UPDATED SOCIAL MEDIA POLICIES RELATED TO ELECTIONS IN THE US

August 2023

Contents

Meta

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.

Ahead of the 2024 presidential election, Meta reinstated Donald Trump and Robert F. Kennedy Jr. on Instagram and Facebook. Trump was reinstated after his 2-year suspension for incitement ahead of January 6 expired. Kennedy, who was suspended in August 2022 for COVID-19 misinformation, was reinstated as he is now an active candidate for the 2024 presidential election. It is unclear whether Meta plans to re-suspend him if he drops out of or loses the race.

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.

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:
1. Threats targeting election officials or are related "to voting, voter registration, or the outcome of an election"; and
2. 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.

Manipulated Media

Meta [removes](#) misleading manipulated media if it "has been edited or synthesized […] in ways that aren't apparent to an average person and would likely mislead someone into thinking that a subject of the video said words that they did not actually say" and if the media is a product of AI or machine learning that "merges, replaces or superimposes content onto a video, making it appear to be authentic." This policy does not apply to content that is parody or satire or a video that has been edited to omit or "change the order of words." If a post containing misleading manipulated media does not meet the standards for removal, they are still eligible for review by Meta's third-party fact-checkers.

Meta has [dodged questions](#) about whether politicians can post AI-generated media or manipulated media without warnings or removals. At the time of writing, it would seem Meta will only fact-check manipulated media posted by politicians, rather than applying their larger manipulated media policy to such content.

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." As of August 2023, this currently applies to the 2021 German federal election and 2014, 2018, and 2022 Brazilian Presidential elections. The US 2020 federal election has been removed from the list. YouTube stated that "while removing [content with misleading claims about the 2020 election] does curb some misinformation, it could also have the unintended effect of curtailing political speech without meaningfully reducing the risk of violence or other real-world harm."

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 identify 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.

Manipulated Media

YouTube's policies regarding manipulated content fall under the platform's [Misinformation Policy](#). The platform defines manipulated content as "content that has been technically manipulated or doctored in a way that misleads users (beyond clips taken out of context) and may pose a serious risk of egregious harm." Some examples that YouTube provides include:

1. Inaccurately translated video subtitles that inflame geopolitical tensions creating serious risk of egregious harm.
2. Videos that have been technically manipulated (beyond clips taken out of context) to make it appear that a government official is dead.
3. Video content that has been technically manipulated (beyond clips taken out of context) to fabricate events where there's a serious risk of egregious harm.

TikTok

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 launched an "Election Center" to connect users who engage with election-related content to "authoritative information and sources in more than 45 languages." The information includes how and where to vote, who and what is on the ballot. As elections were conducted in the US, TikTok displayed results from the Associated Press. The company also worked 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.

Manipulated Media

In April 2023, TikTok updated its Integrity and Authenticity Policy to include synthetic and manipulated media. The Synthetic and Manipulated Media Policy states that "synthetic or manipulated media that shows realistic scenes must be clearly disclosed" and that the platform does not allow synthetic media that "contains the likeness of any real private figures." The platform is more lax for public figures but does not allow "synthetic media of public figures if the content is used for endorsements or violates any other policy" (e.g., hate speech, harassment, sexual exploitation).

TikTok's definition of synthetic media is "content created or modified by AI technology that includes highly realistic digitally-created content of real people, such as a video of a real person speaking words that have been modified or changed." The platform's definition of public figures is "adults (18 years and older) with a significant public role, such as a government official, politician, business leader, and celebrity."

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."

Telegram

<mark>Policies remain unchanged – section can remain as is.</mark>

Manipulated Media
Telegram does not have a Manipulated Media policy.

X (formerly Twitter)

Mis/Disinformation
X'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, dissuades, 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 – unless the accounts are parody, commentary, or fan accounts. Accounts that fall under this latter category must distinguish themselves in their account name and bio, according to the [Misleading and Deceptive Identities Policy](#).

X addresses violations of this policy through:
- Post deletion;
- Requiring the user to remove the content;
- Temporarily locking the user out of their account;
- Profile modifications, if the violating content is within the profile information;
- Labeling Posts to warn they are misleading, or providing links with additional context;
- Turning off the ability to Repost, like, or reply to a Post;
- And locking or suspending the account.

On April 1, X removed blue checkmark badges from verified accounts that were not signed up for paid X Blue subscriptions. Several politicians lost their verification and Musk's new system gave way for a new spread of fake accounts. Musk then implemented [grey checkmarks](#) for government/multilateral organizations and officials. The [features](#) that were put in place for 2022 midterms (including "[prebunks](#)" and state-specific event hubs) no longer exist, and it is unclear if they will be brought back before 2024 elections.

Additionally, Musk introduced [Community Notes](#), which allows X users to add context to potentially misleading Posts (formerly Tweets).

Ad Requirements

In 2023, X lifted the long-time ban on paid ads featuring political content that references political candidates, parties, elections, or legislation in the US. Political campaigning ads can now be promoted through the following ad formats: Promoted Ads, Follower Ads, X Takeover, X Live, DPA, Collection Ads, and X Ad Features. However, political content cannot include:

- False or misleading information about how to participate in an election;
- False or misleading information intended to intimidate or dissuade people from participating in an election;
- False or misleading information intended to undermine public confidence in an election.

Additionally, X states that "advertisers must comply with any applicable laws regarding disclosure and content requirements. Such compliance is the sole responsibility of the advertiser."

X also created a Political Ads Disclosure page to allow anyone to request information on US Political Campaigning ads via Google Form.

Threats/Harassment

X has a Violent Speech Policy that prohibits threats, incitement, glorification, or expressions of desire for violence or harm. Violations of this policy include, but are not limited to:

- Threatening to kill someone;
- Threatening to sexually assault someone;
- Threatening to damage civilian homes, infrastructure, and shelters;
- Wishing, hoping, or expressing desire for harm;
- Inciting, promoting, or encouraging others to commit acts of violence or harm;
- Glorifying, praising, or celebrating acts of violence where harm occurred.

X does not ban all violent rhetoric or threatening content. According to the company, expressions of violent speech when "there is no clear abusive or violent context" or is a "figure of speech, satire, or artistic expression" are not violations.

X also has a Hateful Conduct Policy and Abuse and Harassment 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." X also acts against "reports of accounts targeting an individual or group of people."

Doxing

X's Private Information and Media Policy prohibits users from publishing or posting another individual's private information without express permission. The company also forbids threats of exposing private information or encouraging others to do so.

X 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. Accounts dedicated to sharing someone's "live location" are automatically suspended.

In December 2022, X updated its Private Information and Media Policy to prohibit users from sharing "live information" about an individual, or their "real-time and/or same-day information where there is potential that the individual could still be at the named location." This update came after X [banned](#) @elonjet, an account that used publicly available flight data to track Elon Musk's private jet.

Manipulated Media

X's Synthetic and Manipulated Media Policy prohibits "synthetic, manipulated, or out-of-context media that may deceive or confuse people and lead to harm ('misleading media')." The company defines misleading media as images, videos, audios, gifs, and URLs hosting relevant content that:
- Includes media that is significantly and deceptively altered, manipulated, or fabricated, or;
- Includes media that is shared in a deceptive manner or with false context, and;
- Includes media likely to result in widespread confusion on public issues, impact public safety, or cause serious harm.

When considering whether to label or remove content containing manipulated media, X considers three questions:
1. Is the content significantly and deceptively altered, manipulated, or fabricated?
2. Is the content shared in a deceptive manner or with false context?
3. Is the content likely to result in widespread confusion on public issues, impact public safety, or cause serious harm?

Manipulated media in the form of memes, satire, animations, illustrations, cartoons, commentary, and counter speech – minus some exceptions – do not violate this policy.

If content is found to violate X's Synthetic and Manipulated Media Policy, the platform can:
- Delete the post;
- Apply a label and/or warning message to the post;
- Reduce the visibility of the post on the platform;
- Provide a link to additional explanations or clarifications;
- Temporarily reduce the visibility of the account or lock or suspend the account.

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.

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.

Manipulated Media

Reddit bans content that impersonates individuals or entities in a "misleading or deceptive manner." This includes "deepfakes or other manipulated content presented to mislead, or falsely attributed to an individual or entity."

Gab

Policies remain unchanged – section can remain as is.

Manipulated Media

Gab does not have a Manipulated Media policy.

Truth Social

Policies remain unchanged – section can remain as is.

Manipulated Media

Truth Social does not have a Manipulated Media Policy. It is worth to note that in late May 2023, Trump [posted](#) a fan-made fake video seemingly using AI-generated voices of Adolf Hitler, George Soros, WEF chairman Klaus Schwab, former VP Dick Cheney, and the FBI. The video was attacking Ron DeSantis' Twitter Spaces event with Elon Musk, where DeSantis was slated to announce his candidacy for 2024. This suggests there might not be strict policies on deepfakes or altered content.