Hate Speech & Reporting Tools for Social Media

Community Standards: Hate Speech

www.facebook.com/communitystandards/
hate_speech

Facebook defines hate speech as a direct attack on people based on protected characteristics, including race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, serious disease or disability, and immigration status.

Facebook categorizes **direct attacks** into **THREE tiers of severity**:

Tier One: Content targeting a person or group of people who share one of the above-listed protected characteristics with: (1) any violent speech or support in written or visual form; (2) dehumanizing speech that compares, references, or generalizes the person or group of people to/as: insects; animals culturally perceived as inferior; filth, bacteria, disease, and feces; sexual predators; subhuman; violent and sexual criminals; criminals such as thieves or bank robbers (includes generalizations that all members of the protected group are of this nature); statements denying existence; (3) mocking the concept, events or victims of hate crimes; and (4) designating dehumanizing comparisons, generalizations, or behavioral statements of various groups, including African Americans, Jews, Muslims, Hispanics and women.

Tier Two: an attack that targets a person or group of people who share any of the above-listed protected characteristics, where attack is defined as generalizations of inferiority, including physical, mental, and moral deficiencies; expressions of being less than adequate or deviating from the norm; expressions of contempt or visual equivalent; expressions of dismissal or disgust and cursing.

Tier Three: an attack that calls to exclude or segregate a person or group of people based on the above-listed protected characteristics. This also includes content that describes or negatively targets people with slurs, where slurs are defined as words commonly used as insulting labels for the above-listed characteristics.

Keep Facebook Accountable: Reporting Hate Speech

Report the post by clicking on the three dots in the upper-right corner of the post in question. Click on the "Find Support or Report Post" option, and select "Hate Speech." Click on the option that best describes how the post is hate speech, and click "Send."

Facebook may follow up with a formal report request.

Help Center: Hateful Conduct Policy

help.twitter.com/en/rules-and-policies/hateful-conduct-policy

Twitter users are prohibited to promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. This includes images or symbols in a profile image or header, and conduct applies to both tweets and direct messages.

Twitter prohibits hateful conduct to any individual or group of people within a protected group listed above that contains:

- **Violent threats**
- **Wishes, hopes or calls for serious harm on a person or group or people**
- **References to murder or violence against people or groups that have been victimized or the primary targets**
- **Incitements of fear**
- **Repeated/non-consensual slurs, epithets, racist or sexist tropes, or degrading content**
- **Hateful imagery**

Keep Twitter Accountable: Reporting Hate Speech

Users can report hate speech at help.twitter.com/forms/abusiveuser

Users are asked to share the tweet that they are reporting on the form and may include additional information in an information box.

Community Guidelines

help.instagram.com/477434105621119

Instagram removes content that contains credible threats or hate speech, content that targets private individuals to degrade or shame them, personal information meant to blackmail or harass someone, and repeated unwanted messages.

There is a no-tolerance policy on Instagram for encouraging violence or attack on anyone based on his or her race, ethnicity, national origin, sex, gender, gender identity, sexual orientation, religious affiliation, disabilities, or diseases. When hate speech is being shared to challenge it or to raise awareness, Instagram may allow it. In those instances, Instagram requires that you express intent clearly.

Keep Instagram Accountable: Reporting Hate Speech

If you see something that you think may violate guidelines, fill out the built-in reporting option by tapping on the three dots at the top right of the image, select "Report," select "It's inappropriate," select "Hate speech or symbols," and click "Submit."

Instagram has a global team that reviews reports and works as quickly as possible to remove content that does not meet its guidelines.

Instagram may work with law enforcement when it believes there is a risk of physical harm or threat to public safety.

StandWithUs
SAIDOFF
Legal

CCA
CENTER FOR COMBATING
ANTISEMITISM
A Division of StandWithUs

info@standwithus.com
310.836.6140

StandWithUs, PO Box 341069, Los Angeles, CA 90034-1069

EVELYN &
DR. SHMUEL
KATZ

Help Us Send This Publication To Others