

# SOCIAL MEDIA POLICIES:

## Mis/Disinformation, Threats, and Harassment

Social media has become a key vector in the spread of election-related disinformation and threats. The number of platforms is growing, and each has its own policies concerning election mis- and disinformation, threats, harassment, and doxing. With the help of the Institute for Strategic Dialogue, we have compiled the policies of some of the most used platforms, including Gab, Meta (Facebook, Instagram, and WhatsApp), Reddit, Telegram, TikTok, Truth Social, Twitter, and YouTube. A detailed description of each policy along with links to them on each platform's site are below.

### Gab

#### *Mis/Disinformation*

Gab does not have a stated policy related to elections or to mis- or disinformation.

#### *Ad Requirements*

Gab does not have a stated policy on political ads.

#### *Threats/Harassment*

The platform's [Terms of Service](#) forbids illegal content and "unlawful threats." Additionally, Gab's policy states users "agree not to use" Gab to engage in conduct which, as determined by the company, "may result in the physical harm or offline harassment of the Company, individual users of the Website or any other person (e.g. 'doxing'), or expose them to liability."

#### *Doxing*

Gab's [Terms of Service](#) prohibits users from sharing information that could result in physical harm, offline harassment or expose them to liability (i.e. sharing personal information).

## *Mis/Disinformation*

Meta, which owns Facebook, Instagram, and WhatsApp, says its election-related policies focus on “preventing interference, fighting misinformation, and increasing transparency.”

Prohibited content includes:

1. Posts containing false or misleading information about election dates, locations, times, or eligibility;
2. Posts that feature false or misleading information about the methods of voting or whether a vote will be counted;
3. Misleading posts about whether a candidate is running; and
4. Coordinated calls for voter/elections interference.

Prohibited content may be removed, designated for “reduced” distribution into users’ newsfeeds, or labeled with additional information, likely by third-party [fact checkers](#). Meta [policies](#) on misinformation indicate that content that produces risk of immediate physical harm or interference in electoral processes is more likely to be removed altogether, whereas content containing general misinformation and disinformation is more likely to be limited in its distribution across the platform and subject to fact-checking labels. Posts that violate Meta policies may sometimes remain accessible if the company determines “the public interest outweighs the risk of harm,” though Meta states that posts promoting violence or suppressing voting are not considered for this exemption.

## Exemptions for Politicians

Meta exempts posts and ads by [politicians](#) from its third-party fact-checking program. This policy covers statements, photos, videos, and other content labeled as part of a politician’s platform or campaign. Meta defines “politician” as anyone running for or currently holding elected office, cabinet appointees, political parties, and political party leaders.

Meta policies prohibit politicians from posting misinformation on where, when, or how to vote and content inciting violence, but Meta does not subject direct posts from politicians to the same fact-checking process applied to other content. However, if a politician shares content produced by others that has been found to be false by a third-party fact-checker (e.g., a link to an article containing false claims), Meta will label it with a fact-check and limit its spread. Meta also claims it will “label content from politicians that might violate our policies,” but may allow the offending content to remain on the platform “for public awareness.”

Former candidates and former officials are [subject](#) to regular fact-checking policies.

If a political candidate claims to have won an election before it is officially called by reputable media outlets, Meta will [add](#) a label to their post to state that vote counting is ongoing and no winner has emerged.

If a candidate contests the declared winner of an election, a label will be [added](#) with the name of the winning candidate. Both labels include links to the company’s “Voting Information Center,” which includes detailed information about elections.

## *Ad Requirements*

Advertisers seeking to publish ads on Meta platforms that address “social issues, elections or politics” must be [authorized](#) before doing so. Such ads on Facebook and [Instagram](#) must have a “Paid for by” disclaimer. Meta’s [ad library](#) makes available information about current and past ads, including their received audiences and how much money was spent promoting them.

Meta will forbid the creation of new political, electoral, and social issue ads during the last week of the election in 2022. Old ads can be re-run, but not edited. The ban on new ads will lift the day after Election Day.

## *Threats/Harassment*

Ads or posts that incite violence are not permitted, from any type of user — e.g., individual, politician, organization, or any other. Related to elections, Meta explicitly forbids:

- Threats targeting election officials or are related “to voting, voter registration, or the outcome of an election”; and
- Calls to bring weapons to election-related locations.

## *Doxing*

Meta forbids content that “shares, offers or solicits personally identifiable information or other private information that could lead to physical or financial harm, including financial, residential, and medical information, as well as private information obtained from illegal sources” such as hacking.

Meta [implemented](#) a recommendation by its Oversight Board which removed its “publicly available” exception for private residential information from both Facebook and Instagram, however its policies related to sharing photos of the exteriors of private homes have exceptions for when the property “is the focus of the news story, except when shared in the context of organizing protests against the resident.”

WhatsApp does not have a user-based doxing policy.

## Facebook-Specific Policies

Facebook “demotes” Group content from members who violate Community Guidelines on the platform. This includes restricting their ability to like comment, add new members to a Group or create a new Group. Facebook has also reduced the distribution of political content on users Feed and provides users the option to switch off social issue, electoral, or political ads.

## Instagram-Specific Policies

Instagram’s [policy](#) states it removes “misinformation that could cause physical harm or suppress voting.” However, merely false or misleading content, “including claims that have been found false by independent fact-checkers or certain expert organizations,” is allowed on the platform, though it is not included in algorithm-driven “recommended” content that appears to users via the platform’s “Explore” feed, “Accounts You May Like” suggestions, or the “Reels” tab.

# Reddit

## Mis/Disinformation

Content moderation is handled at Reddit at a site-wide level, community level, and user level. Reddit’s [content policy](#) does not specifically address elections. Communities are given deference to write and enforce their own rules, meaning that policies on election-related misinformation and disinformation can vary wildly. A Reddit spokesperson [told](#) Consumer Reports in 2020 that misinformation on voting was banned, but that it wasn’t included in published policies.

Reddit [enforces](#) its policies through:

- Warnings to cease violating behavior;
- Account suspension;
- Restrictions added to Reddit communities, such as Not Safe for Work (NSFW) tags or [Quarantining](#);
- Content deletion; and
- Community bans.

## Ad Requirements

Reddit’s ad [policy](#) bans “deceptive, untrue, or misleading” advertisements. In addition, [political](#) advertisers must allow users to comment on their ads for at least 24 hours after posting and include clear “paid for by” disclaimers. Reddit forbids political ads from outside of the United States and only accepts ads for campaigns and issues at the federal level. Ads that discourage voting or registering to vote are [not allowed](#).

In 2020, Reddit launched an official [subreddit](#) to list all political ads running on the platform. Posts in this subreddit included information such as the name of the organization, the amount they spent on an ad, and which subreddits were targeted with the ad.

## *Threats/Harassment*

Reddit bans users and communities that “incite violence or that promote hate,” and does not allow confidential information to be posted. There is also a ban on [impersonating](#) another person, including through the use of deepfakes.

## *Doxing*

Reddit’s Content Policy prohibits the “instigation of harassment,” including revealing someone else’s personal confidential information, including “links to public Facebook pages and screenshots of Facebook pages with the names still legible.” Exceptions to the rule can include posting professional links of public figures, such as the CEO of a company, if the post does not encourage harassment or “obvious vigilantism.” Users who violate these policies could be banned from the platform.

# Telegram

## *Mis/Disinformation*

[Telegram](#) does not have a stated policy related to elections or to mis- or disinformation.

## *Ad Requirements*

Telegram does not have a stated policy on political ads.

## *Threats/Harassment*

Telegram’s [Terms of Service](#) prohibits calls to violence. The platform does engage in periodic moderation of hateful and violent channels—including in the aftermath of the Jan 6 Capitol attack—however, moderation is applied irregularly and inconsistently.

## *Doxing*

Telegram does not have a user-based doxing policy.

## *Mis/Disinformation*

TikTok's [Election Integrity](#) policies do not allow for content that spreads distrust in public institutions, claims votes will not be counted, misrepresents election dates or locations, or attempts to suppress votes. Unverified claims, such as early declarations of victory or unconfirmed stories about polling locations, are [made](#) ineligible for recommendation to viewers. Accounts that are entirely dedicated to spreading election related mis- or disinformation are [banned](#).

In 2022, TikTok [announced](#) it will launch an “Election Center” to connect users who engage with election-related content to “authoritative information and sources in more than 45 languages.” The information will include how and where to vote, who and what is on the ballot. As elections are conducted, TikTok will display results from the Associated Press. The company will work with fact-checking groups like PolitiFact.

There are a variety of policy enforcement mechanisms TikTok uses, including:

- Removing content;
- Redirecting search results;
- Restricting discoverability, for example, by making content ineligible for the “For You” page;
- Blocking accounts from livestreaming;
- Removing an account; and
- Banning a device from the platform, in the case of serious violations of community guidelines.

## *Ad Requirements*

TikTok does not allow paid political ads and has vowed to work to tighten an existing loophole in its policies that some content creators used to receive payment in exchange for posting political messages online.

## *Threats/Harassment*

TikTok's [Election Integrity](#) policy does not allow:

- “Newsworthy content that incites people to violence”;
- Livestreams that seek to “incite violence or promote hateful ideologies, conspiracies, or disinformation”; and
- Redirect search results or hashtags that incite violence or are associated with hate speech.

In addition to its election focused policy, TikTok's Community Guidelines prohibit users from using TikTok “to threaten or incite violence, or to promote violent extremist organizations, individuals, or acts.”

Regarding harassment, TikTok [does not allow](#):

- Content that insults another individual, or disparages an individual on the basis of attributes such as intellect, appearance, personality traits, or hygiene
- Content that encourages coordinated harassment
- Content that disparages victims of violent tragedies
- Content that uses TikTok interactive features (e.g., duet) to degrade others
- Content that depicts willful harm or intimidation, such as cyberstalking or trolling
- Content that wishes death, serious disease, or other serious harm on an individual.

### *Doxing*

TikTok's [Community Guidelines](#) state the company forbids threats of hacking or doxing users with the intention to harass or blackmail users that can cause "serious emotional distress and other offline harm." The platform defines doxing as the "act of collecting and publishing personal data or personally identifiable information (PII) for malicious purposes."

## Truth Social

### *Mis/Disinformation*

Truth Social's [moderation](#) page states the company moderates the platform to prevent "illegal and other prohibited content," but that they "cherish free expression." While the [Terms of Service](#) do not contain any references to elections or misinformation, they state users "may not post any false, unlawful, threatening, defamatory, harassing or misleading statements" or post content that is "false, inaccurate, or misleading." It is unknown how "false," "inaccurate," or "misleading" are defined.

### *Ad Requirements*

Truth Social does not have a stated political ads policy.

### *Threats/Harassment*

Truth Social's [Terms of Service](#) prohibits threats and harassment on the platforms including posts or contributions that:

- Depict violence, threats of violence or criminal activity;
- Advocate or incite, encourage, or threaten physical harm against another;
- Are false, unlawful, threatening, defamatory, harassing or misleading statements;
- Use any information from the Truth Social in order to harass, abuse, or harm another person; and
- Are obscene, lewd, lascivious, filthy, violent, harassing, libelous, slanderous, or otherwise objectionable.

## *Doxing*

Truth Social's [Terms of Service](#) page does not explicitly discuss policies regarding doxing. However, it does prohibit the "use of any information obtained from the Service in order to harass, abuse, or harm another person." The platform does not specify what would fall under this categorization.

# Twitter

## *Mis/Disinformation*

Twitter's [Civic Integrity Policy](#) prohibits "manipulating or interfering in elections or other civic processes." The company's definition of civic processes includes political elections, censuses, and referenda or ballot initiatives. Violations of this policy include:

- Content that suppresses participation or misleads users about how to participate in civic processes;
- Content that misleads people about the outcome of an election, or undermines trust in electoral processes; and
- Accounts that pretend to be a political candidate, political party, electoral authority, or government entity.

Twitter has made [changes](#) to their mis/disinformation policy approaching the 2022 midterm election, expanding their Civic Integrity Policy and adding features to the platform.

Twitter will be incorporating "prebunks," in English, Spanish, and other languages, to proactively address topics that could be subjects of misinformation. Prebunks will be placed on timelines and search results when people type related terms, phrases, or hashtags.

Twitter will be providing state-specific pages that display real-time election information of state election officials and local news outlets/journalists. The same will be provided on a national level.

Twitter will be adding candidate account labels to help individuals identify who is running for office. Labels will be applied to those running for US Senate, US House of Representatives, or Governor. This label will appear on their profile and on all their tweets.

In the absence of other violations, Twitter does not classify content that includes inaccurate statements, hyper-partisan content, or parody accounts that discuss elections as misinformation.

Twitter addresses violations of this policy through:

- Tweet deletion;
- Failing to recommend or amplify content with misleading information;
- Redesigning the misleading information label to increase click through rates for debunking content;
- Profile modification, if the violating content is within profile information;
- Labeling tweets to warn it is misleading, or providing links with additional context;
- Prompting users when Twitter determines a tweet has a false or misleading claim;
- Turning off the ability to retweet, like, or reply to a tweet;
- And locking or suspending the account

## *Ad Requirements*

Twitter does not allow [paid ads](#) featuring political content that references political candidates, parties, elections, or legislation. They do not allow any promoted content from politicians or political parties. In the U.S., Political Action Committees (PACs) and 501(c)(4)s are not allowed to advertise on Twitter.

## *Threats/Harassment*

Twitter has a [Violent Threats Policy](#) that prohibits various threatening and harassing activity on the platform. The company states users cannot state intent to inflict violence on a person or group of people. Twitter says examples of intent include “I will”, “I’m going to”, or “I plan to”, as well as conditional statements like “If you do X, I will”. Violations of this policy include, but are not limited to:

- Threatening to kill someone;
- Threatening to sexually assault someone;
- Threatening to seriously hurt someone and/or commit an other violent act that could lead to someone’s death or serious physical injury; and
- Asking for or offering a financial reward in exchange for inflicting violence on a specific person or group of people.

Twitter does not ban all violent rhetoric or threatening content. According to the company, “statements that express a wish or hope that someone experiences physical harm, making vague or indirect threats, or threatening actions that are unlikely to cause serious or lasting injury are not actionable under this policy, but may be reviewed and actioned under those policies.”

Twitter also has a [Hateful Conduct Policy](#) and [Abusive Behavior Policy](#). According to the Hateful Conduct Policy, users “may not 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. We also do not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories.”

In addition to content and behavior mentioned in the company's [Violent Threats Policy](#), directly related to elections, Twitter [prohibits](#):

- Threats regarding voting locations or other key places or events; and
- Inciting unlawful conduct to prevent the procedural or practical implementation of election results.

## *Doxing*

In April 2022, Twitter [updated](#) its private information and media policy to prohibit users from publishing or posting other individual's private information without express permission. The company also forbids threats of exposing private information or encouraging others to do so.

Twitter claims that when a user violates this policy for the first time, the account is required to remove the content and will be temporarily locked. If the violation occurs a second time, the account will be permanently suspended.

# YouTube

## *Mis/Disinformation*

YouTube has a specific elections misinformation [policy](#) that states content that suppresses or prevents voting, undermines election integrity, and otherwise spreads election misinformation is in violation of the platform's community standards and may be removed. The platform's Terms of Service defines election misinformation-related content as "misleading or deceptive," with "serious risk of egregious harm ... [including] technically manipulated content and content interfering with democratic processes." Clips taken out of context or deceptively edited [do not](#) fall under the definition of "manipulated content."

The [policy](#) elaborates on specific forms of elections misinformation, including but not limited to:

- Content that promotes voter suppression or is misleading regarding whether specific candidates are eligible and running;
- Content that encourages individuals to "interfere with democratic processes" which may include "obstructing or interrupting voting procedures";
- Distributing information obtained through hacking, "the disclosure of which may interfere with democratic processes"; and
- Content that undermines the integrity of elections, including claims that fraud or errors are widespread "in certain past certified national elections, including any U.S. previous presidential elections."

YouTube's election misinformation policy also covers external links shared in content posted on YouTube, including URLs and verbal directions to another site. The policy also does not allow users to post previously removed content or share content from terminated or restricted users. If a user's content violates this policy, YouTube states it removes the content and sends an email to the user. On the first violation, the account receives a warning; however, it will receive a strike following each subsequent incident. After three strikes, the channel is terminated.

## Ad Requirements

[Political ads](#) on YouTube must adhere to Google's [ad policies](#), which requires political organizations to go through [verification](#) before running ads on Google platforms. Monetized channels are subject to [eligibility requirements](#).

## Threats/Harassment

YouTube has a [Harmful or Dangerous Content Policy](#), [Hate Speech Policy](#), and [Harassment and Cyberbullying Policy](#). Though not specific to elections, YouTube prohibits:

- Inciting others to commit violent acts against individuals or a defined group of people;
- Promoting violence or hatred against individuals or groups based on age, caste, disability, ethnicity, gender identity and expression, nationality, race, immigration status, religion, sex/gender, sexual orientation, victims of a major violent event and their kin, or veteran status;
- Content that threatens individuals;
- Content that targets an individual with prolonged or malicious insults based on intrinsic attributes, including [protected group status](#) or physical traits.

YouTube does have exceptions for harassment, if “the primary purpose is educational, documentary, scientific, or artistic in nature, we may allow content that includes harassment.” For example, “Content featuring debates or discussions of topical issues concerning individuals who have positions of power, like high-profile government officials or CEOs of major multinational corporations.”

## Doxing

YouTube's [Harassment and Cyberbullying Policies](#) state that users are not allowed to post content that reveals someone's personally identifiable information (PII). Additionally, YouTube explicitly states that abusive behavior, such as doxing, is banned from the site. Exceptions to this include posting widely available information such as the phone number of a business.