicule and belittlement rather than outright aggression.
- Male content creators faced both personal attacks and accusations of aligning with particular ideologies or groups. They also faced hate related to their identity and role, often beyond the substance of their content.

Cooking

- Female content creators in the cooking genre were subjected to personal attacks and were often portrayed as less capable or qualified. On the surface, this seems to go against the traditional view of women as the family cook. CIR researchers reviewed a few examples and saw that many of these personal attacks were related to the individuals' appearance. Further investigation into this trend is needed.
- Male content creators - no data.

Education

- The findings for female content creators in the ‘education’ genre underscore the ongoing struggle women face in academic spaces, where their authority and competence are often questioned and undermined, reflecting broader societal biases that position men as more authoritative in intellectual discourse. The high level of mockery exemplifies this.
- Similarly, male educators faced challenges on their expertise or credibility, with some encountering outright hostility, stereotyping, and ridicule. In comparison with female educators, they received less mockery, insults, and inferiority, yet more association with harmful characteristics.

Fashion

- Female content creators in the ‘fashion’ genre had their appearances criticised in online discourse, and hate reinforced traditional values. The combination of inferiority language and mockery suggests that female content creators’ views on fashion are not taken seriously.

- Male content creators - no data.

Feminism & Women's Rights

- Similarly to the narratives discussed earlier, female content creators in the 'feminism & women's rights' genre were accused of having ulterior motives, such as seeking personal gain or promoting Western ideologies. The findings also reinforce claims that advocating for gender equality results in belittling or dismissal.
- Male content creators in the 'feminism & women's rights' genre faced backlash for supporting gender equality. This aligns with CIR's workshop findings, where participants noted that male feminists experience hostility when advocating for gender rights. Men advocating for gender equality are often subjected to gender-based biases, potentially questioning their motives or authenticity.

Politics & News

- Female content creators in the 'politics and news' genre were discredited, with their opinions dismissed as less valid or informed compared to their male counterparts. The high proportions of aggressive and offensive speech also reflected the direct and often hostile nature of criticism faced by women in political discourse.
- This suggests that men engaging in political discourse were often personally discredited and threatened with aggressive, direct insults. This supports CIR's earlier findings that men are often vilified for their perceived affiliations or beliefs.

Religion

- Female content creators engaging in religious discussions faced heightened hostility, often linked to challenging traditional, patriarchal structures, and were dismissed as unqualified to participate in theological discourse.
- Male content creators engaging in religious discourse had their credibility challenged. Their views were dismissed or belittled, potentially reflecting ideological clashes or challenges to their authority in spiritual matters.

Travel

- Female content creators - no data.
- Male content creators in the 'travel' genre received hate based on assumptions about privilege or alignment with certain cultural or political ideologies.

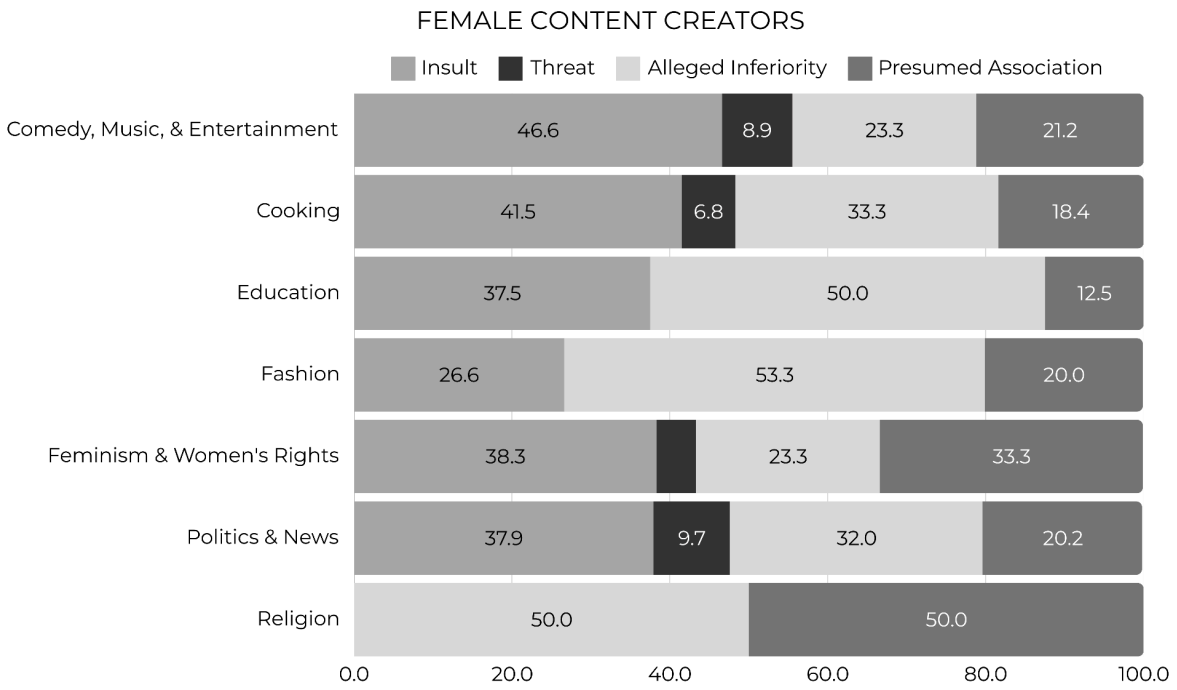


Figure 18: Bar chart comparing the different types of hate received by female content creators, by genre.

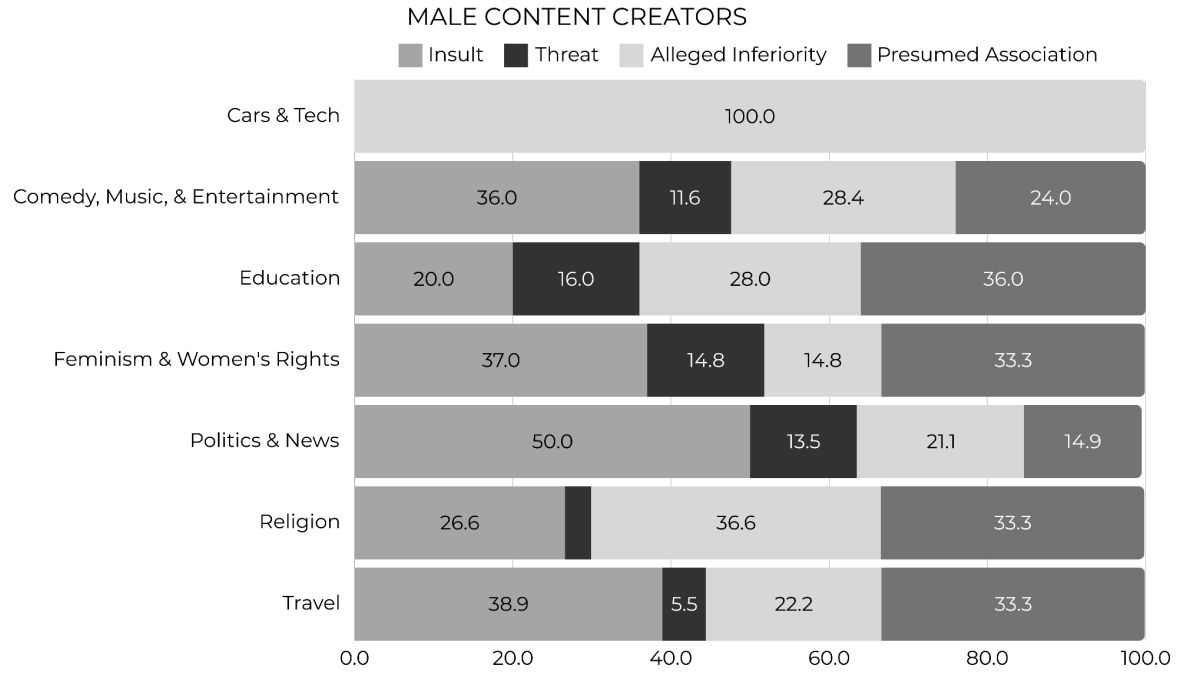


Figure 19: Bar chart comparing the different types of hate received by male content creators, by genre.

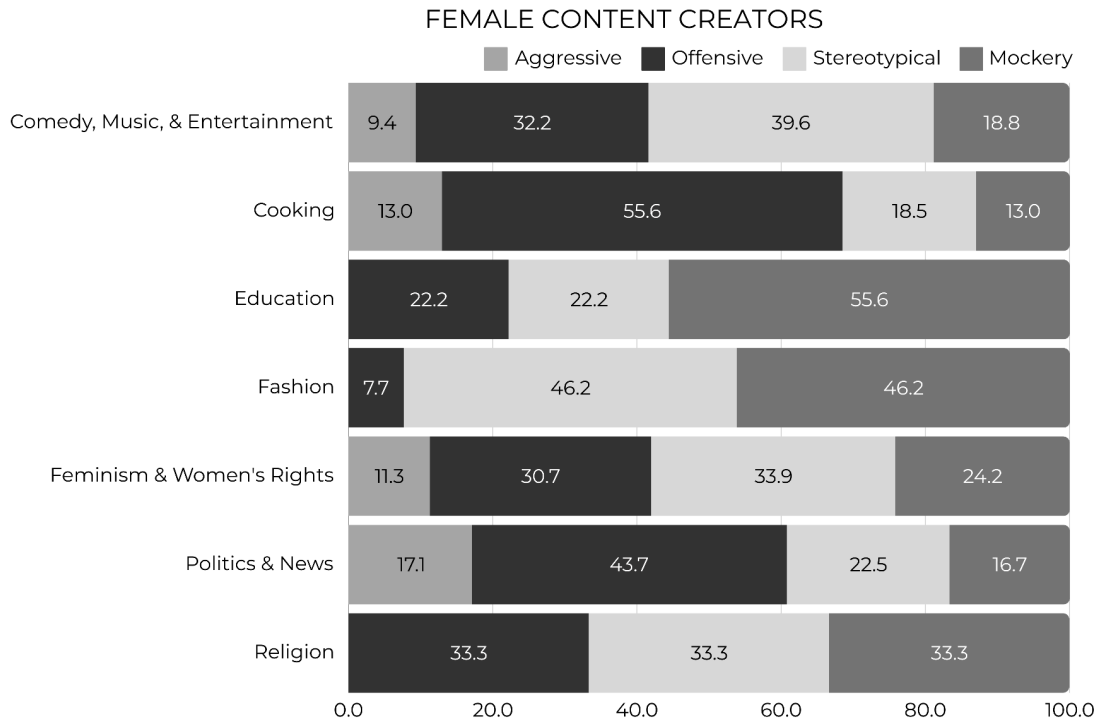


Figure 20: Bar chart comparing the different sentiments of hate received by female content creators, by genre.

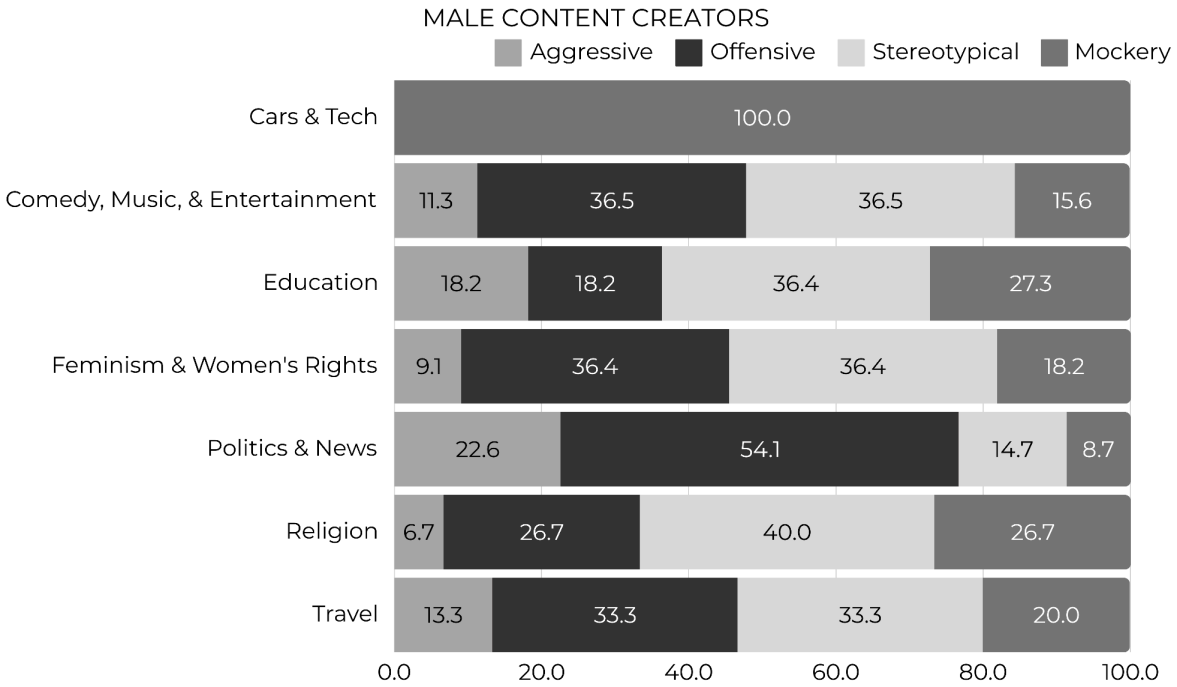


Figure 21: Bar chart comparing the different sentiments of hate received by male content creators, by genre.

4.7 CONTENT CREATORS THAT SEND AND/OR RECEIVE ABUSE

CIR took a deeper dive into the nature of the TikTok accounts and YouTube channels to determine whether the accounts that engaged in hate speech also received patterns of online abuse themselves. CIR categorised them as either senders, receivers, both senders and receivers, or neither sends nor receivers of hate speech (based on an analysis of a sample of their content). This hypothesis was incorrect, with very similar trends across the board, regardless of whether the online user was a sender, receiver, both, or neither.

5. CONCLUSION

TFGBV in Ethiopia is harmful, pervasive, and deeply intertwined with entrenched gender norms and misogynistic attitudes. Women and girls face a barrage of insults, stereotypes, and degrading rhetoric that diminish their voices and restrict their participation in public life.

In response to calls from Ethiopian stakeholders, CIR has researched TFGBV on TikTok and YouTube, uncovering new insights through data-driven analysis and expert workshops. By expanding the evidence base on TFGBV in Ethiopia, this report equips stakeholders with the knowledge needed to drive meaningful policy change and strengthen advocacy efforts.

> If ignored, TFGBV will erase women from public life
>
> WORKSHOP PARTICIPANT, 2025

Women and girls experience pervasive abuse across both TikTok and YouTube. **Insults**, degrading **stereotypes**, and messages suggesting their **inferiority** dominate, with **mockery** often used to undermine them. While **threats** and **aggressive** speech are less frequent, they still pose serious risks to women's safety both online and offline. Although **hate types** were consistent across YouTube and TikTok, the **sentiment** varied, reinforcing the need for both broad and platform-specific interventions. A comparison with CIR's earlier study also reveals slight differences across Facebook, Telegram, X, TikTok, and YouTube, which

workshop participants note could be due to usership and content moderation policies.

There were prominent narratives used to **discriminate** and **shame** women and girls. Criticism of women's societal roles and appearance was common. Women who challenge social norms, including women in power, women's sports, or those advocating for women's rights, faced sexualisation, discrediting, insults, and stigma. Feminism is often framed as a threat to traditional values, and feminists were discredited through accusations of financial fraud or a conflation with lesbianism. These narratives, along with the dismissal of digital rights, have contributed to a digital environment that silences women and limits their participation in public life.

Furthermore, hate manifests in **distinct** ways across **gendered lines**. For example, men and boys are also restricted within rigid gender roles, with hate speech targeting perceived weakness, lack of masculinity, or for supporting gender equality. While both male and female political figures faced online abuse, women were more often undermined based on their appearance or traditional societal roles, and men were criticised for their policies, actions, or political ties.

Abuse targeting men often **weaponised** or **objectified** the female gender as a means of shaming and emasculating them. As a result, women not only experience gender-based violence firsthand, but also become symbolised in hate targeting men, further stigmatising women and girls.

However, hate speech does not simply target women and girls in a vacuum; it shifts and adapts depending on the additional identity markers present (ethnicity, religion, etc.). Narrative analysis and workshop discussions revealed how **gender, ethnicity, religion,** and **skin tone** intersect to shape the nature of online abuse in Ethiopia.

Political events and internal conflicts fuel **'ethnic and gendered'** hate speech, while **religious** narratives often reinforce harmful rhetoric. For example, the association of women with Eve's original sin is frequently used to justify misogyny, and Muslim women face unique forms of abuse, particularly in relation to wearing the hijab. Additionally, **colourism** and anti-Blackness were identified as underexplored but pervasive issues.

Given the rich ethnic and religious diversity in Ethiopia, it is not surprising that each group faces distinct forms of hate speech. For example, women and girls are less likely to face overt **threats** or **aggressive** speech than ethnic groups, yet they are disproportionately subjected to hate involving **stereotypes** and **mockery**.

Ethnicity and politics further fuels inflammatory discourse, with figures from different backgrounds accused of betraying their country or aligning with foreign powers. Historical terms like “banda” and “traitor” are weaponised to incite real-world violence, as seen during the height of hostilities between the TPLF and Government forces. Additionally, religious rhetoric with words like “devil” and “infidel” intensifies divisions. These findings underscore the complex, multi-layered nature of hate speech in Ethiopia, where women’s identities are weaponised, political, ethnic and religious tensions are exacerbated.

Female and male **content creators** face distinct types of **genre-specific** abuse, often linked to industry norms and societal biases. For example, female educational content creators experienced notable levels of **mockery**, highlighting persistent barriers to women’s authority in intellectual spaces, while those in political and news discussions had their credibility questioned, and encountered high levels of **aggression**. Male educators and religious commentators have their credibility challenged, and those advocating for gender equality encountered backlash rooted in gender biases. Male content creators discussing politics or religion received vilification based on perceived affiliations. The patterns highlight the need for a nuanced approach that considers both gender dynamics and the cultural context of each field.

There are **platform-specific** and **overarching trends** in online abuse, signalling the need for both technological solutions and societal shifts to challenge and dismantle the norms that sustain gender-based violence online. For example, Workshop participants believe that TikTok’s design fuels polarisation, while YouTube’s structure allows politically motivated, ethnic and gendered abuse to flourish.

This study highlights a crisis in content moderation and regulatory enforcement, with a widespread **lack of trust** in platforms’ ability to tackle online abuse effectively. Despite existing guidelines against hate speech and other gendered harms, **enforcement remains weak** – stronger action from platforms is essential. An improvement in content moderation, stronger enforcement of regulations, and greater accountability from social media platforms, especially in countries with a diversity of languages, is necessary. Additionally, the persistent and misguided view that the online space is separate from the ‘real world’ must be overcome. Online and offline spaces are intrinsically linked; it is clear that offline discourse impacts online activity, and online harms do not stay online, their impacts are both significant and far-reaching.

Many of these findings align closely with CIR’s earlier research, which revealed that gendered abuse has become normalised, fostering a digital environment where women's contributions are routinely dismissed or ridiculed. By shining a light on

the pervasive forms of gendered abuse on social media, CIR hopes to prevent women and girls from being **pushed further into the margins.**

6. RECOMMENDATIONS

To make a real, lasting impact, any efforts to combat TFGBV must tackle its root causes, including challenging gender stereotypes, addressing gender-based discrimination, and ensuring women and girls have a strong and visible presence in all public spheres. Without targeted action, the normalisation of online abuse will continue to have far-reaching consequences for gender equality and public participation in Ethiopia.

CIR recommends a multi-pronged approach, focussing on education, developing or enacting law and policy, engaging the community, improving platform accountability, carrying out awareness campaigns, supporting vulnerable groups and continued research. As part of this project, CIR worked with stakeholders in Ethiopia to create a policy and community-led recommendations whitepaper (published May 2024) around these key themes. CIR hopes that government institutions can use the findings to inform decision making, that social media companies can use them to inform their content moderation efforts, that civil society can use them in their advocacy, and that the public can use them to call for action.

7. APPENDICES

7.1 GLOSSARY

Term	Definition
Hate Speech	Speech that deliberately promotes hatred, discrimination or attacks against a person or a discernible group of identity, based on ethnicity, religion, race, gender, or disability. (Ethiopian Government)
Online abuse	Online abuse is a broad term which encompasses many types of harmful behaviours that occur on the internet. The 'Online Harassment Field Manual'

	published by PEN America, defines online abuse as "pervasive or severe targeting of an individual or group online through harmful behaviour." This includes, and is not limited to, acts such as hatespeech, doxing, and sexual harassment.
Gender-based violence	"[H]armful acts directed at any individual or a group of individuals based on their gender. It is rooted in gender inequality, the abuse of power and harmful norms". (UN Women, Africa)
Technology-facilitated gender-based violence (TFGBV)	"[A]n act of violence perpetrated by one or more individuals that is committed, assisted, aggravated and amplified in part or fully by the use of information and communication technologies or digital media against a person on the basis of their gender." (UNFPA)

7.2 EXISTING CIR PUBLICATIONS ON TFGBV IN ETHIOPIA:

Research Summary (May, 2024)

- English
- Amharic
- Afaan Oromo
- Tigrigna

Report 1: Silenced, shamed, and threatened (May, 2024)

- English
- Amharic

Report 2: Normalised and Invisible (May, 2024)

- English
- Amharic

Recommendations for combating TFGBV in Ethiopia (May, 2024)

- English
- Amharic

Inflammatory Keyword List (Lexicon)

- CIR Github

Academic Journal article – Journal, *Resources for Annotating Hate Speech in Social Media Platforms Used in Ethiopia: A Novel Lexicon and Labelling Scheme*

- The Resources for African Indigenous Languages Journal

For more information, visit the CIR website.

CONFERENCES:
A conference was held by The Ethiopian Human Rights Defenders Centre, the Ethiopian Women's Human Rights Defenders Network and CIR in Addis Ababa to raise awareness about TFGBV in Ethiopia (May 9, 2024). For more information, see the CIR website.

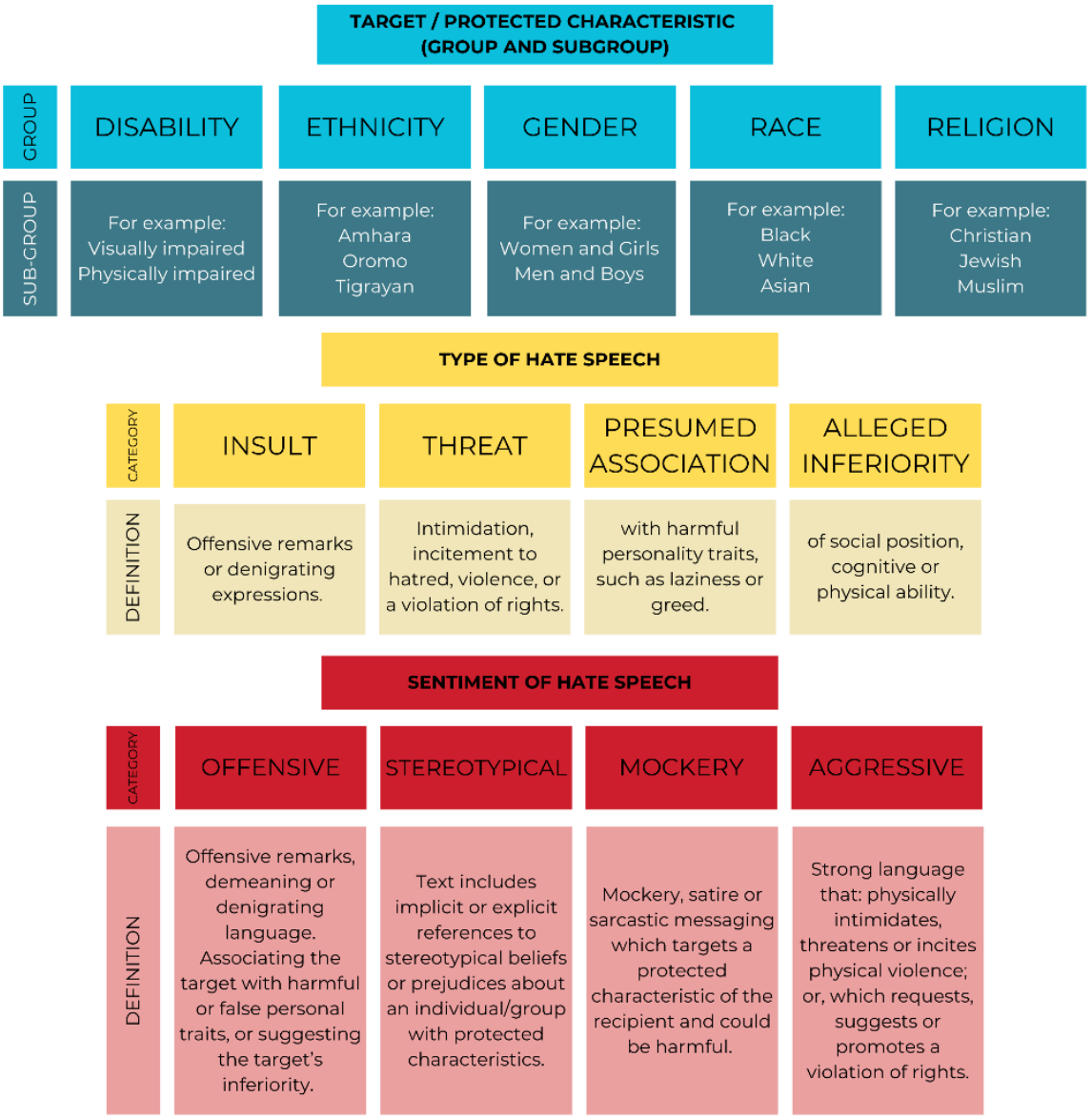
7.3 ANNOTATION PROTOCOL

The following guide was provided to the data annotation team to ensure that data annotation was consistent between annotators and across languages.

THE PROTOCOL

To limit the impact of individual biases during the data annotation process, CIR asks each annotator to follow these annotation protocol, which outline the different variables that should be considered during the labelling of online posts. Please read these guidelines carefully and refer back to them frequently during the study.

If a post contains hate speech (as defined by the Ethiopian Government) three elements need to be labelled using the annotation tool: the target, the type of speech and the sentiment of speech. The categories pertaining to each of these elements are visually depicted in the diagram below.



NOTE TO ANNOTATORS:

The views expressed during this research are not those of the investigators and the research team do not condone the opinions expressed. As the dataset contains real messages/posts sent by users on TikTok and YouTube, this research may expose the annotators to offensive and harmful content. Please take steps to protect yourself from the impacts of vicarious trauma. For example, please take regular breaks. Resources on 'Vicarious Trauma' have been shared with you.

TARGET OF HATE SPEECH

A post that contains hate speech is targeted towards an individual or group with a protected characteristic. To capture this information, the words that convey which individual/group is being targeted should be assigned any of the following labels (in line with the Ethiopian Government's definition of hate speech):

- **Gender:** An individual or group of people of a particular gender. Although not included in the Ethiopian Government's definition, this study includes sexual orientation, e.g., homosexuals in this category. This is because interviewees reported being labelled as homosexual by abusers.
- **Ethnicity:** An individual or group of people who come from a particular place of origin and culture.
- **Religion:** An individual or group of people belonging to a particular religious group.
- **Race:** An individual or group of people possessing distinctive physical traits associated with a particular race.
- **Disability:** An individual or group of people possessing a particular disability.

Example A1: *Stupid women and girls shouldn't talk about political issues. Someone should throw acid at her face, maybe then she will shut up.*

In the above example, the target should be labelled as **Gender** since the words "women and girls", "her" and "she" indicate that women and girls are the target of hate speech.

Example A2: *@airline how about donating flights to deport the Invaders back to their homeland. #DeportThem*

In the above example, the target should be labelled as **Ethnicity** since the word "Invaders" indicates that the hate speech is intended for a group of people whose country of origin is different from that of the speaker (even if the specific group is not explicitly mentioned).

TYPE OF HATE SPEECH

A post that contains hate speech uses language that spreads, incites, promotes, or justifies hatred, discrimination, dehumanisation, intimidation, or violence towards the target (individual/group). To capture this information, the words that convey such language should be assigned any of the following labels.

2.1 Insult: Insults or denigrating expressions against an individual/group due to protected characteristics.

Example B1: *Fucking clueless women and girls should stay in the kitchen and not ruin a good man's name.*

2.2 Threat: Intimidation, threats or incitement to hatred, violence or violation of individuals' rights, due to protected characteristics, such as:

- bodily harm and threats to physical safety (to individual or family members)
- rape threats or sexual harassment
- image-based sexual abuse (also referred to as 'revenge pornography')
- death threats
- arbitrary arrests
- restriction of access to services
- doxing
- online harassment
- creation of videos/memes

Example B2: *I'll fucking kill the RELIGIOUS_GROUP pig.*

2.3 Presumed Association: Presumed association of protected characteristics with any of the following negative connotations:

- propensity to crime or violence
- laziness or other vices such as greed, alcoholism, cleanliness level
- fundamentalism or terrorism
- invasion or conquest of another location
- threat to national security
- threat to welfare/competitors in distribution of resources

Example B3: *Fucking ETHNIC_GROUP always take more than they deserve. Scroungers!*

2.4 Alleged Inferiority: References to the alleged inferiority (or superiority) of individual/group with a protected characteristic, for example, in relation to:

- social position
- credibility (e.g. defamation)
- cognitive ability
- physical ability
- ability to engage in societal activities, e.g. politics
- undesirable or inferior lifestyle/cultural practises
- dehumanisation or association with animals or entities considered inferior

Example B4: *They ETHNIC_GROUP shouldn't be allowed to vote, they don't contribute to our society. They must be silenced.*

SENTIMENT OF HATE SPEECH

The sentiment of hate speech must be labelled as any of:

3.1 Aggressive: This includes strong language that physically intimidates, threatens or incites physical violence against the recipient, or which requests, suggests or promotes a violation of the recipients' rights.

Example C1.a*: I'll fucking kill the RELIGIOUS_GROUP pig.*
Example C1.b: *Stupid women and girls shouldn't talk about political issues. Someone should throw acid at her face, maybe then she will shut up.*

3.2 Offensive: This includes several different forms of speech, from insulting, demeaning or denigrating language, to associating the target (individual or group) with harmful or false personal traits, or suggesting the target's inferiority.

Example C2: *Stupid immigrants... they come and take our resources. They make us weak. They are the enemy.*

3.3 Mockery: This includes jokes, satire or sarcastic messaging which targets a protected characteristic of the recipient and could be harmful. Hateful content is sometimes conveyed using nuances in language, such as sarcasm, mockery, or satire. Previous studies have expressed the importance of not overlooking this form of hate speech.

Example C3: *@xxxxx Ok hoe or whore you choose sweetie?*

3.4 Stereotypical: Text includes implicit or explicit references to stereotypical beliefs or prejudices about an individual/group with protected characteristics.

Example C4: *Fucking clueless women and girls should stay in the kitchen and not ruin a good man's name.*

Please note, the labels are not mutually exclusive. Something can be both offensive and stereotypical.

IMPLEMENTING THE ANNOTATION PROTOCOL: RULES

Dialogue between the annotators and data engineers led to the creation of a series of rules, outlined below, to ensure consistency during annotation.

Rule 1: Take the text at face value

Investigators should not infer meaning from the text; the text must be taken at face value. For example, although the following messages from the dataset contained offensive language, the investigators were advised not to annotate the text as they do not make sense.

Pitchfork Dirty van bitches with all that his bagpipes again, and, to give a toast is pain Surviving on toast is your skin

The highest high, I'm posed to a dyke bitch 'til she wanna Mac go I don't give me scarred I wake

Similarly, if the target is unclear, investigators were advised not to annotate. The following examples exemplify this challenge.

It's just so hard to let you dusty pieces of crap to do the bare minimum when my bitches show out every fucking time….

In the above example, it is unclear whether the individuals targeted here are being targeted because of a protected characteristic.

You are not muslim so crime lenesu normal new

The above example translates roughly to "you are not muslim so crime is normal for them". While 'crime is normal for them' could fall under 'presumed association' of a protected characteristic with crime, it is unclear who the target is. This may have been meant as hate speech, albeit written with a number of grammatical errors, however, no inferences were allowed.

Yehen aswged ant 666 new

The above example translates roughly to: 'Terminate/finish this one, it is 666'. While '666' implies Satanism, and it is inciting violence, the target is unclear.

As this investigation relies solely on textual information, context is lost. As a result, the sentiment of the post may not be captured, such as irony. Additionally, abusive terms may have multiple meanings. For example, the term 'public toilet' is used in gendered abuse. According to CIR workshop participants, this term is used against women and girls who people consider sexually promiscuous. Everyone can use a public toilet. Thus, the use of this word against women and girls implies that anyone can use them and that they are unclean/impure. This is both offensive and derogatory. Without context, however, it is hard to tell if the term 'public toilet' is being used against a woman, or in a different context altogether. As a result, the following example could be gendered abuse, complaining about a woman moaning during intercourse, or it could be talking about a public toilet:

They should make a public toilet that isn't so miserably loud when it flushes

This rule may have resulted in hate speech being excluded from the study. While this could be seen as a limitation, it prevented any personal biases from impacting the annotation process. This ensures that the hate speech dataset that is created is robust and reflects the current definition as set out by the Ethiopian Government.

Rule 2: The importance of protected characteristics

For a post to be classified as hate speech, it should target a protected characteristic (gender, race, religion, ethnicity or disability) under the current Ethiopian Government definition.

Sometimes the text was highly offensive; however, if no protected characteristics were targeted, then this does not classify as hate speech under the current definition. For example, the following is not hate-speech:

> *@USERNAME graduated from cunt university with honors with a pHd in serving*

The annotation task revealed a lot of hate, threats and incitement to violence directed against President Abiy and his political party, or other political organisations. However, political views and affiliations are not protected characteristics under the Ethiopian Government's definition of hate speech. This only classifies as hate speech when a protected characteristic is mentioned. For example, the following is not hate speech:

> *@USERNAME: #AmharaGenocide by the fascist #AbiyAhmedAli #JusticeForEthiopia*

However, if the above example made reference to Abiy's ethnicity, then this would classify as hate speech. For example:

> *@USERNAME: #AmharaGenocide by the fascist Oromo #AbiyAhmedAli #JusticeForEthiopia #OromoFascism*

Similarly, a Telegram/X (formerly Twitter) user may be targeted for their views. 'Viewpoints' (like political affiliations) are not protected characteristics. This does not classify as hate speech under the current definition unless a protected characteristic is targeted.

Sometimes posts contained information and misinformation about war crimes and human rights violations inflicted on an ethnic/religious/gender group and/or committed by certain armed/political groups. While this could amount to incitement to hatred against a particular entity, this is not hate speech unless there is also a protected characteristic targeted within the text.

Rule 3: Dealing with multiple languages

When posts contained multiple languages, the investigators consulted the wider annotation team which included four hate speech experts spanning four languages: Amharic, Afaan Oromo, Tigrigna, and English. If the text included languages that were not included in this study, the investigators were advised not to annotate the text. If the text included any combination of this investigation's four languages, the team annotated the text together. This ensured that the study remained focussed on the four languages being analysed, however it also means that hate speech may have been excluded from the study. Another implication of this is that text may have been pulled within, for example, the English dataset, but it may have contained Afaan Oromo. When analysing the results of the study it is important to bear this in mind.

Rule 4: Copypasta

During the exercise, a number of 'copypasta' texts were identified. Copypasta is a block of text shared on the internet which is literally copy and pasted by multiple users. It often reveals coordination in information sharing. If the copypasta text contained hate speech, as per the guidelines, each instance was annotated. If they didn't conform with the annotation protocol, they were not annotated. Interestingly, this led annotators to identify a number of copypasta websites including Ethiopian Truth.

Rule 5: Additional hate speech terms

When terms which could be indicative of hate speech were identified that were not in the study's hate speech lexicon, the investigators were advised to write these down. These will be added to the final hate speech lexicon that is published alongside the report, to ensure that the lexicon is as comprehensive as possible to aid future research. For example, the terms "flour ranger" and "kufr" were identified within the English dataset and added to the lexicon.

7.4 BIBLIOGRAPHY

Akshita Jha and R. Mamidi (2017) 'When does a compliment become sexist? Analysis and classification of ambivalent sexism using X (formerly Twitter) data', NLP+CSS@ACL, Available here.

Deborah James (1998) 'Gender-Linked Derogatory Terms and Their Use by Women and girls and Men', American Speech, Vol. 73, No. 4 (Winter, 1998), Duke University Press, pp. 399-420.

Gao, L., Kuppersmith, A., & Huang, R. (2017). Recognizing explicit and implicit hate speech using a weakly supervised two-path bootstrapping approach. arXiv preprint arXiv:1710.07394.

Gashe, S. M. (2022). Hate Speech Detection and Classification System in Amharic Text with Deep Learning. List of Amharic Hate Speech Keywords (Lexicons). Available here.

Mekuanent Degu (2022), "Amharic dataset for hate speech detection", Mendeley Data, V3, doi: 10.17632/fhvsvsbvtg.3

Samuel Minale (2022), "Amharic Social Media Dataset for Hate Speech Detection and Classification in Amharic Text with Deep Learning", Mendeley Data, V1, doi: 10.17632/p74pfhz3yx.1

Surafel Getachew (2020), "Amharic Facebook Dataset for Hate Speech detection", Mendeley Data, V1, doi: 10.17632/ymtmxx385m.1

REPORTS AND WEB ARTICLES USED IN THE LEXICON DEVELOPMENT:

CARD's Bi-weekly Social Media Conversation Sensitivity Report, see here.

David Shariatmadari (2016) 'Eight words that reveal the sexism at the heart of the English language' Available here [last accessed 9 Nov 2023].

Hatebase.org (2023) Available here [last accessed 9 Nov 2023].

Hate Speech Dataset Catalogue, Available here [last accessed 9 Nov 2023].

Peace Tech Lab (n.d.) 'Hateful Speech and Conflict in the Federal Democratic Republic of Ethiopia: Alexicon of hateful of inflammatory words and Phrases' Available here [last accessed 9 Nov 2023].

Thalikir 2016 'Everyday misogyny: 122 subtly sexist words about women and girls (and what to do about them)' Available here [last accessed 9 Nov 2023].

The Wilson Centre (2021) 'Malign Creativity: How Gender, Sex, and Lies are Weaponized Against Women and girls Online' Available here [last accessed 9 Nov 2023].

7.5 FUNDING

This material has been funded by UK International Development; however, the views expressed do not necessarily reflect the views of UK government.