

Crisis misinformation policy



1. [Help Center](https://help.twitter.com/) ^ (https://help.twitter.com/)
2. [Platform integrity and authenticity](https://help.twitter.com/en/rules-and-policies#platform-integrity-and-authenticity) ^ (https://help.twitter.com/en/rules-and-policies#platform-integrity-and-authenticity)

Crisis misinformation policy

Armed conflict

Overview

May 2022

We will take action on accounts that use Twitter's services to share false or misleading information that could bring harm to crisis-affected populations.

A crisis is any situation in which there is a widespread threat to life, physical safety, health, or basic subsistence that is beyond the coping capacity of individuals and the communities in which they reside.

In times of crisis — such as situations of armed conflict, public health emergencies, and large-scale natural disasters — false and misleading information has a special capacity to bring harm to vulnerable populations. For these reasons, Twitter is taking action to mitigate misleading information that has the potential to shape crisis dynamics and put vulnerable people in harm's way.

Currently, the scope of this policy includes international armed conflict. In these contexts, we're focusing on misleading information with the capacity to:

- Serve as a pretext for further aggression by armed actors, belligerents, or combatants,
- Trigger forced or anticipatory displacement of vulnerable populations, or lead to increased humanitarian needs,
- Negatively impact the ability of humanitarian protection, human rights, or relief organizations to provide assistance or access affected populations,
- Incite the targeting or surveillance of groups that can be identified based on their political, religious, ethnic or ideological affiliation or membership, or organizations and actors protected by international humanitarian law;
- Disrupt potential ceasefire agreements, peacekeeping operations, or diplomatic solutions to conflict or insecurity, among other matters.

We will update this page in the future as this policy expands to cover additional types of crisis events.

What is in violation of this policy?

Criteria

In order for conflict-related content to be considered violative under this policy, it must:

- advance a claim of fact, expressed in definitive terms;
- be demonstrably false or misleading, based on widely available, authoritative sources; and
- be likely to impact public safety or cause serious harm.

In the context of international armed conflict, we will take action on Tweets that contain:

- False coverage or event reporting, or information that mischaracterizes conditions on the ground as a conflict evolves.
- False allegations regarding use of force, incursions on territorial sovereignty, or around the use of weapons.
- Demonstrably false or misleading allegations of war crimes or mass atrocities against specific populations;
- False information regarding international community response, sanctions, defensive actions, and or humanitarian operations.

What is not a violation of this policy?

The following types of information/expressions are generally not violative of the policy:

- **Strong commentary, opinions, and/or satire**, provided these do not contain false or misleading assertions of fact, are not violative of the policy.
- **Information seeking** (i.e. asking legitimate questions about an event in question) is not violative of the policy, unless there are signals that the content represents a pattern of unhealthy behavior throughout the duration of a crisis.
- **Debunking or fact-checking** (i.e. calling attention to the claim in question to refute it) is not violative of the policy. We allow for reproductions of false information or misleading content that provides analysis on if/how the content is misleading.
- **News reporting** – in most cases, fact-based descriptions of news-worthy events, even if those descriptions contain or involve misleading information, is not violative of the policy.
- **Personal anecdotes or first-person accounts**. We do not enforce the crisis misinformation policy on reports or first-person accounts of specific events or incidents that are otherwise un-verifiable. *Note that other parts of the Twitter Rules, including our policies about misleading and deceptive identities and platform manipulation, will also apply.*

What happens if you violate this policy?

Account eligibility

We aim to reduce harm by limiting impressions on violative content on Twitter. We will prioritize enforcement of highly visible Tweets and Tweets from high profile accounts, such as state-affiliated media accounts, verified, official government accounts, and/or Tweets with a large number of impressions or engagements (Likes, Retweets, etc).

Warning notice

When Tweets include misleading information related to an active international armed conflict, we may place the Tweet behind an interstitial advising viewers that the Tweet has violated our crisis misinformation policy. We will disable the ability for others to Retweet, like, or engage in other ways to prevent the spread of the misleading information and the notice will link to the Twitter Rules. We are doing this to preserve a record of the violative content on the platform while ensuring users are not directly exposed to harmful misleading information.

Reducing visibility of content

Finally, we may temporarily reduce the visibility of Tweets or accounts that we believe with high confidence to be in violation of the crisis misinformation policy. If we determine that the accounts have continued to violate the policy, amplification limitations can be reapplied.

We may reduce visibility by:

- Making Tweets and Retweets from those accounts ineligible for recommendation in certain parts of the Twitter product (such as top Search results)
- Displaying Replies from the account in a lower position in conversations
- Excluding Tweets by the account itself in email or in-product recommendations

Strike system

In instances where accounts repeatedly violate this policy, we will use a strike system to determine if further enforcement actions should be applied. Two notices applied within 30 days will result in a 12-hour account time-out. Three or more notices applied within 30 days will result in 7-day time-outs.

If you believe that your account was locked in error, you can [submit an appeal](https://help.twitter.com/forms/general?subtopic=suspended) (<https://help.twitter.com/forms/general?subtopic=suspended>).

Additional resources

Learn more about our ongoing approach to the war in Ukraine [here](https://blog.twitter.com/en_us/topics/company/2022/our-ongoing-approach-to-the-war-in-ukraine) (https://blog.twitter.com/en_us/topics/company/2022/our-ongoing-approach-to-the-war-in-ukraine), and about our other misinformation policies [here](#).

Also, be sure to check out [our range of enforcement options](#) and [our approach to policy development and enforcement](#).

Share this article



Was this article helpful?^{*}



Submit



© 2022 Twitter, Inc.

[Cookies](https://help.twitter.com/rules-and-policies/twitter-cookies) (<https://help.twitter.com/rules-and-policies/twitter-cookies>)

[Privacy](https://twitter.com/privacy) (<https://twitter.com/privacy>)

[Terms and conditions](https://twitter.com/tos) (<https://twitter.com/tos>)