



Evaluating Platform Election-Related Speech Policies

Platform Policy Details

This document has been updated as of October 28, 2020.

In practice, policies covering election-related content often intersect with other platform policies, such as those addressing misleading information, threats of violence, or health misinformation. However, this overview of platform policies is mostly restricted to those that specifically refer to elections, voting, and election results in order to shed light on how platforms are preparing for the upcoming election. Again, it is important to note that the comprehensiveness of these policies isn't a guarantee of their effectiveness or of their consistent enforcement, and information other than content is also taken into account, such as the gravity of the infringement, the nature of the account posting the content, and prior infringements made by the account posting the content. Nevertheless, these are the policies as they appear in each platform's community guidelines.

This document is intended to contribute to our understanding of the platforms' policies in three ways: it offers a summary of each platform's policies, provides a detailed breakdown of how these policies fall into separate election-related categories, and explains how we arrived at each rating.

Facebook

Facebook's policy remains the most detailed in terms of election-related misinformation and disinformation content; this specificity includes [statements from Facebook's leadership](#) on how consequences of infringement will be implemented, including labeling content the platform chooses to leave up under its "newsworthiness" policy with a [prompt](#) that the content the user is sharing may violate Facebook policies. However, Facebook also has some broad clauses that make it difficult to determine what specific content would fall under the policy. For example, phrasing such as "an individual's ability to participate in an election" reserves more leeway for the platform to act or not act on borderline content.

According to NBC, Facebook has [said](#) it will apply a label to official information about the 2020 election on all posts that mention voting or elections by presidential and congressional candidates as well as elected officials, but will not fact-check the claims made in the post. A notable application of this policy is Facebook's decision to label President Trump's [post](#) on July 30, 2020, that mail-in voting will make the election the most "Inaccurate & Fraudulent" one yet.

On [September 3, 2020](#), Facebook introduced new election-related policies and provided more information about which posts it will remove or label. Facebook expanded the application of its voter suppression policies; it will [remove posts](#) that explicitly and implicitly aim to deter people from voting. The platform also updated its advertising policies, stating it will not accept new political ads in the week before the election. Advertisers will be able to continue running ads they started running before the final week. Facebook noted that it will not update its election-related policies again prior to the November elections, but it released an update to its policies on October 7, 2020.

Facebook's [policy updates on October 7](#) aim to reduce mis- and disinformation on election day. Facebook will remove calls for people to engage in poll watching when the posts suggest that the goal is voter intimidation. After polls close, Facebook will post labels on election-related content noting that votes are still

being counted. Labels will update when major media outlets announce a projected winner. Finally, starting November 3, Facebook will temporarily suspend all ads related to politics or social issues. It is unclear when these will be allowed again, but, given that this policy is meant to reduce election confusion, this may depend on how long it takes to declare a definite winner.

Table key: A platform’s policies may apply to multiple categories; in these cases phrases applicable to the category discussed are in bold. An asterisk indicates our justification for including a phrase in that category. Red text reflects Facebook’s policy updates since August 2020. A date of either [Sept. 03] or [Oct. 7] next to a clause reflects the date of the policy update.

F A C E B O O K	Categories	Policy Language	Rating and Explanation
	Procedural Interference	Do Not Post: <ul style="list-style-type: none"> • “Misrepresentation of the dates, locations, and times, and methods for voting or voter registration or census participation.” • “Misrepresentation of who can vote, qualifications for voting, whether a vote will be counted, and what information and/or materials must be provided in order to vote.” • “Misrepresentation of whether a candidate is running or not.” • “Calls for coordinated interference that would affect an individual’s ability to participate in the census or an election.” <i>(*Depending on the content’s exact wording, this statement could fall under the Participation Interference category.)</i> • “Other misrepresentations related to voting in an official election or census participation may be subject to false news standards, as referenced in section 20” (now section 21). 	Comprehensive. Facebook’s election interference policy is “Comprehensive” in addressing procedural interference because it classifies content that prevents people from engaging in the election process as a type of misinformation prohibited on its platform. Specific language such as “dates, locations, and times, and methods” add to the concrete nature of this category.
	Participation Interference	Do Not Post: <ul style="list-style-type: none"> • “Any content containing statements of intent, calls for action, conditional or aspirational statements, or advocating for high— or mid—severity violence due to voting, voter registration, or the administration of outcome of an election.” [Sept. 03] • “Content stating that census or voting participation may or will result in law 	Comprehensive. Facebook’s community standards related to participation interference is labeled “Comprehensive” because they classify content that would make a voter feel threatened or deterred from participating in the

		<p>enforcement consequences (e.g., arrest, deportation, imprisonment).”</p> <ul style="list-style-type: none"> • “Content claiming that the U.S. Immigration and Customs Enforcement (ICE) is at a voting location.” [Sept. 03] • “Calls for coordinated interference that would affect an individual’s ability to participate in an election.” (<i>*Depending on the content’s exact wording, this statement could fall under the Procedural Interference category.</i>) • “Explicit claims that people will be infected by COVID (or another communicable disease) if they participate in the voting process.” [Sept. 03] • “Statements of intent or advocacy, calls to action, or aspirational or conditional statements to bring weapons to locations, including but not limited to places of worship, educational facilities, or polling places, or locations used to count votes or administer an election* (or encouraging others to do the same).” • “For the following content, we may require more information and/or context in order to enforce <ul style="list-style-type: none"> ○ Threats against election officials.” [Sept. 03] 	<p>electoral process. Though not incorporated yet into policy, Facebook has provided concrete examples of what constitutes voter intimidation.</p>
--	--	--	--

F A C E B O O K	Categories	Policy Language	Rating and Explanation
	<p>Fraud</p>	<p>Do Not Post</p> <ul style="list-style-type: none"> • “Misrepresentation of who can vote, qualifications for voting, whether a vote will be counted, and what information and/or materials must be provided in order to vote.” • “Offers to buy or sell votes with cash or gifts.” • “Statements that advocate, provide instructions or show explicit intent to illegally participate in a voting or census process.” 	<p>Comprehensive. Facebook includes a clear policy on election fraud, classifying content that incites action to illegally participate in the election process and content misrepresenting whether a vote will be counted.</p>
	<p>Delegitimization of Election Results</p>	<p><u>Scenario 1 (corresponding policy):</u></p> <ul style="list-style-type: none"> • “Other misrepresentations related to voting in an official election or census participation may be subject to false news standards, as referenced in section 20” (now section 21). • “We will attach an informational label to content that seeks to delegitimize the outcome of the election or discuss the legitimacy of voting methods, for example, by claiming that lawful methods of voting will lead to fraud. This label will provide basic authoritative information about the integrity of the election and voting methods.” [Sept. 03] 	<p><u>Scenario 1:</u> Non-Comprehensive. It is unclear if Facebook’s policy would apply to the broad claim that the “election is rigged.” Policy action might differ depending on the context of the post, e.g., the speaker, additional content, or virality of the post.</p>

Scenario 2:

- “Misrepresentation of who can vote, qualifications for voting, **whether a vote will be counted**, and what information and/or materials must be provided in order to vote.”
- Facebook will remove implicit misrepresentations about voting that may “mislead you about what you need to do to get a ballot.” [Sept. 03]
- “We will attach an informational label to content that seeks to delegitimize the outcome of the election or discuss the legitimacy of voting methods, for example, by claiming that lawful methods of voting will lead to fraud. This label will provide basic authoritative information about the integrity of the election and voting methods.” [Sept. 03]

Scenario 3:

- “Video that has been edited or synthesized, beyond adjustments for clarity or quality, in ways that are not apparent to an average person, and would likely mislead an average person to believe that a subject of the video said words that they did not say AND is the product of artificial intelligence or machine learning, including deep learning techniques (e.g., a technical deepfake), that merges, combines, replaces, and/or superimposes content onto a video, creating a video that appears authentic.”
- “Other misrepresentations related to voting in an official election or census participation may be subject to false news standards, as referenced in section 20.” (Now section 21).
- “We will attach an informational label to content that seeks to delegitimize the outcome of the election or discuss the legitimacy of voting methods, for example, by claiming that lawful methods of voting will lead to fraud. This label will provide basic authoritative information about the integrity of the election and voting methods.” [Sept. 03]

Scenario 4:

- “Importantly, if any candidate or campaign tries to declare victory before the results are in, we’ll add a label to their post educating that official results are not yet in and directing people to the official results.” [Sept. 03]
- “Other misrepresentations related to voting in an official election or census participation may be subject to false news standards, as referenced in [section 20](#)” (now section 21).

Scenario 2: Updated from Non-Comprehensive to Comprehensive. It is now clear that under the updated policies introduced on September 3, 2020, Facebook will take action against posts that implicitly spread false information about voting or aim to delegitimize the election.

Scenario 3: Updated from Non-Comprehensive to Comprehensive. It is clear that under the updated policies introduced on September 3, 2020, Facebook will take action against posts that aim to delegitimize the election.

Scenario 4: Comprehensive. Under Facebook’s [new policies](#) updated September 3, 2020, the platform explicitly states it will take action on statements by candidates or campaigns attempting to declare victory before election results are officially called.

<p>Physical Confrontations Outside Polling Places</p>	<p><u>Scenario 1 (corresponding policy):</u></p> <ul style="list-style-type: none"> Facebook will “remove calls for people to engage in poll watching when those calls use militarized language or suggest that the goal is to intimidate, exert control, or display power over election officials or voters.” [Oct. 7] <p><u>Scenario 2:</u></p> <ul style="list-style-type: none"> Facebook will “remove calls for people to engage in poll watching when those calls use militarized language or suggest that the goal is to intimidate, exert control, or display power over election officials or voters.” [Oct. 7] <p><u>Scenario 3:</u></p> <ul style="list-style-type: none"> “Statements of intent or advocacy, calls to action, or aspirational or conditional statements to bring weapons 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).” [Oct. 7] “Any content containing statements of intent, calls for action, conditional or aspirational statements, or advocating for high-or mid-severity violence due to voting, voter registration, or the administration of outcome of an election.” [Sept. 03] 	<p><u>Scenario 1:</u> Comprehensive. Facebook’s policy specifically addresses calls to action for people to engage in poll watching. The policy addresses posts that use militarized language, and also those where the goal is to “exert control or display power over election officials or voters,” which may apply in Scenarios 1 and 2.</p> <p><u>Scenario 2:</u> Comprehensive. Facebook’s policy specifically addresses calls to action for people to engage in poll watching. The policy addresses posts that use militarized language, and also those where the goal is to “exert control or display power over election officials or voters,” which may apply in Scenarios 1 and 2.</p> <p><u>Scenario 3:</u> Comprehensive. Facebook’s policy specifically addresses posts that contain calls to action to bring a weapon to a polling place.</p>
--	--	--

<p style="text-align: center;">F A C E B O O K</p>	<p>Consequences of Infringement</p>	<p>Facebook affirms it will remove content that violates its community standards. Additionally, the platform and third-party fact-checkers identify problematic content that doesn't explicitly violate its policies but gets reduced in the newsfeed, "such as misinformation and clickbait."</p> <ul style="list-style-type: none"> ● Facebook will extend its period of "partnering with state election authorities to identify and remove false claims about polling conditions" from 72 hours prior to the election, as it had announced on June 26, 2020, to beginning Sept. 3 and continuing through the election "until we have a clear result." ● Facebook will limit the number of chats that you can forward a message to at one time on Messenger. ● Facebook will remove posts "with claims that people will get Covid-19 if they take part in voting." They will provide a link to authoritative information about Covid-19 to posts that use the virus to discourage voting. ● "If any candidate or campaign tries to declare victory before the results are in, we'll add a label to their post educating them that official results are not yet in and directing people to the official results." ● "We will attach an informational label to content that seeks to delegitimize the outcome of the election or discuss the legitimacy of voting methods, for example, by claiming that lawful methods of voting will lead to fraud. This label will provide basic authoritative information about the integrity of the election and voting methods." ● Facebook will remove implicit misrepresentations about voting that may "mislead you about what you need to do to get a ballot." [Sept. 3] <p>"Other misrepresentations of electoral interference other than what has been specified may be subject to false news standards": Facebook will not remove but will instead significantly reduce such content's distribution by showing it lower in the News Feed.</p> <ul style="list-style-type: none"> ● Noteworthiness Exemption: Facebook labels content it leaves up as noteworthy. "We will treat speech from politicians as noteworthy content that should, as a general rule, be seen and heard. However, ... this will not apply to ads." There is no newsworthiness exemption to content that incites violence or suppresses voting. <p>Policies to introduce friction:</p> <ul style="list-style-type: none"> ● "Other misrepresentations of electoral interference other than what has been specified may be subject to false news standards": Facebook will not remove but will instead significantly reduce such content's distribution by showing it lower in the News Feed. ● Facebook will limit the number of chats a user can forward a message to at one time on Messenger. [Sept. 3]
--	--	---

Twitter

Twitter's policy, like Facebook's, is largely detailed and comprehensive. Most of Twitter's election-related policy comes from its [Civic Integrity policy](#). The lynchpin of this policy is the specificity of the content in question. A spokesperson for Twitter [told Vox](#) that, "Twitter does not take down 'broad, non-specific statements' about the integrity of elections or civic processes." This parameter is important when understanding the policy's use in practice. For example, Twitter took swift action on President Trump's [tweet](#) on May 26, 2020, alleging mail-in-ballots will be "fraudulent" and specifically mentioning that the "Governor of California is sending Ballots to millions of people," but it took no action on a similar President Trump [tweet](#) on July 30, 2020, that made more general claims about voter fraud.

On [September 10, 2020](#), Twitter updated its Civic Integrity Policy to expand the type of content that the platform will take action on. First, Twitter will now label or remove content that aims to delegitimize the election results. Second, Twitter introduced more explicit language about how it will respond to policy violations. Twitter [updated its policy actions](#) in each subsection below from: "You can't share false or misleading information..." to "We will label or remove false or misleading information." Twitter also clarified what happens when a tweet is labeled. For example, in some cases the tweet will have less visibility.

On [October 9, 2020](#), Twitter once again updated the steps it is taking ahead of the 2020 U.S. election. In this post, it shared more details about how it will determine the results of the election in order to effectively enforce its policy on premature claims of results. It also implemented further steps to insert "friction" into posts with potential for virality, including clicking through warnings, pushing users to Quote tweets instead of retweeting, eliminating "liked by" and "followed by" recommendations in timelines, and only surfacing trends in the "For You" tab with additional context. Lastly, it compiled and included a timeline of all election-related policies starting in September 2019 until now.

Red text reflects Twitter policy updates since August 2020. A date of either [Sept. 10] or [Oct. 9] next to a clause reflects the date of the policy update.

T W I T T E R	Categories	Policy Language	Rating and Explanation
	Procedural Interference	<p>Twitter’s policy prohibits false or misleading information about how to participate in an election. This includes but is not limited to:</p> <ul style="list-style-type: none"> • “Misleading information about procedures to participate in a civic process (for example, that you can vote by Tweet, text message, email, or phone call in jurisdictions where these are not a possibility).” • “Misleading information about requirements for participation, including identification or citizenship requirements” • “Misleading statements or information about the official, announced date or time of a civic process.” • “Misleading claims that polling places are closed, that polling has ended or other misleading information relating to votes not being counted.” • “Misleading claims about long lines, equipment problems, or other disruptions at voting locations during election periods.” • “False or misleading information that causes confusion about the laws and regulations of a civic process, or officials and institutions executing those civic processes.” [Sept. 10] • “Tweets meant to incite interference with the election process or with the implementation of election results, such as through violent action, will be subject to removal. This covers all Congressional races and the Presidential Election.” [Oct. 9] (*Depending on the content, this statement could fall under the Procedural Interference category.) 	<p>Comprehensive. Twitter’s Civic Integrity policy on procedural interference is labeled “Comprehensive” because of its specificity on the areas of procedural interference – requirements, date, time, manner – that provides more clarity on what type of content falls within the scope of this policy.</p>

T W I T T E R	Categories	Policy Language	Rating and Explanation
	Participation Interference	<p>Twitter’s policy prohibits false or misleading information intended to intimidate or dissuade people from participating in an election. This includes but is not limited to:</p> <ul style="list-style-type: none"> • “Misleading claims about police, or law enforcement activity related to voting in an election, polling places, or collecting census information.” • “Misleading claims about long lines, equipment problems, or other disruptions at voting locations during election periods.” • “Misleading claims about process, procedures, or techniques which could dissuade people from participating.” • “Threats regarding voting locations or other key places or events (note that our violent threats policy may also be relevant for threats not covered by this policy).” (<i>*This statement falls under the Participation Interference category because of the use of the term “threats.”</i>) • Twitter will remove “Tweets that encourage violence or call for people to interfere with election results or smooth operation of polling places.” • “Tweets meant to incite interference with the election process or with the implementation of election results, such as through violent action, will be subject to removal. This covers all Congressional races and the Presidential Election.” 	<p>Comprehensive. Twitter’s Civic Integrity policy on participation interference is labeled “Comprehensive” because the policy language is clear and detailed. For example, the policy language addresses specific types of speech the platform prohibits, giving weight to instances in which users post misleading content about critical processes, such as polling place closures, law enforcement function, and equipment viability.</p>
	Fraud	<ul style="list-style-type: none"> • “Misleading claims that polling places are closed, that polling has ended or other misleading information relating to votes not being counted.” • “Disputed claims that could undermine faith in the process itself, e.g. unverified information about election rigging, ballot tampering, vote tallying, or certification of election results.” [Sept. 10] 	<p>Non-Comprehensive. Twitter’s Civic Integrity policy on fraud does not address other ways in which users may illegally participate in the electoral process, such as claims of votes being cast illegally.</p>

T W I T T E R	Categories	Policy Language	Rating and Explanation
	Delegitimization of Election Results	<p><u>Scenario 1 (corresponding policy):</u></p> <ul style="list-style-type: none"> -Not in violation of Twitter’s policy: “...broad, non-specific statements about the integrity of elections or civic processes (such as unsubstantiated claims that an election is “rigged”)...” [Sept. 10] • “Disputed claims that could undermine faith in the process itself, e.g. unverified information about election rigging, ballot tampering, vote tallying, or certification of election results.” [Sept. 10] <p><u>Scenario 2:</u></p> <ul style="list-style-type: none"> • “Misleading claims that polling places are closed, that polling has ended or other misleading information relating to votes not being counted.” • “Disputed claims that could undermine faith in the process itself, e.g. unverified information about election rigging, ballot tampering, vote tallying, or certification of election results.” [Sept. 10] <p><u>Scenario 3:</u></p> <ul style="list-style-type: none"> • “Misleading claims that polling places are closed, that polling has ended or other misleading information relating to votes not being counted.” • Twitter’s synthetic and manipulated media policy: “In order for the content to be labeled or removed under this policy, we must have reason to believe that media, or the context in which media are presented, are significantly and deceptively altered or manipulated.” • “We also consider whether the context in which media are shared could result in confusion or misunderstanding or suggests a deliberate intent to deceive people about the nature or origin of the content, for example by falsely claiming that it depicts reality.” • “Disputed claims that could undermine faith in the process itself, e.g. unverified information about election rigging, ballot tampering, vote tallying, or certification of election results.” [Sept. 10] <p><u>Scenario 4:</u></p> <ul style="list-style-type: none"> • “Misleading claims about the results or outcome of a civic process which calls for or could lead to interference with the implementation of the results of the process, e.g. claiming victory before election 	<p><u>Scenario 1:</u> Comprehensive. Twitter’s policy is comprehensive in Scenario 1; however, the action taken by Twitter has been updated. In its latest policy update, Twitter omitted the clause that broad claims about the election integrity are policy-compliant. Twitter’s new policy states that claims need to be “disputed” to qualify for removal if they <i>also</i> could undermine faith in the process, and election rigging is now explicitly cited as an example that could qualify for removal.</p> <p><u>Scenario 2:</u> Updated from Non-Comprehensive to Comprehensive. Twitter’s policy explicitly states it will not allow unverified claims about ballot tampering, which specifically applies to Scenario 2.</p> <p><u>Scenario 3:</u> Comprehensive. While Twitter’s <u>synthetic and manipulated media policy</u> doesn’t mention election-related content explicitly, it clearly states the type of action that would be taken for authentic videos taken out of context that either results in “confusion or misunderstanding” or “suggests a deliberate intent to deceive people about the nature or origin of the content...”</p> <p><u>Scenario 4:</u> Comprehensive. As of Twitter’s policy update on September</p>

		<p>results have been certified, inciting unlawful conduct to prevent a peaceful transfer of power or orderly succession.” [Oct. 9]</p> <ul style="list-style-type: none"> • “People on Twitter, including candidates for office, may not claim an election win before it is authoritatively called. To determine the results of an election in the US, we require either an announcement from state election officials, or a public projection from at least two authoritative, national news outlets that make independent election calls. Tweets which include premature claims will be labeled and direct people to our official US election page.” [Oct. 9] 	<p>10, 2020, Twitter will label or remove tweets from candidates claiming victory before the election results are officially called. Twitter’s October 9 update clarifies even further how it will determine the results of an election in order to properly enforce its policy.</p>
	<p>Physical Confrontations Outside Polling Places</p>	<p><u>Scenario 1 (corresponding policy):</u></p> <ul style="list-style-type: none"> • Twitter will remove “Tweets that encourage violence or call for people to interfere with election results or smooth operation of polling places.” <p><u>Scenario 2:</u></p> <ul style="list-style-type: none"> • Twitter will remove “Tweets that encourage violence or call for people to interfere with election results or smooth operation of polling places.” <p><u>Scenario 3:</u></p> <ul style="list-style-type: none"> • “Tweets meant to incite interference with the election process or with the implementation of election results, such as through violent action, will be subject to removal. This covers all Congressional races and the Presidential Election.” 	<p><u>Scenario 1:</u> Comprehensive. Twitter’s policy specifies it will address posts that encourage or call to action individuals or a group to interfere in the procedures or operations of a polling place, which includes who is allowed to be a poll watcher.</p> <p><u>Scenario 2:</u> Comprehensive. Twitter’s policy specifies it will address posts like that in Scenario 1, that encourage or call to action people to interfere in the operations of a polling place, which includes who is allowed to be a poll watcher.</p> <p><u>Scenario 3:</u> Comprehensive. Twitter’s policy specifically addresses tweets that incite interference “through violent action”; a tweet asking other users to bring a weapon to the polls, would likely fall under this policy.</p>

<p style="text-align: center;">T W I T T E R</p>	<p>Consequences of Infringement</p>	<p>Twitter’s policy states it will take action depending on the severity, type of violation and the accounts’ history of previous violations:</p> <ul style="list-style-type: none"> ● “Tweet deletion: The first time you violate this policy, we will require you to remove this content. We will also temporarily lock you out of your account before you can Tweet again.” ● “Profile modifications: If you violate this policy within your profile information (e.g., your bio), we will require you to remove this content. We will also temporarily lock you out of your account before you can Tweet again. If you violate this policy again after your first warning, your account will be permanently suspended.” ● “Labeling: In circumstances where we do not remove content which violates this policy, we may provide additional context on Tweets sharing the content where they appear on Twitter. This means we may: <ul style="list-style-type: none"> ○ Apply a label to the content where it appears in the Twitter product; ○ Show a warning to people before they share or like the content; ○ Reduce the visibility of the content on Twitter and/or prevent it from being recommended; and/or ○ Provide a link to additional explanations or clarifications, such as in a Twitter Moment or relevant Twitter policies.” ● Labeling: “In most cases, we will take all of the above actions on Tweets we label. We prioritize producing Twitter Moments in cases where misleading content on Twitter is gaining significant attention and has caused public confusion on our service.” [Sept. 10] ● “Permanent suspension: For severe or repeated violations of this policy, accounts will be permanently suspended.” ● Public-Interest Exception: Twitter allows people to view tweets from elected government officials that would otherwise be removed. Instead of removing the material, Twitter places the tweet behind a notice that requires people to click through to view the content. Twitter limits public-interest exceptions to elected and government officials given the significant public interest in knowing and being able to discuss their actions and statements. <p>Policies to introduce friction:</p> <ul style="list-style-type: none"> ● “We currently may label Tweets that violate our policies against misleading information about civic integrity, COVID-19, and synthetic and manipulated media. Starting next week, when people attempt to Retweet one of these Tweets with a misleading information label, they will see a prompt pointing them to credible information about the topic before they are able to amplify it.” ● “Tweets with labels are already de-amplified through our own recommendation systems and these new prompts will give individuals more context on labeled Tweets so they can make more informed decisions on whether or not they want to amplify them to their followers.” ● “We will now add additional warnings and restrictions on Tweets with a misleading information label from US political figures (including candidates and campaign accounts), US-based accounts with more than 100,000 followers, or that obtain significant engagement. People must tap through a warning to see these Tweets, and then will only be able to Quote Tweet; likes, Retweets and replies will be turned off, and these Tweets won’t be algorithmically recommended by Twitter.”
--	--	--

- “Beginning October 20 through at least the end of Election week in the US”:
 - “We will encourage people to add their own commentary prior to amplifying content by prompting them to Quote Tweet instead of Retweet. People who go to Retweet will be brought to the Quote Tweet composer where they’ll be encouraged to comment before sending their Tweet. Though this adds some extra friction for those who simply want to Retweet, we hope it will encourage everyone to not only consider why they are amplifying a Tweet, but also increase the likelihood that people add their own thoughts, reactions and perspectives to the conversation. If people don’t add anything on the Quote Tweet composer, it will still appear as a Retweet.”
 - “We will prevent “liked by” and “followed by” recommendations from people you don’t follow from showing up in your timeline and won’t send notifications for these Tweets. These recommendations can be a helpful way for people to see relevant conversations from outside of their network, but we are removing them because we don’t believe the “Like” button provides sufficient, thoughtful consideration prior to amplifying Tweets to people who don’t follow the author of the Tweet, or the relevant topic that the Tweet is about.”
 - “We will only surface Trends in the “For You” tab in the United States that include additional context. That means there will be a description Tweet or article that represents or summarizes why that term is trending. ... This will help people more quickly gain an informed understanding of the high volume public conversation in the US and also help reduce the potential for misleading information to spread.” [Oct. 9]

YouTube

YouTube's policy is comprehensive for content related to procedural and participation interference but non-comprehensive for content related fraud. In February 2020, YouTube [laid out](#) its policy regarding the upcoming election; a YouTube spokesperson [stated](#) that a video's context and content would dictate whether or not it would be taken down. The spokesperson also stated that the platform would focus on videos that are doctored to mislead viewers beyond clips taken out of context, citing the [doctored video](#) of Nancy Pelosi in which her speech appeared to be slurred, as a violation of the policy. A video of former President Joe Biden that was deceptively cut to wrongly suggest he had made racist remarks would not be removed, because the video was not doctored. Yet regardless of the technical differences, on content alone, both videos are arguably misleading and could create real-world harm.

In late-September 2020, YouTube updated its election-related [policies](#) to include "false claims that could materially discourage voting" and two new examples. For this reason, their rating on participation interference has changed from Non-Comprehensive to Comprehensive, as it specifically addresses content meant to discourage participation in the electoral process. In a blog post on [September 24](#), YouTube also stated that it will provide an information panel under videos that address voting by mail, linking to "authoritative information by the Bipartisan Policy Center, a bipartisan think tank."

A week before the election on October 27, YouTube published a blog post, "[Our approach to Election Day on YouTube](#)" in which the platform details how it will provide authoritative information related to voting and election results. While it clearly named its authoritative source, The Associated Press, for election results, YouTube hasn't addressed what it will do with false or premature claims of victory.

Red text reflects YouTube’s updates since we first published on August 18.

Y O U T U B E	Categories	Policy Language	Rating and Explanation
	Procedural Interference	<ul style="list-style-type: none"> “Content that advances false claims related to the technical eligibility requirements for current political candidates and sitting elected government officials to serve in office. Eligibility requirements considered are based on applicable national law, and include age, citizenship, or vital status.” “Content aiming to mislead voters about the time, place, means or eligibility requirements for voting, or false claims that could materially discourage voting.” “Incitement to interfere with democratic processes: content encouraging others to interfere with democratic processes, such as obstructing or interrupting voting procedures.” <p>Examples of content not to post:</p> <ul style="list-style-type: none"> “Deliberately telling viewers an incorrect election date.” “Telling viewers they can vote through fake methods like texting their vote to a particular number.” “Giving made up voter eligibility requirements like saying that a particular election is only open to voters over 50 years old.” “we remove content falsely claiming that mail-in ballots have been manipulated to change the results of an election” 	<p>Comprehensive. <u>YouTube’s policy</u> on procedural interference is labeled “Comprehensive” because it specifies the type of procedural content related to the election that is prohibited on the platform.</p>
Participation Interference	<ul style="list-style-type: none"> “Content aiming to mislead voters about the time, place, means or eligibility requirements for voting, or false claims that could materially discourage voting.” “Incitement to interfere with democratic processes: content encouraging others to interfere with democratic processes, such as obstructing or interrupting voting procedures.” <p>Examples of content not to post:</p> <ul style="list-style-type: none"> Telling viewers to create long voting lines with the purpose of making it harder for others to vote “Claiming that a voter’s political party affiliation is visible on a vote-by-mail envelope.” 	<p>Comprehensive. YouTube’s policy has changed from Non-Comprehensive to Comprehensive because it addresses content encouraging voters to interfere with participation and content meant to discourage participating in the electoral process.</p>	

Y O U T U B E	Categories	Policy Language	Rating and Explanation
	Fraud	<ul style="list-style-type: none"> “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.” “Distribution of hacked materials: content that contains hacked information, the disclosure of which may interfere with democratic processes, such as elections and censuses.” (*Depending on the content, this statement could fall under the Procedural Interference category.) <p>Examples:</p> <ul style="list-style-type: none"> “Misattributing a 10 year old video that depicts stuffing of a ballot box to a recent election.” “Videos that contain hacked information about a political candidate shared with the intent to interfere in an election.” (*Depending on the content, this statement could fall under the Procedural Interference category.) 	Non-Comprehensive. While YouTube’s policy provides an explicit example of content that would be considered fraud, its overarching policy is labeled “Non-Comprehensive” as it applies to election-related content. This is in part because the policy is more about technical manipulation than about the content itself.

Y O U T U B E	Categories	Policy Language	Rating and Explanation
	Delegitimization of Election Results	<p>YouTube’s policies around delegitimization of election results do not fit under the scenarios below.</p> <ul style="list-style-type: none"> False claims that non-citizen voting has determined the outcome of past elections. Telling viewers to hack government websites to delay the release of elections results <p><u>Scenario 1 (corresponding policy)</u>: None.</p> <p><u>Scenario 2</u>: None.</p> <p><u>Scenario 3</u>:</p> <ul style="list-style-type: none"> Manipulated Media: “Content that has been technically manipulated or 	<p><u>Scenario 1</u>: None.</p> <p><u>Scenario 2</u>: None.</p> <p><u>Scenario 3</u>: Comprehensive. YouTube’s policy clearly states an example of misattributing an out-of-context video to current elections and covers content that has been technically manipulated.</p>

	<p>doctored in a way that misleads users (beyond clips taken out of context) and may pose a serious risk of egregious harm.”</p> <ul style="list-style-type: none"> ○ Example: “Misattributing a 10 year old video that depicts stuffing of a ballot box to a recent election.” <p><u>Scenario 4</u>: None.</p>	<p><u>Scenario 4</u>: None.</p>
<p>Physical Confrontations Outside Polling Places</p>	<p><u>Scenarios 1 and 2 (corresponding policy)</u>:</p> <ul style="list-style-type: none"> ● “Incitement to interfere with democratic processes: content encouraging others to interfere with democratic processes, such as obstructing or interrupting voting procedures.” <p><u>Scenario 3</u>:</p> <ul style="list-style-type: none"> ● “Incitement to interfere with democratic processes: content encouraging others to interfere with democratic processes, such as obstructing or interrupting voting procedures.” ● “Encourage others to go to a particular place to commit violence, to perform violence or to target individuals or groups with violence” 	<p><u>Scenarios 1 and 2</u>: Comprehensive. YouTube’s policy is labeled “Comprehensive” because it includes any content that incites others to interfere with voting procedures.</p> <p><u>Scenario 3</u>: Updated from Non-Comprehensive to Comprehensive. YouTube’s policy is labeled “Comprehensive” because in its blog post published October 27, 2020, the platform stated it’s policy addressing content that encourages others to go to a particular place to commit violence, applies to polling stations</p>
<p>Consequences of Infringement</p>	<p>YouTube’s policy states it will remove content that violates this policy: first-time violations will receive a warning; for violations after that, YouTube issues a strike against the channel. If a channel gets three strikes, it will be terminated.</p> <p>YouTube will also apply an information panel (which functions like a “label” on other platforms) to videos that discuss voting by mail and federal and presidential candidates. This information panel links to “authoritative information from the Bipartisan Policy Center, a bipartisan think tank.” [blog post Oct. 27]</p> <p>According to YouTube’s blog post published October 27, on November 3, YouTube will “surface a new election results information panel at the top of queries related to the election and under videos that discuss the election. This information panel will note that election results may not be final, and link to Google’s election results feature, which will enable you to track election results in real time.”</p>	

Pinterest

On [September 3, 2020](#), Pinterest updated its election-related policies by adding a new subsection under its “Misinformation” policy, titled “Civic participation misinformation.” EIP is updating our scoring of Pinterest’s policies to reflect that the platform now has a comprehensive policy across our four core categories.

Pinterest will now also limit recommendations about “election-related content (like election memes or slogans) in places like home feed and notifications.” The platform will also restrict search autocomplete and search guides – recommended categories related to your search term –for specific election-related terms, although it did not state which terms.

Red text reflects Pinterest policy updates on September 3, 2020.

P I N T E R E S T	Categories	Policy Language	Rating and Explanation
	Procedural Interference	<p>Pinterest will remove or limit distribution of:</p> <ul style="list-style-type: none"> • “False or misleading information about the dates, times, locations and procedure for voting or census participation.” • “Content that misleads voters about how to correctly fill-out and submit a ballot, including a mail-in ballot, or census form.” 	Updated from None to Comprehensive. Pinterest’s policy specifically addresses content that relates to time, place, and manner of voting, including mail-in ballots.
	Participation Interference	<ul style="list-style-type: none"> • “False or misleading content that impedes an election’s integrity or an individual’s or group’s civic participation, including registering to vote, voting, and being counted in a census.” • “False or misleading information about public safety that is intended to deter people from exercising their right to vote or participate in a census.” • “False or misleading information about who can vote or participate in the census and what information must be provided to participate.” • “False or misleading statements about who is collecting information and/or how it will be used.” • “Threats against voting locations, census or voting personnel, voters or census participants, including intimidation of vulnerable or protected group voters or participants.” 	Updated from Non-Comprehensive to Comprehensive. Pinterest’s updated policies address content that may deter people from participating in the election.

Fraud	<ul style="list-style-type: none"> • “False or misleading content that impedes an election’s integrity or an individual’s or group’s civic participation, including registering to vote, voting, and being counted in a census.” • “Content that encourages or instructs voters or participants to misrepresent themselves or illegally participate” 	<p>Updated from Non-Comprehensive to Comprehensive: Pinterest’s updated policy accounts for fraud, such as voters misrepresenting themselves or illegally participating in the election.</p>
Delegitimization of Election Results	<p><u>Scenario 1 (corresponding policy):</u></p> <ul style="list-style-type: none"> • “Content apparently intended to delegitimize election results on the basis of false or misleading claims.” <p><u>Scenarios 2 and 3:</u></p> <ul style="list-style-type: none"> • “False or misleading content that impedes an election’s integrity” • “Content apparently intended to delegitimize election results on the basis of false or misleading claims.” • “Content that encourages or instructs voters or participants to misrepresent themselves or illegally participate” <p><u>Scenario 4:</u></p> <ul style="list-style-type: none"> • “Content apparently intended to delegitimize election results on the basis of false or misleading claims.” 	<p><u>Scenario 1:</u> Updated from None to Non-Comprehensive. It is unclear if Pinterest’s policy will directly apply to the broad claim that the “election is rigged.” This specific scenario will likely rely on interpretation by the moderation team, specifically on the perception of the user’s intent and image accompanying the text, given the type of content on Pinterest.</p> <p><u>Scenarios 2 and 3:</u> Updated from Non-Comprehensive to Comprehensive. Pinterest’s updated policy comprehensively addresses Scenarios 2 and 3.</p> <p><u>Scenario 4:</u> Non-Comprehensive. While Pinterest’s updated policy addresses delegitimization of election results, the policy does not address this specific scenario.</p>
Physical Confrontations Outside Polling Places	<p><u>Scenarios 1 and 2 (corresponding policy):</u></p> <ul style="list-style-type: none"> • “False or misleading content that impedes an election’s integrity or an individual’s or group’s civic participation, including registering to vote, voting and being counted in a census.” • “Threats against voting locations, census or voting personnel, voters or census participants, including intimidation of vulnerable or protected group voters or participants.” 	<p><u>Scenario 1:</u> Comprehensive. Pinterest’s policies are labeled “Comprehensive” because they cover content encouraging users to misrepresent themselves or illegally participate.</p> <p><u>Scenario 2:</u> Comprehensive. Pinterest’s policies are labeled “Comprehensive”</p>

		<p><u>Scenario 3:</u></p> <ul style="list-style-type: none"> • “False or misleading content that impedes an election’s integrity or an individual’s or group’s civic participation, including registering to vote, voting and being counted in a census.” • “Threats against voting locations, census or voting personnel, voters or census participants, including intimidation of vulnerable or protected group voters or participants.” 	<p>because they address participants misrepresenting themselves and content that impedes others’ ability to vote.</p> <p><u>Scenario 3:</u> Non-Comprehensive. Pinterest’s policies are labeled “Non-Comprehensive” because while they address posts that include threats against voting locations, Pinterest’s policies do not specifically mention calls to action that may incite violence.</p>
	<p>Consequences of Infringement</p>	<p>Pinterest’s policy states, “We remove or limit distribution of false or misleading content that may harm Pinners’ or the public’s well-being, safety or trust.”</p>	

Nextdoor

Nextdoor updated its policies since EIP first published this PDF in August 2020. Its new policies now cover more categories of election-related content, but its policies are still non-comprehensive.

Red text reflects Nextdoor’s policy updates as of August, 2020.

N E X T D O O R	Categories	Policy Language	Rating and Explanation
	Procedural Interference	<ul style="list-style-type: none"> “bans any inaccurate content about the time, place, means, or eligibility requirements to vote in any local or national elections in the U.S.” “False or misleading information that could prevent or discourage people from voting, cause their votes not to be counted, or interfere with the election process.” 	Updated from Comprehensive to Non-Comprehensive. Nextdoor’s policies are labeled “Non-Comprehensive” because they do not address the full scope of the voting process, such as time, place and manner. Nextdoor’s policy does not specify what is considered interfering “with the election process.”
	Participation Interference	<ul style="list-style-type: none"> “False or misleading information that could prevent or discourage people from voting, cause their votes not to be counted, or interfere with the election process.” 	Updated from None to Non-Comprehensive. Nextdoor’s policies are labeled “Non-Comprehensive” because they do not address threats to personal safety and participation in the election process.
	Fraud	<ul style="list-style-type: none"> “Messages that call for or could incite interference with the vote counting process.” 	Updated from None to Non-Comprehensive. Nextdoor’s policies are labeled “Non-Comprehensive” because while they address interference with the vote counting process, they do not address voters casting ballots in an illegal manner.
Delegitimization of Election Results	<p><u>Scenarios 1, 2 and 3 (corresponding policy):</u></p> <ul style="list-style-type: none"> “False or misleading information that could prevent or discourage people from voting, cause their votes not to be counted, or interfere with the election process.” 	<p><u>Scenarios 1, 2 and 3:</u></p> <p>Non-Comprehensive. While Nextdoor’s policy addresses content that could “cause their votes not to be counted,” and “false or misleading claims about the election results” its policy does not address specific claims that mean to delegitimize the election</p>	

		<ul style="list-style-type: none"> • “False or misleading claims about the results of an election that could lead to interference with the election process.” <p><u>Scenario 4:</u></p> <ul style="list-style-type: none"> • “False or misleading claims about the results of an election that could lead to interference with the election process.” 	<p>process.</p> <p><u>Scenario 4:</u> Nextdoor’s policies are labeled “Non-Comprehensive” because a misleading declaration of victory would not necessarily lead to election interference. The policy also does not specify if this includes a candidate making the claim.</p>
<p>Physical Confrontations Outside Polling Places</p>		<p><u>Scenarios 1 and 2 (corresponding policy):</u></p> <ul style="list-style-type: none"> • “Messages that call for or could incite interference with the vote counting process.” <p><u>Scenario 3:</u></p> <ul style="list-style-type: none"> • “Messages that call for or could incite interference with the vote counting process.” 	<p><u>Scenarios 1 and 2:</u> Non-Comprehensive: Nextdoor’s policies are labeled “Non-Comprehensive” because while they address interference with the vote counting process, they do not address interference with the process of voting.</p> <p><u>Scenario 3:</u> Non-Comprehensive. Nextdoor’s policies are labeled “Non-Comprehensive” because they do not address threats or violence.</p>
<p>Consequences of Infringement</p>	<p>Nextdoor’s policy states it will remove content, as appropriate, that violates its election misinformation policy.</p>		

TikTok

On [October 7, 2020](#), TikTok published a post providing more insight into how it will enforce its election-related policies, and in doing so provided more clarity into how it approaches each of the four categories. It includes not only its election-related policies but also its approach and rationale for each policy and consequences for infringement – a detailed transparency, similar to that of Twitter’s, that is not shared by many other platforms. In addition, it discusses its policies related to accounts that repeatedly violate these policies or whose sole purpose is to spread mis- or disinformation. It has also introduced how it will use friction and curation methods to limit the spread of mis- and disinformation, including reducing discoverability and redirecting search results and hashtags. Lastly, it has compiled a timeline of new policies and enforcement steps taken in a step towards transparency.

Red text reflects TikTok’s policy updates on October 7, 2020 & October 28.

T I K T O K	Categories	Policy Language	Rating and Explanation
	Procedural Interference	<ul style="list-style-type: none"> “Content that misleads community members about elections or other civic processes.” “Claims relating to polling stations on election day that have not yet been verified.” “Content that misrepresents the date of an election.” 	Non-Comprehensive. TikTok’s policy on procedural interference is labeled “Non-Comprehensive” because it does not address the full scope of the voting process, such as place and manner. The statement referring to “claims related to polling stations” may address place of voting misinformation, such as a polling location change, but this is not clear.
	Participation Interference	<ul style="list-style-type: none"> “Attempts to intimidate voters or suppress voting.” <p>TikTok will redirect search results with terms associated with “incitement to violence.”</p> <p>TikTok will block future livestreaming from an account whose livestream “seeks to incite violence or promote hateful ideologies, conspiracies, or disinformation.”</p>	Updated from None to Non-Comprehensive. TikTok’s policies on participation interference are labeled “Non-Comprehensive” because while they address this interference, they do not sufficiently define what voter intimidation or voter suppression is outside of incitement to violence.

	<p>TikTok will add a banner pointing viewers to our election guide content with...“attempts to dissuade people from voting by exploiting COVID-19 as a voter suppression tactic.”</p>	
Fraud	<ul style="list-style-type: none"> “False claims that seek to erode trust in public institutions, such as claims of voter fraud resulting from voting by mail or claims that your vote won’t count.” (*Depending on the content’s exact wording, this statement could fall under the Procedural Interference category.) “Disinformation around voter fraud, such as ballot harvesting.” 	<p>Updated from None to Non-Comprehensive. TikTok’s policies on fraud are labeled “Non-Comprehensive” because while they adequately cover claims of electoral fraud, such as ballot harvesting, they do not address content encouraging others to cast ballots in an illegal manner, or other types of voter fraud.</p>
Delegitimization of Election Results	<p><u>Scenario 1 (corresponding policy):</u></p> <ul style="list-style-type: none"> “Content that misleads community members about elections or other civic processes.” <p><u>Scenario 2:</u></p> <ul style="list-style-type: none"> “Content that misleads community members about elections or other civic processes.” “False claims that seek to erode trust in public institutions, such as claims of voter fraud resulting from voting by mail or claims that your vote won’t count.” <p><u>Scenario 3:</u></p> <ul style="list-style-type: none"> “Content that misleads community members about elections or other civic processes.” “False claims that seek to erode trust in public institutions, such as claims of voter fraud resulting from voting by mail or claims that your vote won’t count.” “Digital Forgeries (Synthetic Media or Manipulated Media) that mislead users by 	<p><u>Scenario 1:</u> Updated from None to Non-Comprehensive. While TikTok’s policy addresses misleading content about elections generally, it is unclear if its policies would apply to the broad claim that the “election is rigged.”</p> <p><u>Scenario 2:</u> Updated from Non-Comprehensive to Comprehensive. TikTok’s policy is labeled comprehensive because it addresses false claims aimed at public institutions and claims that a vote won’t count.</p> <p><u>Scenario 3:</u> Updated from Non-Comprehensive to Comprehensive. TikTok’s policy is labeled “Comprehensive” because it addresses false claims and digital forgeries meant to mislead users.</p> <p><u>Scenario 4:</u> Updated from None to Comprehensive. TikTok’s policy explicitly states that it will address content that shares a premature declaration of victory.</p>

		<p>distorting the truth of events and cause harm to the subject of the video or society.”</p> <p><u>Scenario 4:</u></p> <ul style="list-style-type: none"> • “Reviewed content that shares unverified claims, such as a premature declaration of victory before results are confirmed.” 	
<p>Physical Confrontations Outside Polling Places</p>		<p><u>Scenarios 1 and 2 (corresponding policy):</u></p> <ul style="list-style-type: none"> • TikTok will remove content including any “attempts to intimidate voters or suppress voting.” <p><u>Scenario 3:</u></p> <ul style="list-style-type: none"> • TikTok will redirect search results with terms associated with “incitement to violence” related to the elections. 	<p><u>Scenarios 1 and 2:</u> Non-comprehensive. TikTok does not specifically address calls to action in which individuals may interfere in the procedures or operations of a polling place.</p> <p><u>Scenario 3:</u> Comprehensive. TikTok’s policy is comprehensive because it specifically mentions incitements to violence as related to the elections.</p>
<p>Consequences of Infringement</p>		<p>TikTok’s policy includes various levels of consequences based on the infringement committed. The consequences include:</p> <ul style="list-style-type: none"> • “Remove content for violation of our Community Guidelines.” • “Redirect search results and hashtags to our Community Guidelines.” • “Reduce content discoverability, including by redirecting search results or making such content ineligible for recommendation into anyone’s For You feed.” • “Block the account from future livestreaming.” • “Remove the account and its content.” • “Ban the device, including all linked accounts, and block the ability to create future accounts from that device.” 	

Snapchat

When we first published our analysis of the platforms’ election-related policies, Snapchat belonged to the group of platforms that did not have policies that addressed this content. However, in mid-September 2020, Snapchat updated its policies to address content that aims to undermine the integrity of civic processes. While this update is constructive, its policy around “civic processes” is unclear, beginning with the fact that Snapchat does not define what a civic process is, or what undermining its integrity looks like on its platform.

Red text reflects Snapchat’s policy updates September 2020.

S n a p c h a t	Categories	Policy Language	Rating and Explanation
	Procedural Interference	<ul style="list-style-type: none"> “We prohibit spreading false information that causes harm or is malicious, such as denying the existence of tragic events, unsubstantiated medical claims, or undermining the integrity of civic processes.” 	Non-Comprehensive. Snapchat’s new election-related policy is labeled “non-comprehensive” because while it addresses content that aims to undermine the integrity of civic processes, it does not provide clarity into what “undermining integrity” looks like on its platform
	Participation Interference	None	None
	Fraud	<ul style="list-style-type: none"> “We prohibit spreading false information that causes harm or is malicious, such as denying the existence of tragic events, unsubstantiated medical claims, or undermining the integrity of civic processes.” 	Non-Comprehensive. Snapchat’s policy is labeled “non-comprehensive” because while claims of fraud are generally considered falling under undermining the “integrity of civic processes,” the policy is not clear that it specifically addresses this content.
	Delegitimization of Election Results	<p>Scenario 1, 2, and 3 (corresponding policy):</p> <ul style="list-style-type: none"> “We prohibit spreading false information that causes harm or is malicious, such as denying the existence of tragic events, unsubstantiated medical claims, or undermining the integrity of 	Non-Comprehensive. Snapchat’s policy is labeled “non-comprehensive” for the first 3 scenarios. While these claims of delegitimization of election results can be considered falling under “undermining the integrity of civic processes,” it is unclear from the policy language that it will apply.

	civic processes.”	<u>Scenario 4</u> : None	<u>Scenario 4</u> : None
Physical Confrontations Outside Polling Places	<u>Scenario 1, 2, and 3 (corresponding policy)</u> : None		Scenario 1, 2, and 3: None
Consequences of Infringement	According to Snapchat’s Community Guidelines , the platform may “remove the offending content, terminate or limit the visibility of your content, and/or notify law enforcement.” The Guidelines do not explicitly say what would happen to the election-related content on its platform.		