

Violence and Incitement

Policy details

Policy Rationale

We aim to prevent potential offline violence that may be related to content on our platforms. While we understand that people commonly express disdain or disagreement by threatening or calling for violence in non-serious and casual ways, we remove language that incites or facilitates violence and credible threats to public or personal safety. This includes violent speech targeting a person or group of people on the basis of their protected characteristic(s) or immigration status. We remove content, disable accounts and work with law enforcement when we believe there is a genuine risk of physical harm or direct threats to public safety. We also try to consider the language and context in order to distinguish casual or awareness-raising statements from content that constitutes a credible threat to public or personal safety. In determining whether a threat is credible, we may also consider additional information such as a person's public visibility and the risks to their physical safety.

In some cases, we see aspirational or conditional threats of violence, including expressions of hope that violence will be committed, directed at terrorists and other violent actors (e.g., "Terrorists deserve to be killed," "I hope they kill the terrorists"). We deem those non-credible, absent specific evidence to the contrary.

We Remove:

We remove threats of violence against various targets. Threats of violence are statements or visuals representing an intention, aspiration, or call for violence against a target, and threats can be expressed in various types of statements such as statements of intent, calls for action, advocacy, expressions of hope, aspirational statements and conditional statements.

We do not prohibit threats when shared in awareness-raising or condemning context, when less severe threats are made in the context of contact sports, or certain threats against violent actors, like terrorist groups.

Universal protections for everyone

Everyone is protected from the following threats:

- Threats of violence that could lead to death (or other forms of high-severity violence)
- Threats of violence that could lead to serious injury (mid-severity violence). We remove such threats against public figures and groups not based on protected characteristics when credible, and we remove them against any other targets (including groups based on protected characteristics) regardless of credibility

- Admissions to high-severity or mid-severity violence (in written or verbal form, or visually depicted by the perpetrator or an associate), except when shared in a context of redemption, self-defense, contact sports (mid-severity or less), or when committed by law enforcement, military or state security personnel
- Threats or depictions of kidnappings or abductions, unless it is clear that the content is being shared by a victim or their family as a plea for help, or shared for informational, condemnation or awareness-raising purposes

Additional protections for Private Adults, All Children, high-risk persons and persons or groups based on their protected characteristics:

In addition to the universal protections for everyone, all private adults (when self-reported), children and persons or groups of people targeted on the basis of their protected characteristic(s), are protected from threats of low-severity violence.

Other Violence

In addition to all of the protections listed above, we remove the following:

- Content that asks for, offers, or admits to offering services of high-severity violence (for example, hitmen, mercenaries, assassins, female genital mutilation) or advocates for the use of these services
- Instructions on how to make or use weapons where there is language explicitly stating the goal to seriously injure or kill people, or imagery that shows or simulates the end result, unless with context that the content is for a non-violent purpose such as educational self-defense (for example, combat training, martial arts) and military training
- Instructions on how to make or use explosives, unless with context that the content is for a non-violent purpose such as recreational uses (for example, fireworks and commercial video games, fishing)
- Threats to take up weapons or to bring weapons to a location or forcibly enter a location (including but not limited to places of worship, educational facilities, polling places or locations used to count votes or administer an election), or locations where there are temporary signals of a heightened risk of violence.
- Threats of violence related to voting, voter registration, or the administration or outcome of an election, even if there is no target.
- Glorification of gender-based violence that is either intimate partner violence or honor-based violence

For the following Community Standards, we require additional information and/or context to enforce:

We Remove:

- Threats against law enforcement officers or election officials, regardless of their public figure status or credibility of the threat.
- Coded statements where the method of violence is not clearly articulated, but the threat is veiled or implicit, as shown by the combination of both a threat signal and contextual signal from the list below.

Threat: a coded statement that is one of the following:

- Shared in a retaliatory context (e.g., expressions of desire to engage in violence against others in response to a grievance or threat that may be real, perceived or anticipated)
- References to historical or fictional incidents of violence (e.g., content that threatens others by referring to known historical incidents of violence that have been committed throughout history or in fictional settings)
- Acts as a threatening call to action (e.g., content inviting or encouraging others to carry out violent acts or to join in carrying out the violent acts)
- Indicates knowledge of or shares sensitive information that could expose others to violence (e.g., content that either makes note of or implies awareness of personal information that might make a threat of violence more credible. This includes implying knowledge of a person's residential address, their place of employment or education, daily commute routes or current location)

Context

- Local context or expertise confirms that the statement in question could lead to imminent violence.
 - The target of the content or an authorized representative reports the content to us.
 - The target is a child.
- Implicit threats to bring armaments to locations, including but not limited to places of worship, educational facilities, polling places or locations used to count votes or administer an election (or encouraging others to do the same) or locations where there are temporary signals of a heightened risk of violence.
 - Claims or speculation about election-related corruption, irregularities, or bias when combined with a signal that content is threatening violence (e.g., threats to take up or bring a weapon, visual depictions of a weapon, references to arson, theft, vandalism), including:

- Targeting individual(s)
- Targeting a specific location (state or smaller)
- Where the target is not explicit
- References to election-related gatherings or events when combined with a signal that content is threatening violence (e.g., threats to take up or bring a weapon, visual depictions of a weapon, references to arson, theft, vandalism).
- Threats of high- or mid-severity violence in the defense of self or another human when the criteria below are met.
 - Against a person (excluding persons identifiable by name or face, people targeted based on their protected characteristics, and children)
 - In the context of home entry or interpersonal violence that is proportional to the violence responded to and is an immediate threat
 - The potential impact on voice outweighs the risk of imminent violence

User experiences

See some examples of what enforcement looks like for people on Facebook, such as: what it looks like to report something you don't think should be on Facebook, to be told you've violated our Community Standards and to see a warning screen over certain content.