

# US Election Platform Preparedness

**Comparative Analysis** 

Isabelle Frances-Wright, Ellen Jacobs, Clara Martiny, Max Read, Ella Meyer



Amman | Berlin | London | Paris | Washington DC

Copyright © Institute for Strategic Dialogue (2024). ISD-US is a non-profit corporation with 501(c)(3) status registered in the District of Columbia with tax identification number 27-1282489. Details of the Board of Directors can be found at www.isdglobal.org/isd-board. All Rights Reserved.

# **Election Preparedness Comparative Analysis**

## **Information Integrity (User Generated Content)**

		Meta	X	YouTube	Snapchat	TikTok
1.	Does the platform remove (or downrank) content that includes false or misleading information on how to participate in an election (includes times, dates, methods and eligibility requirements)?	Yes	Yes	Yes	Yes	Yes
2.	Does that platform remove (or downrank) content that falsely asserts there will be widespread election fraud or vote counting errors?	Yes	Partial	Yes	Yes	Yes
3.	Does the platform remove (or downrank) content that contains false or misleading information on candidate eligibility requirements?	Partial	Partial	Yes	Partial	Yes
4.	Does the platform remove false claims regarding the 2020 election results or redirect users to accurate information about the 2020 election results?	Partial	No	Partial	Yes	Yes
5.	Will the platforms remove or downrank content which claims victory before the election is called?	Partial	Unclear	Unclear	Yes	Yes
6.	Is the platform partnering with external experts to identify coordinated influence operations? Have details on the scope and results of those partnerships been publicly disclosed?	No/ Unclear	No/ Unclear	No/ Unclear	No/ Unclear	No/ Unclear
7.	Are there policy exemptions (or instances where EDSA exceptions may be made) for political figures and candidates?	Yes	Yes	Yes	No	Yes
8.	Is the platform partnered with third party fact checkers? If so, what are the turnaround times?	Yes / Unclear	No / N/A	Yes / Unclear	Yes / Unclear	Yes / Unclear
9.	Does the violation of misinformation policies related to the election or civic integrity result in account suspension? If so, after how many violations?	Yes / Unclear	Unclear / Unclear	Yes / Yes	Yes / Unclear	Yes / Unclear

Will the platform be promoting, verified, trusted election information from third party sources to platform users?	Yes	Yes	Yes	Yes	Yes
Are the platform's integrity and authenticity policies applied to livestreaming product features (both audio and audiovisual)?	Yes	Yes	Yes	Yes	Yes
Does the platform remove or downrank synthetically generated or misleadingly edited content of candidates, elected officials or individuals involved in the administration of elections, that could be reasonably judged to cause harm to those individuals?	Partial	Partial	Partial	Yes	Unclear
itical Ads					
	Meta	X	YouTube	Snapchat	TikTok
Does the platform allow political ads?	Yes	Yes	Yes	Yes	No
If so, are political ads subject to the same or more stringent content policies regarding misinformation?	More Stringent	Equally Stringent	Equally Stringent	Equally Stringent	N/A
Are political ads eligible for 3rd party fact-checking?	No	No	Unclear	Yes	N/A
Are political ads required to disclose funding sources?	Yes	Partial	Yes	Yes	N/A
Are political ads subject to the same or more stringent content policies, specifically regarding hate and extremist content?	More Stringent	Equally Stringent	Equally Stringent	Equally Stringent	N/A
Are political ads permitted to use synthetic media or misleadingly edited media?	Yes (requires disclosure)	Yes	Yes (requires disclosure)	No	N/A
Is there a publicly available ads library?	Yes	Partial	Yes	Partial	No
Are content creators required to disclose when political content has been sponsored, and who has sponsored the content?	Yes	Partial	Partial	N/A	N/A
te Speech and Violent Extremism (User Generat					
	Meta	X	YouTube	Snapchat	TikTok
Is the platform putting in place prevention measures for domestic hate and extremism, such as product interventions to disincentivize hateful behavior?	Partial	No	Partial	No	Partial
	election information from third party sources to platform users?  Are the platform's integrity and authenticity policies applied to livestreaming product features (both audio and audiovisual)?  Does the platform remove or downrank synthetically generated or misleadingly edited content of candidates, elected officials or individuals involved in the administration of elections, that could be reasonably judged to cause harm to those individuals?  If so, are political ads subject to the same or more stringent content policies regarding misinformation?  Are political ads eligible for 3rd party fact-checking?  Are political ads required to disclose funding sources?  Are political ads subject to the same or more stringent content policies, specifically regarding hate and extremist content?  Are political ads permitted to use synthetic media or misleadingly edited media?  Is there a publicly available ads library?  Are content creators required to disclose when political content has been sponsored, and who has sponsored the content?  The Speech and Violent Extremism (User Generative Speech and Violent Extremism (User	Are political ads eligible for 3rd party fact-checking?  Are political ads eligible for 3rd party fact-checking?  Are political ads subject to the same or more stringent content policies, specifically regarding hate and extremist content?  Are political ads permitted to use synthetic media or misleadingly edited media?  Are political ads permitted to use synthetic media or misleadingly edited content of candidates, elected officials or individuals involved in the administration of elections, that could be reasonably judged to cause harm to those individuals?  Partial  Meta  Does the platform allow political ads?  Yes  If so, are political ads subject to the same or more stringent content policies regarding misinformation?  Are political ads eligible for 3rd party fact-checking?  Are political ads required to disclose funding sources?  Yes  Are political ads subject to the same or more stringent content policies, specifically regarding hate and extremist content?  Are political ads permitted to use synthetic media or misleadingly edited media?  Are political content policies, specifically regarding hate and extremist content?  Yes  Are content creators required to disclose when political content has been sponsored, and who has sponsored the content?  Yes  Are Speech and Violent Extremism (User Generated Content) Meta  Is the platform putting in place prevention measures for domestic hate and extremism, such as product interventions to disincentivize hateful	election information from third party sources to platform users?  Are the platform's integrity and authenticity policies applied to livestreaming product features (both audio and audiovisual)?  Personal Partial Par	election information from third party sources to platform users?  Are the platform's integrity and authenticity policies applied to livestreaming product features (both audio and audiovisual)?  Does the platform remove or downrank synthetically generated or misleadingly edited content of candidates, elected officials or individuals involved in the administration of elections, that could be reasonably judged to cause harm to those individuals?  Meta X YouTube  Partial Partial Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial  Partial	election information from third party sources to platform users?  Are the platform users?  Are the platform's integrity and authenticity policies applied to livestreaming product features (both audio and audiovisual)?  Does the platform remove or downrank synthetically generated or misleadingly edited content of candidates, elected officials or individuals involved in the administration of elections, that could be reasonably judged to cause harm to those individuals?  Partial  Partial

29.	Is there a dedicated US based election integrity team, if so, how large is that team?	No/ Unclear	No/ Unclear	Yes / Unclear	No/N/A	Yes
		Meta	X	YouTube	Snapchat	TikTok
Re	sourcing (Internal and External)					
28.	Does the platform remove or downrank content that calls for or incites coordinated harassment (both online and offline)?	Yes	Yes	Yes	Yes	Yes
27.	Are the platform's hate speech, violence and extremism policies applied to comment features?	Yes	Yes	Yes	Yes	Yes
26.	Are the platform's hate speech, violence and extremism policies applied to livestreaming product features (both audio and audiovisual)?	Yes	Yes	Yes	Yes	Yes
25.	Are there any measures in place to specifically protect candidates and election officials from targeted hate speech, harassment, or threats on the platforms?	Partial	No	No	No	No
24.	Does the platform remove (or downrank) content which includes violent threats?	Yes	Yes	Yes	Yes	Yes
23.	Are there policies against promotion or incitement of election violence, militarization, and calls for intimidation (i.e., bringing weapons to polling stations)?	Yes	Yes	Yes	Yes	Yes
22.	Is the company working/partnering with external experts to support in proactively identifying hate speech and violent extremist content on its platform/s and corresponding policy gaps throughout the election cycle?	Yes/ Unclear	Yes/ Unclear	Yes/ Unclear	Yes/ Unclear	Yes / Unclear

#### **30.** Are Spanish speakers included on that team? N/A N/A Unclear N/A Yes **31.** Have there been staffing cuts to election and civic integrity teams since 2022? Yes Yes Unclear No No **32.** Is there a viral content review system in place? If so, are minimum thresholds or definitions for virality disclosed? Yes/ No/ No/ Yes/ Yes/ Unclear Unclear Unclear Unclear Unclear

## **Transparency**

	Meta	X	YouTube	Snapchat	TikTok
33. Will the platform be releasing findings on identified disinformation networks on an ongoing basis (monthly or more) throughout the election cycle?	No	No	No	No	Yes
34. Will findings on malicious actors and/or harmful content trends relevant to civic discourse be shared on an ongoing basis ((monthly or more) during the election cycle?	No	No	No	No	Yes
<b>35.</b> Have civil society and academic researchers been provided API access for the purposes of independent election integrity research?	Partial	Partial	Partial	No	Partial
State Affiliated Media					
	Meta	X	YouTube	Snapchat	TikTok
<b>36.</b> Does the platform clearly label state affiliated media at the post/video level?	Yes	Yes	Yes	No	Yes
<b>37.</b> Does the platform clearly label state affiliated media at the account level?	Yes	Yes	No	No	Yes
<b>38.</b> Does the platform prohibit state affiliated media from advertising within the US?	Yes	Yes	No	Partial	Yes
<b>39.</b> If yes, does the platform label such ads?	N/A	N/A	Yes	Unclear	N/A

# **Information Integrity**

Does the platform remove (or downrank) content that includes false or misleading information on how to participate in an election (includes times, dates, methods and eligibility requirements)?

#### Meta

**Yes.** According to its Misinformation Policy, Meta removes false claims about how to participate in election processes, including times, dates, methods, voter registration, eligibility requirements, and more.

#### X

**Yes.** X's Civic Integrity Policy prohibits "manipulating or interfering in elections or other civic processes." and "verifiably false or misleading information about how to participate in an election or other civic process". The company's definition of civic processes includes political elections, censuses, and referenda or ballot initiatives. If content is found in violation of the policy, X will have the content "restricted on X by: Excluding the post from search results, trends, and recommended notifications, removing the post from the For you and Following timelines, restricting the post's discoverability to the author's profile, restricting Likes, replies, reposts, quotes, bookmarks, share, pin to profile, or Edit post, downranking the Post in replies. Posts enforced under this policy will receive a label informing both Post authors and viewers that we have limited the Post's visibility."

#### YouTube

Yes. Under its Election Misinformation Policy, YouTube will remove content that aims to "mislead voters" about how to participate in an election.

#### **Snapchat**

**Yes.** According to its Harmful False or Deceptive Information Policy, Snapchat removes content that includes misleading information about election or civic procedures (dates, times, eligibility).

#### TikTok

**Yes.** TikTok does not <u>allow</u> false or misleading claims that misrepresents the date of an election or provides false information on how to register to vote and will remove it for violation of their Community Guidelines.

Does that platform remove (or downrank) content that falsely asserts there will be widespread election fraud or vote counting errors?

#### Meta

**Yes.** According to Meta's <u>Voter or Census Interference</u> <u>policy</u> misinformation about "whether a vote will be counted" will be removed.

#### X

**Partial.** X will restrict reach of "verifiably false or misleading information about the circumstances surrounding a civic process intended to intimidate or dissuade people from participating in an election or other civic process."

#### YouTube

**Yes.** YouTube <u>prohibits</u> content that includes "false claims that could materially discourage voting, including those disputing the validity of voting by mail; and content that encourages others to interfere with democratic processes

#### **Snapchat**

**Yes.** Snapchat removes content <u>delegitimizing civic</u> <u>processes</u>, including false or misleading claims about election results.

#### TikTok

Yes. TikTok previously stated on an election integritywe resource they removed content that includes "removes content thatincludes "false claims that seek to erode trust in public institutions, such asclaims of voter fraud resulting from voting by mail or claims that your votewon't count; content that misrepresents the date of an election; attempts to intimidate voters or suppress voting; and more." While this webpagehas since been removed, TikTok confirmed to ISD that the policy remains the same.

Does the platform remove (or downrank) content that contains false or misleading information on candidate eligibility requirements?

#### Meta

Partial. Meta removes misinformation about "whether

a candidate is running or not," but does not specifically prohibit false information regarding candidate eligibility requirements within its Misinformation Policy. However, this content may be restricted in reach as part of Meta's fact checking process.

Partial. X will restrict the reach of misleading information about "procedures to participate in a civic process," which could include running for office, but does not specifically state this.

#### YouTube

Yes. YouTube will remove content that violates their election misinformation policy, which includes spreading false claims "related to the technical eligibility requirements for current political candidates." YouTube will also penalize the channel if this is not their first offense.

## **Snapchat**

Partial. Snapchat's Community Guidelines expresslyprohibit misinformation related to "eligibility requirements for participation"in civic processes, which could include running for office, but does not specifically state this.

#### **TikTok**

**Yes.** TikTok states they do not allow misinformation about "the qualifications for candidates to run for office

Does the platform remove false claims regarding the 2020 election results or redirect users to accurate information about the 2020 election results?

#### Meta

Partial. There is no reference to past or future election results in Meta's misinformation policy. However, content may be restricted in reach as part of Meta's fact checking program "We have nearly 100 fact-checking partners around the world that address viral misinformation. When they rate content as false, we move it lower in Feed by default and show additional information so people can decide what to read, trust, and share." The focus here however appears to be content that has gained significant traction, as opposed to content containing false claims with lower reach. It should be noted however that ads which include "premature claims of victory" or "call into question the legitimacy of an upcoming or ongoing election" are not prohibited. This would likely cover some of the content referenced in this question.

No. X makes no reference to prior election results in its civic integrity policy -- in fact, in 2022, the platform announced it was no longer enforcing its civic integrity policy in relation to false or misleading information about the 2020 election results.

#### YouTube

Partial. YouTube allows users to "debate political ideas, even those that are controversial or based on disproven assumptions". However, YouTube stated to ISD that they "display an information panel above related search results (e.g., search for "US 2020 election" on YT), and below relevant videos."

#### **Snapchat**

Yes. Snapchat prohibits the "content aiming to delegitimise democratic institutions on the basis of false or misleading claims about election results"

#### **TikTok**

Yes. TikTok states in its election and civic integrity policy that it does now allow content that includes misinformation about the "final results or outcome of an election".

Will the platforms remove or downrank content which claims victory before the election is called?

#### Meta

Partial. Meta does not penalize user-generated content that claims victory before an election is called (but does prohibit paid ads that do). If there is a claim within a piece of content applicable for its Fact checking program (which does not include direct speech from political candidates), and receives certain ratings, it may be downranked.

### X

Unclear. According to the X Rules, users "may not use X's services for the purpose of manipulating or interfering in elections or other civic processes". However, it is not clear whether prematurely claiming victory in an election qualifies as electoral interference under their policy. Additionally, X stated to ISD that "community notes may also be applied to false claims of victory".

#### YouTube

**Unclear.** During the 2020 US election, YouTube aimed to connect people to "authoritative news sources" and linked election results to videos about the 2020 election.

#### **Snapchat**

**Yes.** Snapchat prohibits the "content aiming to delegitimise democratic institutions on the basis of false or misleading claims about election results".

#### **TikTok**

**Yes.** TikTok stated to ISD it "will reduce discoverability of content with unverified claims about elections, such as declarations of victory before results are confirmed, by making it ineligible for the For You feed or removing the content.".

Is the platform partnering with external experts to identify coordinated influence operations? Have details on the scope and results of those partnerships been publicly disclosed?

#### Meta

**Yes/No.** While Meta stated to ISD that "We work with governments, law enforcement agencies, nonprofits, civil rights groups and other tech companies to stop emerging threats", no specific partnerships have been disclosed.

#### X

**Yes/No.** While X stated to ISD that it "coordinates with external experts via cross-platform information sharing partnerships", no specific partnerships have been disclosed.

#### YouTube

Yes/unclear. For foreign and domestic coordinated influence operations aiming to interfere in democratic processes, YouTube works with Google's Threat Analysis Group to "identify bad actors and terminate their channels." In the TAG bulletins, Google sometimes credits third-parties with surfacing leads, an example being: "We received leads from LinkedIn and Graphika that supported us in this investigation." However, no specific partnerships could be identified. Google did also state to ISD that "Through TAG, we also work with other technology companies to share intelligence and best practices, and share threat information with law enforcement."

## **Snapchat**

**Yes/No.** Snapchat stated it "maintains partnerships with ... as well as research organizations such as the Atlantic Council's Digital Forensic Research Lab, and others. Snap leverages third-party vendor services to assess information threats, and Snap also participates in cross-industry signal-sharing initiatives It is unclear who

these experts are. Details on the scope and outcomes of these partnerships is unclear.

#### **TikTok**

**Yes/No.** TikTok stated to ISD that it "works with a variety of experts to identify potential policy-violating material and coordinated disinformation campaigns on the platform." However, details on the scope of these partnerships is unclear.

Are there policy exemptions (or instances where EDSA exceptions may be made) for political figures and candidates?

#### Meta

Yes. According to Meta's page on Fact-Checked Misinformation in its Transparency Center, opinion content and speech from politicians is not eligible to be fact-checked. Additionally, content which Meta judges to be "newsworthy" will be left up even if it contradicts Meta's Community Guidelines, and this may apply to content from politicians.

Former candidates, former officials or organizations not affiliated with candidates such as advocacy groups and super PACs are not exempt from fact-checking (i.e. are not defined as political figures).

#### X

**Yes.** X has a <u>page</u> for defining public-interest exceptions for content on the platform.

In certain and rare instances, X may leave up a post from "an elected or government official that would otherwise violate our policies, given the significant public interest in knowing and being able to discuss their actions and statements." For example, on January 6, 2021, X decided to leave up Trump's tweets despite them inciting violence. According to X, senior leaders from the Trust & Safety team make the final decision to remove the post or apply the exception.

X provides a notice providing context about the rule violation and limits the ability to engage with the post.

Their criteria for public-interest exceptions is:

1) the post violates one or more X Rules; 2) the post was shared by a high-profile account; and, 3) The account represents a current or potential member of a local, state, national, or supra-national governmental or legislative body — currently elected/appointed, candidates or nominees, or registered political parties'.

They claim they are less likely to make exceptions when the post includes references to: terrorism/violent extremism, violence, illegal or certain regulated goods or services, suicide or self harm, election integrity, and private information.

#### YouTube

**Yes.** Sometimes, content may be exempt from policy enforcement if the content is Educational, Documentary, Scientific, and/or Artistic (EDSA exception).

#### **Snapchat**

No.

#### **TikTok**

**Yes.** Content which seems to violate the guidelines is allowed to remain if it is deemed 'public interest', in the following categories: Documentary, Educational, Medical and Scientific, Counterspeech, Satirical, or Artistic.

TikTok enforces "different account restrictions in keeping with (TikTok's) commitment to human rights and free expression." If political figures reach the strike limit they will not appear in the For You and Following feeds for 90 days. Political figures are defined as Government, Politician, and Political Party Accounts (GPPPA), and are subject to the same policies as well as further restrictions regarding monetization, fundraising and music use.

Is the platform partnered with third party fact checkers? If so, what are the turnaround times?

#### Meta

Yes/Unclear. Meta had partnered with over 90 organizations certified by the International Fact-Checking Network (IFCN) in over 60 languages. Viral content and content that is provably falsifiable is prioritized.

Meta also stated to ISD that they have "built the largest independent fact-checking network of any platform with 11 independent partners in the US, including seven who cover content in Spanish."

While falsifiable content is not removed by factcheckers, it is demoted and labelled as false, with a link to the fact-checker's debunking article. Users who previously shared the article are notified and duplicate content is labelled by Al. The removal of content only occurs when it has violated community standards, which is separate from Meta's fact checking program. It is unclear what the turnaround times are.

#### X

No/N/A. X relies on Community Notes, where users can leave notes on a X post providing context, a different point of view, or a link debunking the claim made in the post, not third-party fact-checkers. Users can sign up to be part of the group of contributors. It is unclear what the turnaround times are though X stated to ISD that "notes may appear within several hours of a Post being published".

#### YouTube

Yes/Unclear. YouTube seems to automate fact checking using the publicly available Schema. org ClaimReview markup which any publisher can participate in if they are a verified signatory of the International Fact-Checking Network Code of Principles.

In November 2022, Google and YouTube announced a \$13.2 million grant to the International Fact-Checking Network to support fact-checking organizations around the world. They've also invested in media literacy programs.

YouTube adds fact checks from third-party publishers and information panels with credible links to thirdparty sources. For example, following the reversal of Roe v. Wade in the US, YouTube announced it was cracking down on abortion misinformation and added information panels about abortion to abortion-related

It is unclear what the turnaround times are. In January 2022, over 80 fact-checking organizations signed a letter highlighting YouTube's "insufficient" response to misinformation.

#### Snapchat

Yes/Unclear. Snapchat stated to ISD that they have "partnerships with IFCM-member fact-checking organizations, including Poynter, Boom, and Logically Facts".

#### **TikTok**

**Yes/No.** In their Election Integrity, Safety Center page, TikTok states "Through our Global Fact-Checking Program, we partner with 19 fact-checking organizations who assess the accuracy of content on TikTok in more than 50 languages. These independent partners do not moderate content directly on TikTok, but assess whether a claim is true, false, or unsubstantiated so that our moderators can take action based on our Community Guidelines. They also share proactive insight reports that help us detect harmful misinformation and anticipate misinformation trends. Learn more."

Does the violation of misinformation policies related to the election or civic integrity result in account suspension? If so, after how many violations?

Yes/Unclear. After 'repeated warnings' for violating platform policies, accounts will be restricted. These restrictions include, for example, a 30-day ban on creating content after the 10th 'strike'.

In times of "civil unrest" accounts of public figures who share inciteful content "may" also be restricted. Donald Trump was banned for 2 years for "severe" violation of policy as a result of the January 6 insurrection.

It is unclear at what point Meta disables the offending account.

It is unclear exactly what qualifies a strike, since a post which violates community guidelines may vary in 'severity' of violation.

**Unclear/Unclear.** According to X's enforcement policies, accounts may be suspended if X determines the user "has engaged in repeated violations of" their policies or violated specific policies that cause significant risk to X or other users. X does not mention their Civic Integrity policy in these "specific policies." X also stated to ISD that "We do not disclose details on our internal enforcement guidance and frameworks (e.g. details of a particular strike system)."

#### YouTube

Yes/Yes. If users violate Community Guidelines (including election misinformation policies) once, they will get a warning with no penalty. Users will be offered a policy training to allow the warning to expire after 90 days, but if the same policy is violated within 90 days, the warning will not expire and the channel will be given a strike. If a different policy is violated, the channel will receive another warning.

3 strikes within 90 days results in the termination of a channel.

YouTube reserves the right to terminate channels or accounts for repeated violations of Community Guidelines, or a single case of "severe abuse."

#### **Snapchat**

Yes/Unclear. Snapchat has a strike system for violations of policy, which include against misinformation. The violation count is unclear. If a user acquires "too many strikes over a defined period of time" the account is banned.

#### TikTok

**Yes/Unclear.** Users are banned if they accrue enough strikes from violating community guidelines, for repeated violations or a severe single violation such as promoting violence. TikTok also stated to ISD "Content such as CSAM, the promotion of terrorism or credible threats of violence are zero-tolerance violations that result in immediate suspension. For violations when using certain features, such as LIVE or direct messages, we may also temporarily restrict access to that feature

to ensure the account doesn't immediately reengage in violative behavior."

## Will the platform be promoting verified, trusted election information from third party sources to platform users?

#### Meta

Yes, according to Meta's blog post highlighting how the platform is preparing for the 2024 US elections, Meta will be connecting users with "details about voter registration and elections from their state and local election officials." Meta will also keep its Voting Information Center live on Facebook, and redirect users to official information about voting when they search for terms related to the 2024 election on Facebook and Instagram.

#### X

Yes. X stated to ISD that it will "prompt users to find information from primary sources - such as vote. gov - when searching for election related information. X will also integrate reminders related to voter registration and voting". X also plans to continue scaling Community Notes, which are crowd-sourced.

#### YouTube

Yes. In their blog post highlighting plans for the 2024 US elections, YouTube stated their systems would "recommend election news and information from authoritative sources and display information panels at the top of search results and below videos to provide even more context."

#### Snapchat

Yes. Snapchat has partnered with Vote.org to provide and facilitate access to voting information and encourage civic participation in the US.

#### **TikTok**

Yes. TikTok has launched a US Elections Center in partnership with Democracy Works to provide users with reliable voting information for the 2024 elections.

Are the platform's integrity and authenticity policies applied to livestreaming product features (both audio and audiovisual)?

Yes. Under Meta's Terms of Service, all users of any Meta Products must abide by the Community Standards, which would include livestreaming product features.

Additionally, Facebook Live policies and guidelines state all livestreams must comply with Meta's Community Standards.

Yes. According to X's Terms of Service, all users using X services must comply with X's Rules and Policies. Additionally, X links its Sensitive Media policy in its blog post about using X Live.

#### YouTube

Yes. All of YouTube's Community Guidelines apply to live streams. YouTube also provides moderators with a set of tools to moderate live chats during live streams.

The platform also plans to keep all of its election misinformation policies in place for 2024, which included giving live election results.

#### **Snapchat**

Yes. In Snap's Terms of Service, the company states that all users using any of Snap's services must comply with the Community Guidelines, which have a clause on election integrity.

#### **TikTok**

**Yes.** In TikTok's Terms of Service, the platform states that use of TikTok's services requires compliance with the Community Guidelines, which includes election integrity policies.

Does the platform remove or downrank synthetically generated or misleadingly edited content of candidates, elected officials or individuals involved in the administration of elections that could be reasonably judged to cause harm to those individuals?

#### Meta

Partial. Videos are only removed if the content violates a non-Al related policy. Meta stated to ISD that "for misleading manipulated media that does not otherwise violate Meta's Community Standards, it may be subject to label and reduced distribution through our factchecking program"...

Partial. X removes any misleading media that could cause "serious harm." Satire and memes are not included in this policy. Depending on the content, this type of media may be actioned under other X policies.

#### YouTube

**Partial.** YouTube does not allow video content that has been technically manipulated (usually beyond clips taken out of context) to fabricate events where there is a serious risk of egregious harm,

Recently, YouTube announced that it was requiring creators to "disclose to viewers when realistic content [...] is made with altered or synthetic media," and will display a more "prominent" label on videos touching on more "sensitive topics," including elections.

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 (usually beyond clips taken out of context) and may pose a serious risk of egregious harm." The platform does not, however, define "egregious harm". Some examples that YouTube provides include:

"Inaccurately translated video subtitles that inflame geopolitical tensions creating serious risk of egregious harm."

"Videos that have been technically manipulated (usually beyond clips taken out of context) to make it appear that a government official is dead."

"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."

## **Snapchat**

Yes. See above: Snapchat prohibits manipulated content for "false or misleading purposes," which would cover harm.

#### **TikTok**

Partial. TikTok's recently updated guidelines on synthetic or manipulated media say that Al generated content of public figures is permitted if the content is clearly labeled as such, and in the context of elections does not include a "political endorsement". However, it is unclear under TikTok's definition whether an election worker would be considered a "public figure" or how "political endorsement" is defined.

Realistic scenes must be clearly disclosed as manipulated. This can be done using a sticker or caption saying 'synthetic', 'fake', 'not real', or 'altered' for example.

## **Political Ads**



## Does the platform allow political ads?

#### Meta

Yes.

## YouTube

Yes.

**Yes.** In late August 2023, X started allowing political content ads in Argentina, Brazil, Canada, Chile, Colombia, Ecuador, Israel, Japan, Mexico, Peru, United Kingdom and the United States and political campaign ads in Argentina, Colombia, Ecuador, Japan, Mexico, Peru, United Kingdom and the United States.

#### **Snapchat**

Yes.

#### **TikTok**

No.

## If so, are political ads subject to the same or more stringent content policies regarding misinformation?

#### Meta

**More stringent.** Meta has additional policies (that would be categorized by ISD as relating to platform integrity) for political ads. Meta prohibits ads that "discourage voting, prematurely claim victory, attempt to delegitimize the election, or are inconsistent with health authorities on voting safely. "

Social issue ads, which "seek to influence public opinion through discussion, debate or advocacy for or against important topics, such as health and civil and social rights" are subject to the same policies.

**Equally stringent.** All X Ads are subject to the X Rules, which include X's civic integrity policy. Additionally, X states political content ads 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."

#### YouTube

**Equally stringent.** YouTube enforces all of their Community Guidelines consistently and without regard to an ad's political viewpoint.

Ads on YouTube are subject to Google Ads policies, which ban granular microtargeting and require clear disclosures.

#### **Snapchat**

**Equally stringent.** Political advertising must comply with Snap's Terms of Service, Community Guidelines, and Advertising Policies.

#### TikTok

**N/A.** Political ads and social issue ads are not permitted.

# Are political ads eligible for 3rd party fact-checking?

No. Posts and ads from politicians are not eligible to be fact-checked. Specifically, the content that Meta states is ineligible for fact checking includes: "words a politician says, whether in text, photo, video, video clip, or any other content that is clearly labeled as created by, on behalf of, or directly quoting the politician or their campaign."

However, "When a politician shares a specific piece of content – for example, a link to an article, video or photo created by someone else that has been previously debunked on Facebook and Instagram – we will demote that content, display a warning and reject its inclusion in ads. This is different from a politician's own claim or statement".

No. X has transitioned to Community Notes for factchecking all content, which does include ads.

#### YouTube

**Unclear.** YouTube will remove content that is in violation of their election misinformation policy, including ads. It is unclear whether political ads are applicable for 3rd party fact-checking. YouTube also requires disclosures for synthetic content.

#### **Snapchat**

Yes. Political ads are fact-checked via a third-party partnership with the Poynter Institute.

#### **TikTok**

N/A.

## Are political ads required to disclose funding sources?

Yes. Meta has an authorization process and 'paid for by' disclaimer. They are then put in an Ad Library and can be accessed by researchers without an application process.

#### X

Partial. Advertisers are responsible for complying with any local laws regarding disclosure and content requirements. This applies to political content ads and political campaign ads. X relies on local laws rather than enshrining this policy into their guidelines.

#### YouTube

Yes. Google requires all US "election ads contain a disclosure identifying who has paid for the ad".

#### **Snapchat**

Yes. They must include a 'paid for' label disclosing the funding source.

## **TikTok**

N/A.

Are political ads subject to the same or more stringent content policies, specifically regarding hate and extremist content?

#### Meta

More stringent. Meta policy does not allow "Statements that suggest the inferiority of immigrants, migrants, refugees and asylum seekers, as well as expressions of contempt, dismissal or disgust directed at them" or "ads that claim a group of people on the basis of things like race, ethnicity, or religion engage in behavior that threatens the physical safety, health, or survival of others."

They also do not allow ads that would deny voting to groups of people based on protected characteristics.

"In addition to the requirements in our Community Standards, Ads can't: Use slurs in any context, even those generally allowed in the Community Standards. We require additional information and/or context to

enforce the following advertising rules. Ads can't: Claim people are threats to the safety, health, or survival of others based on their personal characteristics

Include generalizations that state inferiority, other statements of inferiority, expressions of contempt, expressions of dismissal, expressions of disgust, or cursing based on immigration status"

#### X

**Equally stringent.** All ads must comply with X Ads Policies, X's Terms of Service, and X Rules, which includes certain hate and extremist content.

#### YouTube

**Equally Stringent.** All ads must comply with YouTube's Community Guidelines and Google Ad policies.

#### **Snapchat**

**Equally Stringent.** "No content that harasses, intimidates, or threatens" is permitted, but "attack" ads which target political candidates are not "categorically" banned so long as they do not reference the target's personal life or contravene "other guidelines".

#### **TikTok**

N/A.

# Are political ads permitted to use synthetic media or misleadingly edited media?

#### Meta

Yes. According to its Advertising Standards, Meta states that all advertisers running ads across Meta technologies must follow Community Standards. The Meta Community Standards includes a policy on Misinformation which states "we may place an informative label on the face of content – or reject content submitted as an advertisement – when the content is a photorealistic image or video, or realistic sounding audio, that was digitally created or altered and creates a particularly high risk of materially deceiving the public on a matter of public importance". Meta has also, since January of this year, required "advertisers have to disclose when they digitally create or alter a political or social issue ad in certain cases"

Yes. X states that all ads must follow the X Rules, which include a Synthetic and Manipulated Media policy, however the use of Al-generated content within political ads "is not in and of itself violative" and does not require a label, as stated by X to ISD.

#### YouTube

Yes. Verified political advertisers may use synthetic media as long as it is clearly disclosed. This includes content that "that makes it appear as if a person is saying or doing something they didn't say or do".

#### **Snapchat**

No. Snapchat states that all its ads must adhere to Community Guidelines, which prohibit "manipulating content for false or misleading purposes,".

#### **TikTok**

N/A.



Is there a publicly available ads library for US ads?

#### Meta

**Yes.** The Meta Ad Library provides information about spend, reach, funding entities and more for ads about social issues, elections or politics.

**Partial.** X has a downloadable political disclosures report that is updated every two business days. However its ads repository is only available for certain regions and does not include the US.

#### YouTube

Yes. Through the Google Ads Transparency Center, Google provides "extra transparency on political ads" and requires verification of advertisers.

#### **Snapchat**

Partial. Snap has a downloadable report specific to political ads.

#### **TikTok**

No. There is no publicly available ads library for US content, only for the UK and the EU. While TikTok does not permit political ads, the EU and UK ads library has shown that this policy hasn't always been consistently enforced.

Are content creators required to disclose when political content has been sponsored, and who has sponsored the content?

#### Meta

Yes. Meta stated to ISD that "Branded content may only be posted with the use of the branded content tool, and creators must use the branded content tool to tag the featured third-party product, brand or business partner.

Branded content involving Government, elections or politics has further restrictions: Current elected and appointed government officials, political candidates, political parties and political committees may not (themselves) promote branded content. In the US, Creators (who are not among the categories in (1), may promote branded political content, only if the business partner has been authorised to run ads about social issues, elections or politics. Authorization to run ads on social issues, elections, or politics requires ads to have a disclaimer with the name and entity that paid for the ads, among additional restrictions defined in the policy".

#### X

Partial. X requires users to disclose whether the content is sponsored using the hashtag "#ad", but not who the end advertiser is.

#### YouTube

Partial. YouTube states that "You and the brands you work with are responsible for understanding and complying with local and legal obligations to disclose Paid Promotion in their content. Some of these obligations include when and how to disclose, and to whom to disclose.".

#### **Snapchat**

N/A. Snap "Snap restricts the paid promotion of political messaging to traditional ad formats".

#### **TikTok**

N/A. TikTok prohibits creators "being paid to make branded political content".

# **Hate Speech and Violent Extremism**

Is the platform putting in place prevention measures for domestic hate and extremism, such as Product interventions to disincentivize hateful behavior?

#### Meta

Partial. On Instagram, comment warnings remind people of the consequences of bullying when people attempt to post potentially offensive comments or captions, identified by Al. Domestic hate and extremism overlaps with cyber-bullying but is not limited to it.

#### YouTube

Partial. YouTube includes a feature that prompts users to reconsider posting if their comment may be offensive to others.

No.

#### **Snapchat**

No.

#### **TikTok**

Partial. Comment warnings ask users to reconsider when posting a comment which could be unkind and reminds them of the community guidelines.

In 2021 TikTok announced a partnership with the Cyberbullying Research Center (CRC), whose mission is "to develop sound research about cyberbullying and related forms of abuse/misuse". It seems that TikTok is approaching the issue through an anti-cyber bullying lens, which overlaps with domestic hate and extremism but is not limited to it.

Is the company working/partnering with external experts to support in proactively identifying hate speech and violent extremist content on its platform/s and corresponding policy gaps throughout the election cycle? Have details on the scope and results of those partnerships been publicly disclosed?

#### Meta

Yes/Unclear. Meta states: "We routinely commission independent research from think-tanks, academics and NGOs on various topics of violent extremist and

terrorist use of the internet in order to help our industry understand and make progress on these important issues. In 2017, we joined forces with YouTube, Microsoft, and Twitter to create the Global Internet Forum to Counter Terrorism (GIFCT), an organization built with the mission to prevent terrorists and violent extremists from exploiting digital platforms. GIFCT also supports the Global Network on Extremism and Technology (GNET), the academic research arm of the Global Internet Forum to Counter Terrorism (GIFCT) that aims to better understand the ways in which terrorists use technology. In 2022, Meta announced a research partnership with the Middlebury Institute of International Studies' Center on Terrorism, Extremism, and Counterterrorism to analyze evolving trends in violent extremism and effective tools that help communities combat it." However, it is not entirely clear what the scope and outputs are of these partnerships.

#### X

Yes/Unclear. X stated to ISD: "Yes, we will be working with partners to identify emerging language tied to hate speech and violent extremism for internal enforcement". However, it is not entirely clear what the scope and outputs are of these partnerships. .

#### YouTube

Yes/Unclear. Through its priority flagger program, YouTube "continue to invest in the network of over 300 government partners and NGOs who bring valuable expertise to our enforcement systems, including through our Priority Flagger program". YouTube is also a founding member of GIFCT.

Additionally, Jigsaw, a think tank created by Google, is made up of Google engineers, researchers, and policy experts working to address various issues (extremism, disinformation, cybersecurity, etc.). In January 2023, the Jigsaw team was reduced by 30-50%.

#### **Snapchat**

**Yes/Unclear.** Snapchat claim to be "engaging actively with researchers, NGOs and other stakeholders from across the election integrity, democracy, and information integrity communities" in developing their safeguarding policies.

#### **TikTok**

Yes /Unclear, TikTok stated to ISD that it "works in

close collaboration with a number of NGOs and external experts to safeguard our community against hate speech and violent extremist content. TikTok continues to rely on threat detection partners." TikTok also states it is a member of Tech Against Terrorism which "provides practical and operational support to the Trust and Safety teams tasked with preventing violent extremists from using our platform to perpetrate harm." However, it is not entirely clear what the scope and outputs are of these partnerships.

TikTok recently launched a US Elections Integrity Advisory Group "comprised of experts in elections and civic integrity, hate, violent extremism, and voter protection issues and are consulted regularly to provide insights and recommendations to optimize our elections integrity approach."

Are there policies against promotion or incitement of election violence, militarization, and calls for intimidation (i.e., bringing weapons to polling stations)?

Yes. Threats against election officials and militarized language are prohibited, as are calls for election violence, militarization or intimidation of voters.

Meta also prohibits "Coded statements where the method of violence is not clearly articulated, but the threat is veiled or implicit, as shown by the combination of both a threat signal and contextual signal."

Calls to bring weapons to high-locations, such as places of worship or schools, are also prohibited.

#### X

Yes. Users cannot threaten to damage infrastructure that is essential to civic activities under the violent speech policy. Users also cannot engage in or promote behaviors that may "coerce others to refrain from participating in a civic process." This includes inciting violence towards a polling station and vote counting locations, inciting disruption or destruction of electoral procedures, inciting others to harass poll workers or voters, promoting firearms near polling locations, and threats regarding voting locations.

#### YouTube

**Yes.** Content containing incitement to interfere with democratic processes is not allowed on the platform under the election misinformation policy. Additionally, content inciting others to commit violent acts against individuals or a defined group of people is not allowed under the violent or graphic content policies.

#### **Snapchat**

**Yes.** Intimidation and incitement to violence are prohibited.

#### **TikTok**

Yes. TikTok prohibits "Promoting or inciting violence, such as encouraging an attack or others to attack, praising a violent act, or recommending people bring weapons to a location to intimidate others."

Newsworthy content which incites people to violence is not allowed.

TikTok maintains a "zero-tolerance policy" on violent extremism, which means that violent extremists, extremist organizations including those designated by the UN, violent political organizations, hateful organizations (those which target individuals based on protected characteristics) and violent political organizations are banned and users are not allowed to promote them.

Does the platform remove (or downrank) content which includes violent threats?

## Meta

Yes. Under its Violence and Incitement policy, Meta removes threats of violence that could lead to death, serious injury, threats of violence related to voting, and more.

## X

Yes. Under its Violent Content policy, X does not allow content that includes violent threats, including damage to "infrastructure that is essential to daily, civic, religious, or business activities" and also "wishing, hoping or expressing desire for harm".

#### YouTube

Yes. Under its Harassment & Cyberbullying policy, YouTube does not allow users to threaten someone's physical safety or doxing.

#### **Snapchat**

Yes. Under its Threats, Violence & Harm policy, Snapchat prohibits intimidation or threats to a person, a group of people, or someone's property.

#### **TikTok**

Yes. Under its Violent and Criminal Behavior policy, TikTok does not allow any violent threats or incitement to violence.

Are there any measures in place to specifically protect candidates and election officials from targeted hate speech, harassment, or threats on the platforms?

Partial. Meta provides resources to outline how women running for office can stay safe across Meta platforms. However, this seems to be a tailored guide to Meta's policies and practices rather than a bespoke measure or intervention.

**No/Unclear.** ISD was unable to identify measures specifically designed to protect candidates or election officials.

#### YouTube

No/Unclear. ISD was unable to identify measures specifically designed to protect candidates or election officials

#### Snap

**No/Unclear.** ISD was unable to identify measures specifically designed to protect candidates or election officials

#### **TikTok**

**No/Unclear.** ISD was unable to identify measures specifically designed to protect candidates or election officials

Are the platform's hate speech, violence and extremism policies applied to livestreaming product features (both audio and audiovisual)?

Yes. Under Meta's Facebook Live policies and guidelines, all livestreams must comply with Meta's Community Standards.

Yes. According to X's Terms of Service, all users using X services must comply with X's Rules and Policies. Additionally, X links its Sensitive Media policy in its blog post about using X Live.

#### YouTube

**Yes.** YouTube explicitly states that its Community Guidelines apply to "all types of content" on the platform, including videos, comments, links, and thumbnails.

## **Snapchat**

**Yes.** In Snap's Terms of Service, the company states that all users using any of Snap's services must comply with the Community Guidelines, which includes hate and extremism policies.

#### **TikTok**

Yes. TikTok explicitly states that its Community Guidelines apply to "everyone and everything" on their platform.

TikTok has a comment prompt for users to reconsider unkind language, and asserts that it removes "misleading or impersonating accounts" from the platform. Accounts which incite violence are banned from livestreaming.

Are the platform's hate speech, violence and extremism policies applied to comment features?

#### Meta

Yes. Meta explicitly states its Community Standards apply to everyone and all types of content.

#### X

Yes. X's Platform Use Guidelines state that post-level enforcement includes replies and quotes.

#### YouTube

Yes. YouTube explicitly states that its Community Guidelines apply to "all types of content" on the platform, including videos, comments, links, and thumbnails.

#### **Snapchat**

Yes. Snapchat explicitly states that its Community Guidelines apply to "all content", including forms of communication such as text, images, generative Al, links or attachments, emojis, Lenses, and other creative tools or behavior on Snapchat.

#### **TikTok**

Yes. TikTok explicitly states that its Community Guidelines apply to "everyone and everything" on their platform.

Does the platform remove or downrank content that calls for or incites coordinated harassment (both online and offline)?

#### Meta

Yes. In 2021, the platform announced it launched a new policy to protect users from "mass harassment and intimidation from multiple accounts." Meta removes coordinated efforts of mass harassment that target

"individuals at heightened risk of offline harm," even if the content on its own doesn't violate policies.

#### X

Yes. According to its Abuse and Harassment policy, X prohibits targeted harassment or calling for others to harass an individual or group of people that may urge offline action.

#### YouTube

Yes. According to its Harassment & Cyberbullying policy, YouTube prohibits content that encourages "brigading," which the platform defines as "when an individual encourages the coordinated abuse of an identifiable on or off YouTube."

## **Snapchat**

Yes. Snapchat has a Harassment & Bullying policy that prohibits "bullying or harassment of any kind," Snapchat stated to ISD that "this includes any coordinated or uncoordinated harassment, online or offline.".

#### **TikTok**

Yes. According to its Harassment and Bullying policy, TikTok does not allow promotion of coordinated harassment of a person or attempting to create conflict between people.

# **Resourcing (Internal and External)**

## Is there a dedicated US based election integrity team, if so, how large is that team?

#### Meta

**No/Unclear.** Meta states "We will continue to run our Election Operation Center which brings together subject matter experts from across the company – including from our threat intelligence, data science, engineering, research, operations and legal teams - for real-time monitoring so that we can address potential abuse flowing across our network". However this appears to be a cross functional working group, as opposed to a dedicated team.

**No/Unclear.** According to Elon Musk, X's Election Integrity team no longer exists. X also stated to ISD that "This workstream falls under our Safety team which is US-based, and is supported by cross-functional members of our Elections Working Group".

#### YouTube

Yes/Unclear. Google has a Global Election Integrity team (last mentioned in 2022).

#### **Snapchat**

No/N/A. While Snapchat states "we have convened a cross-functional election integrity team, including misinformation, political advertising, and cybersecurity experts, to monitor all relevant developments in connection with elections throughout the world in 2024", this appears to be a cross functional working group, as opposed to a dedicated, permanent team.

#### **TikTok**

Yes/Unclear. TikTok states "We continue to expand and invest in our Election Integrity team, which is tasked with identifying and responding to risks posed by upcoming elections based on current and historic factors. Our team is staffed by multi-disciplinary experts in democracy, elections, civil society, and technology" TikTok further stated to ISD "we have a dedicated USbased election integrity team and expect to spend at least \$2 billion on trust and safety efforts in 2024".

# Are Spanish speakers included on that

#### Meta

**N/A.** See previous question.

#### X

**N/A.** See previous question.

#### YouTube

Unclear.

#### **Snapchat**

N/A.

#### **TikTok**

Yes. TikTok stated to ISD "we have Spanish speakers on our US-based election integrity team".

Have there been staffing cuts to election and civic integrity teams since 2022?

Yes. Meta laid off several members of teams that countered mis- and disinformation in the 2022 US midterms ahead of 2024.

Yes. X claimed it was "currently expanding" their "safety and elections teams" to focus on manipulated media, inauthenticity, and emerging threats, but appeared to cut half of the election integrity team in late September 2023.

## YouTube

Unclear. While cuts at google in 2023 have included Trust & Safety, and now upcoming cuts in 2024, is unclear if the individuals affected were working on election related harms.

#### **Snapchat**

No.

#### **TikTok**

No.

## Is there a viral content review system in place? If so, are minimum thresholds or definitions for virality disclosed?

#### Meta

Yes/Unclear. Meta stated to ISD "Our independent fact-checking partners review and rate viral misinformation, including whether it was created or edited by digital tools such as Al, and we do not allow an ad to run if it's rated as False, Altered, Partly False, or Missing Context". It is unclear at what point content is considered "viral".

Yes/Unclear. X stated to ISD that "Trending content is reviewed for potential policy violations". It is unclear what threshold content must meet to be considered trending, though X did indicate to ISD that is in part dependent on the overall volume of conversation within a region.

#### YouTube

No/Unclear. No reference could be found viral content safety functions.

## Snap

No/Unclear. Snapchat claims that its architecture prevents unvetted content from going viral. However, the platform now allows non-curated content into the 'Spotlight' feature, which features user-generated content.

## **TikTok**

Yes/Unclear. A safety check occurs "when a video reaches a certain threshold of views."

## **Transparency**

Will the platform be releasing findings on identified disinformation networks on an ongoing basis (monthly or more) throughout the election cycle?

#### Meta

No. Regularly published reports include an Adversarial Threat Report and data on Meta's identification and removal of Coordinated Inauthentic Behavior (CIB) networks. Reports can be found on Meta's Threat Disruptions page in the Transparency Centre, however these reports are released quarterly, and not immediately at the end of each quarter.

#### X

No. Twitter formerly published public datasets of influence operations which was discontinued in 2022 and replaced with a research consortium to provide the same data to researchers. However, it appears under X this initiative is no longer operational. However, X stated to ISD "Our goal is to return to a more regular cadence of global public disclosures."

#### YouTube

**No.** Google publishes a quarterly TAG (threat analysis group) bulletin which details coordinated influence operations and Google's enforcement actions, however these reports lack the detail of Meta's and are not at a consistent, real time cadence throughout the election cycle.

#### **Snapchat**

No. Snapchat does not release these types of reports, due to (as stated to ISD) "Snapchat has not identified any evidence of disinformation networks on our platform; our architecture makes the use of our platform by such networks highly inefficient".

## **TikTok**

Yes. In 2022 TikTok pledged to "publish insights about the covert influence operations we identify and remove from our platform globally" in quarterly reports. However, the reports lack in-depth information, sufficing to state the target audience of the operation, whom it portrayed in a negative or positive light, and the number of accounts in the network and its followers. In a recent elections blogpost, TikTok committed to introducing dedicated "covert influence operations

reports to further increase transparency, accountability, and sharing with the industry."

Will findings on malicious actors and/or harmful content trends relevant to civic discourse be shared on an ongoing basis (monthly or more) during the election cycle?

#### Meta

No. Meta shares findings about threats with 'industry peers' and security researchers, as well as reports on Coordinated Inauthentic Behavior (CIB) and cyber espionage operations in its quarterly Adversarial Threat Report, but these threat reports do not address violative content trends in the context of an election, only behaviors or actions from accounts or groups.

No. As mentioned above, X previously published biannual Transparency Reports highlighting data, trends, and insights into global Legal Requests, X Rules enforcement, and Security & Integrity issues. However this detailed format has not been used since December 2021, and the last report was a high-level content enforcement table published via a blog post in April 2023.

#### YouTube

No. Google publishes transparency reports that feature the amount of content removed under different policies. However, these reports do not include details of specific violative content trends relevant to civic integrity being actioned.

#### **Snapchat**

No.

#### **TikTok**

Yes. On September 4th, TikTok announced that "To bring ongoing transparency to our work, today we've launched a new US Election Integrity Hub in our Transparency Center. We'll be providing continuous updates on steps we're taking to protect TikTok during the elections, including misinformation we're taking action on". Based on the updates provided in the initial rollout, it appears these updates will be at least weekly.

Have civil society and academic researchers been provided API access for the purposes of independent election related (Includes hate speech and violent extremism) research?

#### Meta

Partial. Meta engaged in partnerships with academic researchers and provides API tools for researchers who request access. The authentication process seeks to ensure that bad actors do not access user data in the vein of the Cambridge Analytica scandal, including hostile state operations. However, researchers also reported not being able to access the APIs, and accused Meta of "trying to block research on its platform quite systematically" in a "consistent war on independent media."

Criteria for access include that researchers must "be affiliated with a qualified academic institution or a qualified research institution. Researchers from different disciplinary and professional backgrounds are welcome to apply."

Meta Content Library and Content Library API currently covers the US and most other geographies. Data which could identify users is excluded.

#### X

Partial. X ended free access to its API in February 2023 and launched paid tiers in March. The X team claimed they were "looking at new ways to continue serving" the academic community, but in the meantime, the Free tier was available for use. The Free access only allows data on 1.500 Tweets/month.

#### YouTube

**Partial.** While YouTube allows academics to apply they do not allow civil society to apply. YouTube invites eligible researchers to apply to use YouTube's API with "scaled access to YouTube's public data corpus" for a variety of research topics.

Criteria for access include that researchers are "affiliated with an accredited, higher-education institution that can grant degrees."

YouTube also allows developers access to their public YouTube Data API.

Google also awards Trust & Safety research grants open to professors at degree-granting universities and academic research institutions. It is not clear if they have funded research proposals specific to harms related to the US election.

#### **Snapchat**

No.

#### **TikTok**

Partial. TikTok provides academic researchers with API access to research the TikTok platform and TikTok's content moderation, but previously did not provide access to civil society. TikTok did however state to ISD that "We recently updated the eligibility criteria to further expand access to the Research API and are currently beta testing this service for US-based researchers in civil society."

## State Affiliated Media



## Does the platform clearly label state affiliated media at the post/video level?

#### Meta

Yes. Meta "labels state controlled media on Facebook." Instagram and Threads so that users know when content is from a publication that may be under the editorial control of a government" The label state-controlled is employed so that editorially independent but government funded journalistic institutions are not put in the same category of state propaganda arms such as RT.

#### X

Yes. X distinguishes between state affiliated media entities, government accounts and individuals associated with state affiliated media entities.

#### YouTube

Yes. YouTube labels videos that are uploaded by news broadcasters which receive some level of government or public funding.

### **Snapchat**

No.

#### **TikTok**

Yes. TikTok labels "accounts and videos".



Does the platform clearly label state affiliated media at the account level?

#### Meta

Yes.

Yes. State affiliated media accounts are also not amplified or recommended by X.

#### YouTube

No. While a label is applied under an account's individual video posts, it is not applied to the account page.

#### **Snapchat**

No.

#### TikTok

Yes. TikTok's policy is to "label accounts run by entities

whose editorial output or decision-making process is subject to control or influence by a government."

Does the platform prohibit foreign state affiliated media from advertising within the US?

Yes. Meta blocks "ads from state controlled media outlets targeting people in the US."

### X

Yes. According to X, "state-affiliated media may not purchase advertisements. This policy extends to individuals reporting on behalf of, or who are directly affiliated with such entities. State-affiliated media is defined as outlets where the state exercises control over editorial content through financial resources, direct or indirect political pressures, and/or control over production and distribution. Unlike independent media, state-affiliated media frequently use their news coverage as a means to advance a political agenda. The promotion of content from state affiliated media is also prohibited."

#### YouTube

No. There is no policy specifically prohibiting state affiliated media from advertising within the US provided their ads comply with Google's policies.

#### **Snapchat**

Partial. "Political advertisements served by Snap may not be paid for directly or indirectly by non-resident foreign nationals or entities — in other words, by people or entities that aren't residents of the country where the advertisement will run." However, there does not seem to be restrictions on non-political ads from stateaffiliated media accounts, unless from a country subject to US sanctions.

#### **TikTok**

Yes. TikTok stated to ISD "TikTok expanded our stateaffiliated media policies so that state-affiliated media accounts that advertise on our platform will only be able to advertise to audiences in the market where their parent entity is registered. This means that stateaffiliated media registered in foreign countries cannot advertise in the US."



## If yes, does the platform label such ads?

## Meta

N/A

## X

N/A

## YouTube

Yes. YouTube stated to ISD that "For all ads from verified advertisers on our platforms, we include the name and country where the advertiser is based within the My Ad Center panel, accessible via the three-dot or info icon on the ad."

## **Snapchat**

Unclear.

#### **TikTok**

N/A



Amman | Berlin | London | Paris | Washington DC

Copyright © Institute for Strategic Dialogue (2024). ISD-US is a non-profit corporation with 501(c)(3) status registered in the District of Columbia with tax identification number 27-1282489. Details of the Board of Directors can be found at www.isdglobal.org/isd-board. All Rights Reserved.