

Confronting political violence against women

Using safety tools on Facebook, Instagram, WhatsApp and Messenger



Contents

01	Introduction	03
02	Tools to prevent and contain political violence against women	05
	📄 Reporting content	06
	🗨️ Hate speech	08
	👥 Bullying and harassment	10
	📄 Misinformation	13
03	Preventing political violence against women	14



1

Introduction

1

Introduction

At Meta, we are committed to protecting your voice and helping you connect and share safely. Women should have equal access to all of the economic opportunities, education and social connections that the Internet provides, which is why we take a comprehensive approach to making our technologies a safer place for women, including writing, creating policies, engaging with experts and developing cutting-edge technology to help prevent abuse from happening in the first place.

Our Community Standards and guidelines outline the rules for what is and isn't allowed across Meta technologies. They're developed by our policy teams and include rules against behaviours that disproportionately affect women, such as the sharing of non-consensual intimate imagery, which is illegal in many places around the world. They also include rules against harassment, such as sending multiple unwanted messages to a person who's made it clear that they don't want to receive them. We have processes that incorporate technologies, such as artificial intelligence that, together with our teams, work to eliminate and reduce harmful content.



In addition to proactive detection, policies and technology to stop threatening behaviour towards women, we offer a number of tools to help people take control of their Meta experience. In parallel, we also continue to refine those tools, as well as our policies and resources, continuously adapting to both online and offline changes and adopting feedback from external specialists such as our expert advisors on global women's safety. This guide provides tools and tips for engaging safely with community members on:



2

Tools to prevent and contain political violence against women

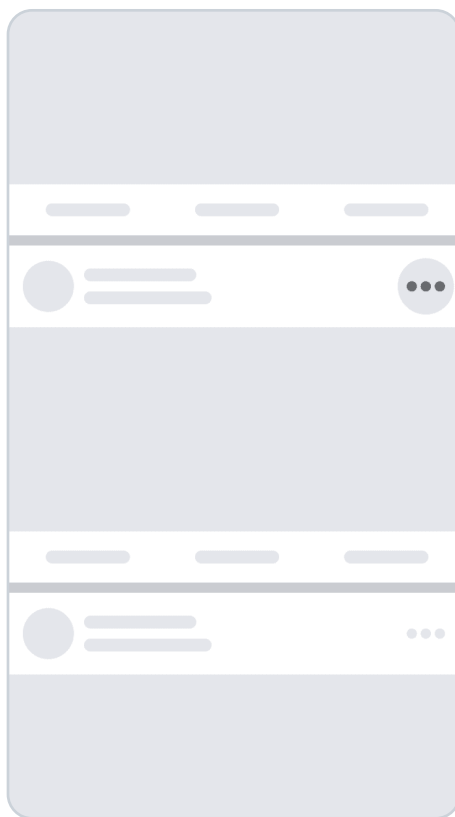
2

Tools to prevent and contain political violence against women

Political violence against women is expressed through different manifestations. The following tools can help prevent or contain some of the violence that occurs in digital spaces, and are a complement to reports or actions that can bring attention to corresponding authorities.

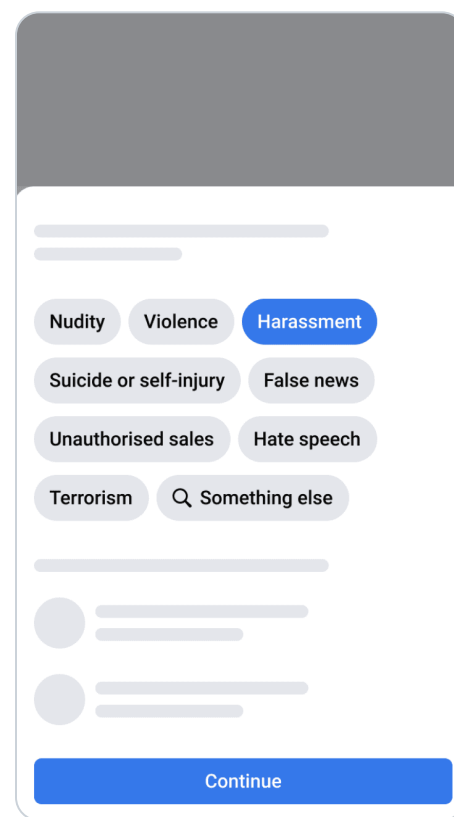


REPORTING CONTENT



01 Universal entry point:

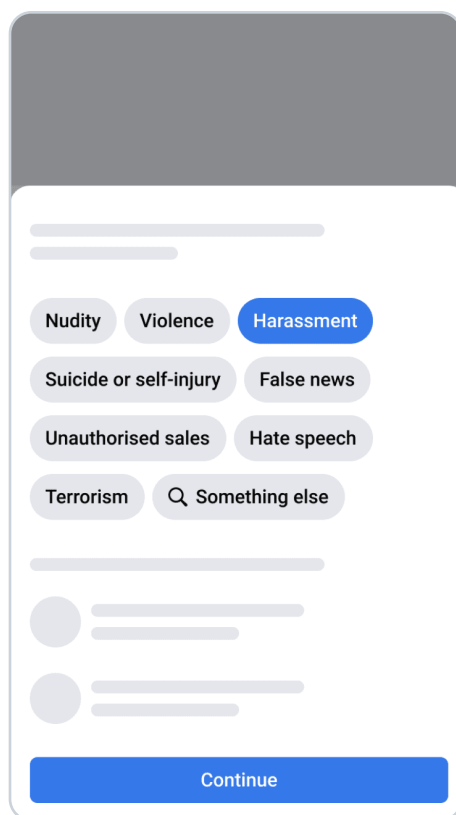
We have an option to report – whether it's on a post, a comment, a story, a message or something else.



02 Begin the report:

We help people report things that they don't think should be on our technologies.

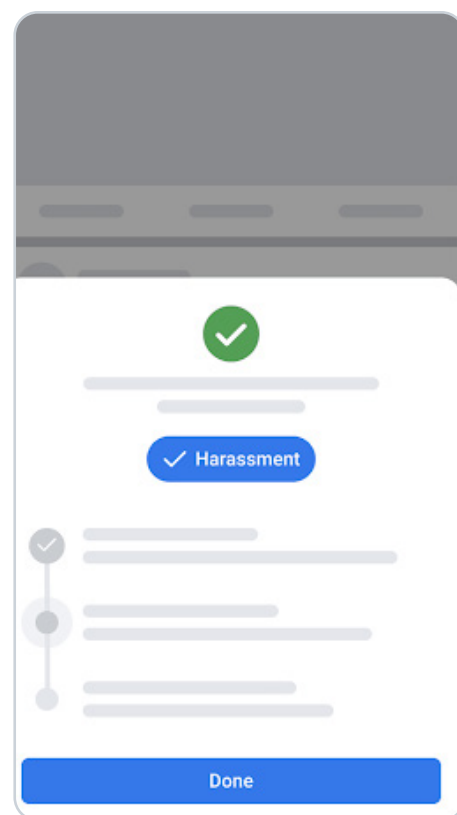
REPORTING CONTENT (CONTINUED)



A mobile app screen showing the 'Select a problem' step. At the top, there are two horizontal lines for text input. Below these are several buttons: 'Nudity', 'Violence', 'Harassment' (highlighted in blue), 'Suicide or self-injury', 'False news', 'Unauthorised sales', 'Hate speech', 'Terrorism', and 'Q Something else'. At the bottom, there is a blue button labeled 'Continue'.

03 Select a problem:

We ask people to tell us more about what's wrong. This helps us send the report to the right place.



A mobile app screen showing the 'Report submitted' step. At the top, there is a green checkmark icon. Below it, there is a blue button labeled '✓ Harassment'. To the left of the main content area is a vertical progress indicator with three circles; the middle circle is filled, indicating the current step. Below the progress indicator are three horizontal lines for text input. At the bottom, there is a blue button labeled 'Done'.

04 Report submitted:

After these steps, we submit the report. We also lay out what people should expect next.



For more information on reporting content, view our policies in the [Meta Transparency Centre](#).



HATE SPEECH

People use their voice and connect more freely when they don't feel attacked on the basis of who they are. That is why we don't allow hate speech on our technologies. It creates an environment of intimidation and exclusion, and in some cases may promote offline violence. We define hate speech as a direct attack against people – rather than concepts or institutions – on the basis of what we call protected characteristics: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease.

TOOLS TO PREVENT HATE SPEECH

- ✓ **Avoid unwanted messages**
Learn more for [Facebook](#), [Instagram](#), [Messenger](#) and [WhatsApp](#)
- ✓ **Moderation Assist:** Prevent the display of offensive comments.
Learn more for [Facebook](#) and [Instagram](#)
- ✓ **Disable comments:** You can turn off commenting to prevent people from commenting on an individual post. When you turn off commenting, any comments that are currently on your post will be hidden.
Learn more for [Facebook](#) and [Instagram](#)
- ✓ **Pin positive comments:** Promote positive behaviour.
Learn more for [Instagram](#)

HATE SPEECH (CONTINUED)

TOOLS TO CONTAIN HATE SPEECH

-  **Report offensive comments and threatening messages:** Identify and remove content if it violates our [Community Standards](#).
Learn more for [Facebook](#) and [Instagram](#)
-  **Block profiles, Pages, groups and apps:** When you block someone, they can no longer see your posts, tag your Page, invite your Page to events or groups or start a conversation with your Page. Public content may still be visible to people who are logged out of a blocked account.
Learn more for [Facebook](#), [Instagram](#) and [WhatsApp](#)
-  **Restrict someone:** Restrict comments from someone without them knowing.
Learn more for [Facebook](#) and [Instagram](#)
-  **Manage spam comments:** Spam involves contacting people with unwanted content or requests. This includes sending bulk messages and excessively posting links or images to people's timelines.
Learn more for [Facebook](#) and [Instagram](#)

On Facebook, prevalence of hate speech in Q4 2023 was 0.01%-0.02%. In other words, of every

10K

content views, an estimate of one to two would contain hate speech.²



2



BULLYING AND HARASSMENT

Bullying and harassment happen in many places and come in many different forms, from making threats and releasing personally identifiable information to sending threatening messages and making unwanted malicious contact. We do not tolerate this kind of behaviour because it prevents people from feeling safe and respected on Meta technologies.

We distinguish between public figures and private individuals because we want to allow discussion, which often includes critical commentary of people who are featured in the news or who have a large public audience. For public figures, we remove attacks that are severe as well as certain attacks where the public figure is directly tagged in the post or comment. We define public figures as state- and national-level government officials, political candidates for those offices, people with over 1 million fans or followers on social media and people who receive substantial news coverage.

TOOLS TO PREVENT BULLYING AND HARASSMENT



Comment warning: When someone tries to post a potentially offensive comment, we'll show a warning reminding them of our Community Guidelines and letting them know that we may remove or hide their comment if they proceed.



Tag and mention controls: You can choose whether you want everyone, only people you follow or no one to be able to tag or mention you in a comment, caption or story.

BULLYING AND HARASSMENT (CONTINUED)

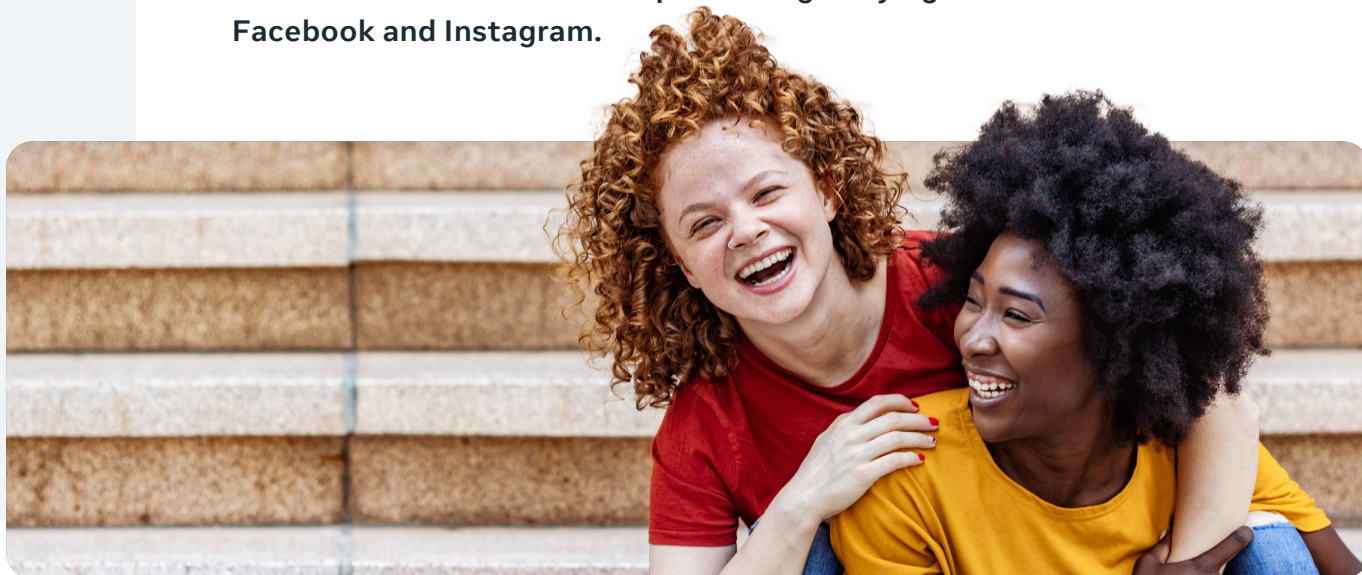
TOOLS TO PREVENT BULLYING AND HARASSMENT

- ✓ **Hidden words:** Comments and message requests that don't go against our Community Guidelines but may be considered inappropriate or offensive can be filtered out.
- ✓ **Limits:** Limit unwanted interactions for a period of time by automatically hiding comments and message requests from people who don't follow you.
- ✓ **Restrict:** When you restrict someone, their comments on your posts will only be visible to that person.
- ✓ **Locked profiles:** Lock your profile is available in some regions. It helps prevent unwanted people from seeing your photos, posts and information about you.
- ✓ **Limits to address unwanted direct messages:** New feature designed to help better protect our community from unwanted images and videos in your direct messages.

Invitation model: Send people one message – which will act as an invitation to keep chatting. You will only be able to continue to send them a direct message when they've accepted your request to chat.

Text only: This one invitation message is limited to text only. That means that you can only send images, videos or voice notes to someone who doesn't follow you after they've accepted your request to chat.

Learn more about how we're preventing bullying and harassment on Facebook and Instagram.



BULLYING AND HARASSMENT (CONTINUED)

TOOLS TO CONTAIN BULLYING AND HARASSMENT AND PREVENT IMAGE ABUSE

- ✓ **Report:** If you see something that violates our Community Standards, report it so that our team can review it and take action. Our teams review reports 24/7 in more than 70 languages and will take action on violating content and behaviour.

Learn more for [Facebook](#) and [Instagram](#)

- ✓ **Block:** When you block someone, they won't be able to see your profile, posts or stories.

Learn more for [Facebook](#) and [Instagram](#)

- ✓ **Report nude and sexual photos or videos of yourself or threats to share** these images and videos to our technologies to prevent them from being reshared.

Learn more for [Facebook](#), [Instagram](#) and [Messenger](#)

- ✓ **Use [StopNCII.org](#) to prevent a nude or sexual photo or video of yourself from being shared online:**

- If you have an intimate image or video that was taken when you were over 18 and are concerned that it will be shared or reshared online, you can take steps to help prevent further circulation through [StopNCII.org](#).

- The tool uses hash-generating technology that assigns a unique hash value (a numeric code) to an image, creating a secure digital fingerprint. Tech companies participating in [StopNCII.org](#) receive the hash and can use that hash to detect if someone has shared or is trying to share those images on their platforms.

On Instagram, we found and actioned on

95.3%
of violating content before people reported it.³

On Facebook, prevalence of hate speech in Q4 2023 was 0.01%-0.02%. In other words, of every

10K
content views, an estimate of one to two would contain hate speech.³



MISINFORMATION

Meta is committed to stopping the spread of misinformation. We use a combination of enforcement technology, human review and independent fact-checkers to identify, review and take action on this type of content.



REMOVE

We value both free expression and keeping people safe, so we remove misinformation from Meta technologies in limited cases:

- When misinformation has the potential to cause imminent physical harm.
- When misinformation has the potential to interfere with suppress voting.
- When videos are manipulated in ways that would not be apparent to an average person – and would mislead an average person to believe that a subject of the video said words they did not say – and were created using artificial intelligence that realistically combines, replaces or superimposes content onto them.



Taking down violating content



REDUCE

When one of our fact-checking partners rates something as false, we reduce the content's distribution in Feed and other places so that fewer people see it. When Pages and groups repeatedly share such content, we reduce the distribution of all of their posts in Feed and remove them from the recommendations that we show people.



INFORM

We apply notices to fact-checked posts and send notifications to the people who posted it. That way, people can see what the fact-checkers concluded and decide for themselves what to read, trust or share. We also partner with organisations around the world to promote news literacy.



Reducing the distribution of problematic content



Providing context on sensitive or misleading content

03

Preventing political violence against women

3

Preventing political violence against women

Our commitment to women's safety is longstanding. Over the years, we've sought the help of experts in the field to ensure that our technologies are safe for women and have:

- Developed strong policies to protect women from online abuse.
- Built tools to empower users with control over their experience online.
- Launched technology to combat the sharing of non-consensual intimate images.

HOW YOU CAN HELP PREVENT POLITICAL VIOLENCE AGAINST WOMEN

**Raising awareness:**

Political violence against women is a problem that has not had sufficient visibility, and statistical information remains scarce.

**Use our technologies to expose acts of violence:**

Speaking out and sharing first-person stories of situations in which you have experienced political violence is important in the fight against it.

**Familiarise yourself with our Community Standards:**

It is our responsibility to keep our services free from abuse. This is why we developed a set of Community Standards that describe what is and is not allowed on our technologies.

**Engage with Meta to protect yourself during election cycles:**

Political accounts may face additional threats during an election cycle and may need additional protection. Facebook Protect was created, offering candidates, elected officials, federal and state departments and agencies and party committees, as well as the people who work with them, a way to better protect their Facebook accounts.