+accountabletech

Democracy by Design: Social Media's Policy Scores

EXECUTIVE SUMMARY

An estimated 2 billion voters around the world will head to the polls this year to cast their ballots amidst groundbreaking technological change. Social media companies have a responsibility to protect the integrity of our elections and ensure that their platforms are not harnessed by bad actors to undermine trust in elections. But with so much at stake, are they ready?

To answer that question, Accountable Tech researchers scored the election preparedness of ten leading online platforms: Facebook, Instagram, Threads, YouTube, TikTok, Snapchat, Discord, LinkedIn, Nextdoor, and X (formerly Twitter). This scorecard measures to what extent these platforms' policies meet recommendations made in Democracy By Design roadmap – actionable, high-impact, and content-agnostic steps to protect the integrity of elections.

Unfortunately, the research paints a grim picture. **With 2024 elections well underway, all ten platforms are failing on election preparedness based on our analysis.**

Specific findings include:

- **Out of a possible 100% match to the Democracy By Design recommendations, no platform scores above 62%.** Nextdoor performs the worst in our analysis, with a 17% preparedness score.
- **Insufficient guardrails to stop the spread of manipulated content depicting public figures, like deepfakes:** Just 20% of platforms – TikTok andSnapchat – have policies on the books that would prohibit deceptively manipulated media of public figures. That means that the vast majority of these platforms do not prohibit deepfake videos or manipulated images depicting false campaign events, candidates spewing false information, or other types of deceptive depictions of candidates for office, election officials or other government figures, which we're already seeing proliferate as elections heat up.
- **Platform features enable AI-generated political ads to be micro-targeted to voters:** Nearly every social media platform which allows political advertising does not explicitly prohibit AI-generated ads from being micro-targeted to voters.
- **Lack of transparency on performance and engagement related to election-related posts:** No platform provides transparent access to data related to the highest-performing and highest-engagement election-related posts, advertisements, accounts, URLs, and groups. That means that voters, independent researchers, and election officials are left in the dark about how election-related information spreads across platforms.
- **Insufficient "friction" to stop the spread of misleading election information:** A majority of platforms do not have policies in place to put posts that contain misleading or unverified election information behind click-through warning labels that include clear context and fact. Without these labels, election misinformation is able to be spread more quickly and magnify threats. Of the platforms with algorithmic recommendation feeds, only Snapchat deploys a viral circuit breaker to stop the fast-spread of posts before they can be reviewed. None of the platforms with public feeds limit sharing after multiple reshares, a simple content-neutral design measure that could significantly mitigate against virality of potentially harmful information.
- **A lack of transparency, including an opacity around policy enforcement and safety teams:** Some platforms, like Meta, which have previously come under intense scrutiny for their role in amplifying the spread of electoral disinformation narratives, have numerous policies, but it's impossible to know how they are being enforced. Platforms have wide latitude when it comes to enforcement, and there is reason for skepticism that they meaningfully follow through. This is made more concerning because of industry-wide layoffs and cuts to election integrity safety teams – including the complete dismantling of X's election integrity team.

SCORECARD AND METHODOLOGY

The following scorecard was developed by Accountable Tech's team of researchers who analyzed the publicly available policies and design of Facebook, Instagram, Threads, YouTube, TikTok, Snapchat, Discord, LinkedIn, Nextdoor, and X (formerly Twitter) to judge their election preparedness in alignment with the three major planks of the Democracy by Design framework:

- **Bolstering resilience,** which focuses on 'soft interventions' that introduce targeted friction and context to mitigate harm;
- **Countering election manipulation,** which outlines bulwarks against evolving threats posed by malign actors and automated systems; and
- **Paper trails,** which highlights key transparency measures needed to assess systemic threats, evaluate the efficacy of interventions, and foster trust.

This scorecard tracks the policies of these ten platforms as of February 16, 2024. The policies of each platform were analyzed and assigned a score for their compliance with each of the recommendations outlined in the Democracy By Design roadmap. Platforms that had relevant and sufficient policies in place for each recommendation received 4 points, platforms with partial but insufficient policies in place received 2 points, and platforms with no policies in place corresponding to Democracy By Design's recommendations received 0 points. Because both the core function of these platforms and the level of publicly available information about their policies varies, "not applicable" classifications were assigned to account for platform-specific nuances (see each category, below).

In some cases, Meta did not specify whether or not their policies were specific to Instagram. For certain policies that specifically mention Instagram and Threads alongside Facebook, this report assumes that other policies refer to Facebook.

The scores calculated below reflect the percent total of the points platforms earned for election preparedness divided by the total amount of points possible.

ELECTION PREPAREDNESS SCORECARD

How do platforms' policies stack up in terms of election preparedness?

Category	Facebook	Instagram	Threads	YouTube	TikTok	Snapchat	Discord	Linkedin	Nextdoor	X
Bolstering resilience	57.14%	28.57%	14.29%	28.57%	57.14%	58.337%	14.29%	14.29%	7.14%	35.71%
Countering Election Manipulation	50%	50%	62.50%	60%	75%	50%	25%	40%	50%	30%
Paper trails	50%	33.33%	8.33%	16.67%	58.33%	16.67%	16.67%	16.67%	0%	0%
Total	**52.78%**	**36.11%**	**23.53%**	**33.33%**	**61.76%**	**41.18%**	**17.65%**	**22.22%**	**16.67%**	**22.22%**

It's worth noting that some platforms have gone to great lengths to address some of the planks outlined in the scorecard – like Meta's fact-checking labels – but fail to adequately address others. While some of these platforms may consider the recommendations outlined in the Democracy By Design roadmap to be too high a bar to clear, these recommendations should be considered the bare-minimum effort for mitigating the novel threats that elections in 2024 face. What's more, these recommendations are rooted in platforms' own product design and policy toolkits, and in many cases, empirically backed by their own research.

We hope each of these platforms implement changes to their policies and designs to avoid political landmines and broadly resonate across nations with distinct laws and cultures, and platforms with incongruous architecture and resources – a consensus roadmap to enhance systemic resilience against election threats.

BOLSTERING SYSTEM RESILIENCE

Targeted Friction and Context to Mitigate Threats

Election integrity vulnerabilities on social platforms often stem from their own architecture: The features designed to make platforms frictionless and maximize engagement – from recommendation algorithms to reshare buttons – can serve to warp discourse and undermine democracy.

Extensive research from experts and tech companies themselves indicates that content-agnostic soft interventions that introduce targeted friction and context – bolstering the resilience of their systems and better informing users – can significantly mitigate threats.

Bolstering System Resilience	Meta	YouTube	TikTok	Snapchat	Discord	LinkedIn	Nextdoor	X
1a. Use pop-ups asking users if they want to read an article before sharing, or alerting users if articles are old	Facebook: Yes Instagram: No Threads: No	Partial	Partial	N/A	No	No	No	Partial
1b. Clearly label accounts of government officials and use interstitials to alert users who try to read or share content from state-run media	Facebook: Partial Instagram: Partial Threads: Partial	Partial	Yes	No	No	No	No	Partial
1c. Place posts that contain misleading or unverified election information behind click-through warning labels that include clear context and fact	Facebook: Yes Instagram: No Threads: No	No	Partial	Partial	Partial	Partial	No	Partial
1d. Append distinct verification badges to the official accounts of local, state, and national election authorities	Facebook: Partial Instagram: Partial Threads: Partial	No	Partial	Partial	No	No	No	Partial
1e. Utilize virality circuit breakers to automatically flag fast-spreading posts and trigger a brief halt on algorithmic amplification	Facebook: No Instagram: No Threads: No	No	Partial	Yes	No	No	No	No
1f. Restrict rampant resharing during election season by removing simple share buttons on posts after multiple levels of sharing	Facebook: No Instagram: No Threads: No	No	No	Partial	No	No	No	No
1g. Implement clear strike systems to deter repeat offenses, curtail the outsized impact of malign actors, and better inform users	Facebook: Yes Instagram: Yes Threads: No	Yes	Yes	Yes	Partial	Partial	Partial	Partial
Total	Facebook: **57.14%** Instagram: **28.57%** Threads: **14.29%**	**28.57%**	**57.14%**	**58.33%**	**14.29%**	**14.29%**	**7.14%**	**35.71%**

COUNTERING ELECTION MANIPULATION

Safeguards Against Malign Actors and Automated Systems

Malign actors have weaponized social platforms to meddle in elections, attack democracy, and erode our shared reality. Now their capacity for manipulation has been turbocharged by new technology, including powerful algorithms and generative AI tools tailormade for high-impact, low-cost influence operations.

Platforms must do everything in their power to thwart unlawful efforts to interfere with elections or individuals' free exercise of their right to vote – including efforts to intimidate voters or mislead them on how to participate – and to counter manipulation more broadly.

Countering Election Manipulation	Meta	YouTube	TikTok	Snapchat	Discord	LinkedIn	Nextdoor	X
2a. Prohibit the use of generative AI or manipulated media to falsely depict election irregularities.	Facebook: Partial Instagram: Partial Threads: Partial	Partial	Yes	Yes	No	Partial	Partial	Partial
2b. Prohibit the use of generative AI or manipulated media to fraudulently misrepresent the speech or actions of public figures in video, audio, or images.	Facebook: Partial Instagram: Partial Threads: Partial	Partial	Yes	Yes	No	Partial	Partial	Partial
2c. Prohibit the use of generative AI or manipulated media to create personalized political or issue ads that microtarget voters with distinct content generated by using their personal data.	Facebook: No Instagram: No Threads: Yes	Yes	Yes	Partial	Yes	Yes	Yes	No
2d. Implement strong provenance standards and require clear disclosures within any political or issue ad that features AI-generated images, video, or audio.	Facebook: Yes Instagram: Yes Threads: N/A	Yes	N/A	No	N/A	No	No	No
2e. Prompt users as election season starts, outlining the main parameters of core algorithmic systems and giving users autonomy to adjust ranking criteria so their recommendations are not based on profiling or optimized for engagement.	Facebook: Partial Instagram: Partial Threads: Partial	No	No	No	No	No	Partial	Partial
Total	Facebook: **50%** Instagram: **50%** Threads: **62.50%**	**60%**	**75%**	**50%**	**25%**	**40%**	**50%**	**30%**

PAPER TRAILS

Fostering Trust through Meaningful Transparency

Just as machine voting systems are backed by paper trails to ensure our elections are trustworthy and secure, platforms that shape the information ecosystem must finally open up the black box and start showing their work. The status quo has yielded distrust from all sides, with partisans suspecting foul play, neutral observers unable to make informed evaluations, and non-English speakers left behind.

Platforms must begin disclosing meaningful data to voters, independent researchers, and election officials regularly in order to engender trust and substantiate integrity measures.

Paper Trails	Meta	YouTube	TikTok	Snapchat	Discord	LinkedIn	Nextdoor	X
3a. Clearly and accessibly detail in one place all election-related (or applicable platform-wide) policies and approaches.	Facebook: **Partial** Instagram: **No** Threads: **No**	**Partial**	**Partial**	**Partial**	**No**	**No**	**No**	**No**
3b. Release regular transparency reports during election season – broken down across widely spoken languages – detailing high-performing and violative content, enforcement, and resource allocation.	Facebook: **Partial** Instagram: **Partial** Threads: **No**	**Partial**	**Partial**	**Partial**	**Partial**	**Partial**	**No**	**No**
3c. A snapshot of the week's highest-performing election-related content – the posts, advertisements, accounts, URLs, and groups that generated the most engagement.	Facebook: **Partial** Instagram: **No** Threads: **No**	**No**	**No**	**No**	**No**	**No**	**No**	**No**
3d. Aggregate data about enforcement of all election-related policies, including specifics about the measures detailed in previous sections and their effectiveness.	Facebook: **Partial** Instagram: **Partial** Threads: **No**	**No**	**Yes**	**No**	**Partial**	**No**	**No**	**No**
3e. Details about the number of full-time staff and contractors working on election integrity, broken down by department role and primary languages spoken.	Facebook: **Partial** Instagram: **Partial** Threads: **Partial**	**No**	**Partial**	**No**	**No**	**No**	**No**	**No**
3f. Provide independent researchers with direct access to platform data to inform studies, threat analysis, and systemic impact assessments.	Facebook: **Partial** Instagram: **Partial** Threads: **No**	**No**	**Yes**	**No**	**No**	**Partial**	**No**	**No**
Total	Facebook: **50%** Instagram: **33.33%** Threads: **8.33%**	**16.67%**	**58.33%**	**16.67%**	**16.67%**	**16.67%**	**0%**	**0%**

BOLSTERING SYSTEM RESILIENCE

Targeted Friction and Context to Mitigate Threats

1a. Use pop-ups asking users if they want to read an article before sharing, or alerting users if articles are old

Facebook: Yes

- Facebook has reportedly tested out interstitials that use article sharing prompts, date indicators, or other warnings. Currently its policies include the use of warning screens for sensitive and graphic material, viral challenges that might be harmful, and for deceptive manipulated media. If someone shares an article more than 90 days old on Facebook, they will see a notification of the article's age.

Instagram: No

- Meta's policies mention Instagram warning screens for "sensitive and graphic material", but notedly do not include Instagram when referring to notifications for old articles. Instagram's policies do not include anything about interstitials related to old articles or prompts to read something before sharing.

Threads: No

- Threads is supposedly included in Instagram's policies, however nothing regarding interstitials could be found in Threads policies.

YouTube: Partial

- YouTube adds context from machine learning systems to prioritize authoritative sources around videos "where accuracy and authoritativeness are key, including news, politics, medical, and scientific information." Most of its examples apply to health misinformation or non-political conspiracy theories. It does not appear to alert users to the date of old videos.

TikTok: Partial

- TikTok introduced pop-ups for unverified content in 2021 and lives up to the intent of this recommendation by adding "warning labels to content related to unfolding or emergency events which have been assessed by our fact-checkers but cannot be verified as accurate." These labels prompt users to reconsider sharing with "cancel" being the easier option. Presumably this would include outdated videos circulating in the context of current "emergency events." In January of 2024, they reiterated this: "When content is unverified, we may label it, reduce its reach by making it ineligible for For You feeds, and prompt people to reconsider sharing it until there's more information."

Snapchat: N/A

- Snapchat does not have an open news feed, and as primarily a video platform it does not give users the option of sharing articles. Snapchat says that it employs AI review as a first level of content moderation for every public post on Spotlight, and that there is human review once a piece of content starts to gain viewership.

Discord: No

- Discord does have pop-ups that prompt users before they share content. In their community guidelines, the platform asks users to not share certain information, but they don't place any warning labels on the content.

LinkedIn: No

- LinkedIn does not have pop-ups asking users if they want to share information before they do so. They only encourage users in their policies to not share content they are skeptical of.

Nextdoor: No

- Nextdoor allows campaign news and stories about national elections only in groups, but does not specify that they have any interstitial on old news.

X: Partial

- X has a program called "Community Notes," which allows users (not employees) to append additional information to posts on the platform. X says that they do not write, rate, or moderate notes. It is possible that outdated articles could draw a community note to point out the date, but it's far from guaranteed.

1b. Clearly label accounts of foreign governments using interstitials to alert users who try to read or share content from state-run media

Facebook: Partial

- Meta does label state-controlled media, however it does not appear to alert users if they share content from those pages.

Instagram: Partial

- Meta's state-controlled media policies include Instagram.

Threads: Partial

- Threads is included in Instagram's state-controlled media policy, which also does not appear to alert users sharing content from such an account.

YouTube: Partial

- YouTube "may" append an information panel to the watch page of channels that are "owned by a news publisher that is funded by a government, or publicly funded." However, while the platform includes an information panel about publishers under videos, which includes a link to the publisher's Wikipedia page, it does not appear to alert users who share such videos.

TikTok: Yes

- TikTok's state-affiliated media policy is comprehensive. The platform'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."

Snapchat: No

- Snapchat does not make mention of warnings or labeling on accounts run by state-media.

Discord: No

- Discord's policies do not mention warnings or labeling on accounts run by state-media.

LinkedIn: No

- LinkedIn's policies do not mention warnings or labeling on accounts run by state-media.

Nextdoor: No

- The way that Nextdoor is designed should make it more difficult for state-run media to set up accounts, but the platform does not appear to have any protections in place to inform users should information from a state-media source be shared on the platform.

X: Partial

- X has a government-media policy, which says labels appear on government, state-run media, and the relevant accounts for individuals associated with those entities from "China, France, Russia, the United Kingdom, the United States, Canada, Germany, Italy, Japan, Cuba, Ecuador, Egypt, Honduras, Indonesia, Iran, Saudi Arabia, Serbia, Spain, Thailand, Turkey, and the United Arab Emirates." They say other countries will be included in the future. The policy does not include the use of interstitials.

1c. Place posts that contain misleading or unverified election information behind click-through warning labels that include clear context and fact

Facebook: Yes

- Facebook applies the strongest warning label to "content rated False or Altered" by their fact-checkers. A 2022 policy update also says that people who post content that earns a fact-checking label are notified.

Instagram: No

- While Meta's policies for warning screens include "misleading content" for Facebook, on Instagram it appears that only content that is "sensitive or graphic" in nature is placed behind a warning screen.

Threads: No

- Threads does not appear in Meta's labeling policies.

YouTube: No

- YouTube's electoral misinformation policy states that content that aims to "mislead voters about the time, place, means, or eligibility requirements for voting; 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" are disallowed. A blog post in December of 2023 states, "Content that misleads voters on how to vote or encourages interference in the democratic process is not allowed on YouTube. And we quickly remove content that incites violence, encourages hatred, promotes harmful conspiracy theories, or threatens election workers." The platform encourages users to report videos that violate this policy. It will remove such videos and issue a strike against the creator. However, YouTube rolled back its policy around false claims related to the 2020 Presidential Election in June of 2023 and will no longer remove videos that advance such disinformation narratives, nor does it provide a warning label on such content.

TikTok: Partial

- TikTok adds "warning labels to content related to unfolding or emergency events which have been assessed by our fact-checkers but cannot be verified as accurate." These labels appear for users watching a video, to the creator, and in a clearer way when someone tries to share such a video. However, these labels do not appear to include clear context and facts.

Snapchat: Partial

- Snapchat does not place click-through warning labels. Instead, they vet all content before it is published publicly. Any unverified information spread on Snapchat would only be able to be shared directly from one individual to another.

Discord: Partial

- It is unclear whether Discord applies warning labels to election misinformation in particular, but there have been reports that Discord tags certain users engaged in what they deem to be suspicious activity. Related, Discord's policy states that they will "remove harmful misinformation about civic processes when [they] become aware of it," but does not outline any policies on warning labels.

LinkedIn: Partial

- LinkedIn policy against content that "interfere[s] with or improperly influence[s] an election or other civic process," and says "depending on the severity of violation, we may limit the visibility of certain content, label it, or remove it entirely." It's unclear when a label would be applied and what context it would include.

Nextdoor: No

- Nextdoor says it "technology and member reports to identify and remove content, as appropriate, that violates our election misinformation policy," on its Election FAQ page, however the link for election misinformation policy directs to a page called "Discuss important topics in the right place," which does not mention the words "misinformation" or "disinformation" anywhere on it. There does exist a page - found via Google search that does outline four types of electoral misinformation that are disallowed, however there do not appear to be any interstitials to inform users.

X: Partial

- X says it will label "disputed or misleading information," and if the content in question "could lead to harm" it may warrant additional context. It's possible that the Community Notes program could also add that context but not guaranteed. The platform also has a notable exception in "certain and rare instances" for leaving violative content up from an elected or government official if X determines it is in the public interest. In that case, the content would be placed behind an interstitial that states that the post violates X Rules, but it would not include factual context.

1d. Append distinct verification badges to the official accounts of local, state, and national election authorities and to elected officials and candidates for office

Facebook: Partial

- Meta has a paid verification program called Meta Verified for the accounts of public figures, celebrities and global brands, which provides a verification badge and certain benefits to paid subscribers. Separately, there is a legacy verification program to certify authenticity of pages and notable accounts. It is not specific for elected officials or election authorities, but in 2020 Facebook did proactively work to protect the pages of candidates and their campaigns.

Instagram: Partial

- Meta's policies about verification specify Instagram. Instagram's policies detail a separate verification process for its platform, but does not name election authorities, elected officials or candidates explicitly.

Threads: Partial

- There is no direct link from meta.com for Threads policies. Existing verification badges will transfer to Threads, however the paid Meta Verified subscription does not carry over to Threads.

YouTube: No

- YouTube has a policy to verify "the official channel of a creator, artist, company, or public figure" with more than 100,000 subscribers, but it doesn't differentiate accounts of election authorities, many of which are unlikely to have 100,000 subscribers.

TikTok: Partial

- During the 2022 elections, TikTok trialed requiring government accounts to be verified. TikTok explicitly includes election authorities as government entities that can be verified, which stands out from other platforms. However, it appears that verification is now voluntary and so government entities would need to go through the process in order to attain a badge.

Snapchat: Partial

- Snapchat does not detail any policies to specifically allow for verification badges for election authorities. It uses a verification system for other types of influencers and brands that appears to also apply to political candidate accounts.

Discord: No

- Discord does not add any verification badges to profiles for election authorities, elected officials or candidate accounts. Discord paused its "Verified Server Applications" as of July 2023, which indicated a server's affiliation with a business, brand, or figure of public interest.

LinkedIn: No

- Government officials can verify their accounts by providing ID to get a check-mark next to their account. LinkedIn also has a policy to verify accounts of individuals to "provide authenticity signals to others that you're who you say you are". However, they do not have a specific policy towards election authorities, nor is it clear if LinkedIn automatically prompts government officials to verify their accounts or if they must do so themselves.

Nextdoor: No

- Local candidates are allowed to introduce themselves in the main feed, but it's unspecified if they are labeled as such in other spaces such as on groups.

X: Partial

- The platform can assign a grey checkmark to government accounts and officials, but they must be requested – they are not proactively granted. The platform also gives a blue checkmark to its paid subscribers, which has confused users as it was previously only given to accounts the platform determined were authentic, notable and active.

1e. Utilize virality circuit breakers to automatically flag fast-spreading posts and trigger a brief halt on algorithmic amplification

Facebook: No

- It is unclear whether they are using virality circuit breakers, as there were previous reports of circuit breakers being tested within the company but have not stated such in its policies.

Instagram: No

- Meta has third-party fact checkers to identify false information, altered content or content with missing context, and reduce visibility on its various feeds and stories, but they do not appear to have a content neutral mechanism to break virality for fast-spreading posts while they can be reviewed.

Threads: No

- Presumably Threads' recommendation algorithm is similar to Instagram's. Nothing is specified in Meta's policies about a content-neutral viral circuit breaker.

YouTube: No

- YouTube's recommendation engine is driven by a number of factors including an evaluation of authoritativeness, which is assessed by human reviewers and determines how likely a video is to contain harmful misinformation, or comes close to violating YouTube's policies (called "borderline" content). However, those human evaluations are then used to train the recommendation system and scaled to all videos across YouTube. There is no content-neutral viral circuit breaker.

TikTok: Partial

- In describing how the recommendation engine works, TikTok mentions that videos that have just been uploaded "may not be eligible" for recommendation. In a separate post from 2021, TikTok says that "Sometimes fact checks are inconclusive or content is not able to be confirmed, especially during unfolding events. In these cases, a video may become ineligible for recommendation into anyone's For You feed to limit the spread of potentially misleading information." While this policy could be more specific and does seem to apply to specific types of content, it does indicate that TikTok has safeguards in place to break virality during crisis situations.

Snapchat: Yes

- Snapchat deploys human review "once a piece of content gains more viewership" and stops it from reaching a bigger audience until it has been approved. Snap emphasizes that "Across our app, we don't allow unvetted content the opportunity to 'go viral.'"

Discord: No

- Discord has policies for content moderators to manually remove posts, but they don't appear to have any systems that automatically pause amplification while posts can be reviewed.

LinkedIn: No

- LinkedIn states they use AI systems and developed "proactive and reactive models" to identify violative content on the platform, but it does not have a content-neutral method of pausing viral amplification.

Nextdoor: No

- Nothing in Nextdoor's policies indicates there is any sort of viral circuit breaker, but given the nature of the platform's recommendation system and its sharing tools, trend-type recommendations may not have the same power as they do on other platforms.

X: No

- Nothing in X's policies indicates the use of a virality circuit breaker, and trending topics on the platform can amplify trend-type recommendations before violative content is flagged for review. While the platform's enforcement options for violative content include limiting visibility and restricting discoverability by removing violative posts from trends and recommendations, and the synthetic and manipulated media policy indicates that there might be visibility restrictions and recommendation pauses for borderline content as well, it appears that would only apply after content has been reviewed for potential violation of that specific policy.

1f. Restrict rampant resharing during election season by removing simple share buttons on posts after multiple levels of sharing

Facebook: No

- Facebook does not specify any restrictions on resharing during election season in its policies, although it does in some cases append fact-check labels to shared posts.

Instagram: No

- Instagram is included in Meta's enforcement of fact checker policy, but does not appear to apply content-neutral sharing restrictions.

Threads: No

- Threads share policies are not clear from Instagram's help center.

YouTube: No

- Nothing in YouTube's policies indicates the share button would be removed from borderline content. However, unlike many other platforms where the intent of this recommendation is aimed to limit sharing viral content that is being amplified on the same platform, YouTube's share tools are designed to share to other platforms.

TikTok: No

- TikTok does not appear to remove the share function, but it does append warning labels to content and has "cancel" as the default option when such labels appear. Furthermore, the platform indicates that this design has proven to reduce shares by 24%.

Snapchat: Partial

- Snapchat appears to keep the share button available on public videos regardless of their reach, but share options are only for direct messages.

Discord: No

- Discord doesn't have any policies to remove rampant resharing, but they do claim to "remove harmful misinformation about civic processes when [they] become aware of it. [They] will determine the truthfulness of claims by looking at independent, third-party sites like PolitiFact, FactCheck.org, and Snopes."

LinkedIn: No

- LinkedIn's policies do not specify any restrictions on resharing during election season. However, does warn users that they're leaving the platform when they click on external links.

Nextdoor: No

- Nextdoor does allow for reposting, which seems intended largely for users to repost their own posts or share to other social media platforms. It does not indicate it would restrict this option during election season.

X: No

- X does not appear to remove share buttons after several levels of resharing, and its visibility-reduction features are limited to content which has already been reviewed by a content moderator.

1g. Implement clear strike systems to deter repeat offenses, curtail the outsized impact of malign actors, and better inform users

Facebook: Yes

- A detailed strike system is listed on Meta's support page, escalating for malign actors and repeat offenders. It also includes a policy regarding restrictions for "accounts of public figures during civil unrest."

Instagram: Yes

- Meta's includes Instagram in its "restricting accounts" policies.

Threads: No

- Presumably Threads accounts, which are connected to Instagram, would be affected by strikes on Instagram, including accounts being disabled, however it is not specified in Instagram's policies.

YouTube: Yes

- YouTube has a three-strike system for violations of the platform's Community Guidelines. The first time a channel uploads violative content, there is a warning with no penalty. After the second strike, the account will have temporary restrictions for one week. Channels that receive three strikes in a 90 day period will be deleted. The exception to the three-strike rule is for channels that have a single severe abuse of the platform rules or are dedicated to violating YouTube's policies, which are immediately terminated.

TikTok: Yes

- TikTok details a clear enforcement system which includes strikes based on the level of harm. They also make clear to creators what policies have been violated, and where their account stands.

Snapchat: Yes

- Snapchat details a three step system for violative content: "Every strike is accompanied by a notice to the Snapchatter; if a Snapchatter accrues too many strikes over a defined period of time, their account will be disabled."

Discord: Partial

- Discord used to have a three-strike system for repeat offenders, but this October they updated their "warning system" policy to be more expansive: "Discord's violations are not strikes and there is no simple formula from number of violations to specific penalties. We weigh the severity and context of each violation and we look at a user's history of past violations when calculating the user's account standing." The platform reserves the right to place permanent suspensions for severe harms.

LinkedIn: Partial

- LinkedIn only has broad policies to deter folks who violate LinkedIn's community policies: "Depending on the severity of violation, we may limit the visibility of certain content, label it, or remove it entirely. Repeated or egregious offenses will result in account restriction." There is nothing clearly labeling about a strike system or the way in which LinkedIn penalizes users.

Nextdoor: Partial

- Nextdoor does disable accounts for violating its community guidelines, and it does give users whose accounts have been disabled the opportunity to appeal, but it's not clear if they receive a warning before the account is disabled, or on what basis their accounts are restored – and if repeat offenses result in permanent disabling of the account.

X: Partial

- X's enforcement actions become stronger for repeat violations of policies, but outside of post-specific actions, there does not appear to be clear indicators for accounts who might incur a strike.

COUNTERING ELECTION MANIPULATION

Safeguards Against Malign Actors and Automated Systems

2a. Prohibit - in policy and practice - the use of generative AI or manipulated media to: falsely depict election irregularities, or otherwise intentionally undermine faith in the process or results.

Facebook: Partial

- Meta states that they will remove any content on Facebook that violates their Community Standards, whether or not it was created by AI. Meta has a detailed list of "Voter or Census Interference" content they would remove, but they do not explicitly include manipulated media that would falsely depict election irregularities unless it directly affects current voting conditions. Meta's Manipulated Media policy was recently criticized by the Oversight Board, which found the policy is "lacking in persuasive justification, is incoherent and confusing to users, and fails to clearly specify the harms it is seeking to prevent.

Instagram: Partial

- On February 6, Meta announced that Facebook, Instagram and Threads will all label AI-generated content when industry standards can detect it. Unless it fits the very specific definitions of the existing Manipulated Media policy (it depicts someone saying something they did not and was created by AI and appears authentic), a deepfake depicting election irregularities would not be removed from Facebook or Instagram.

Threads: Partial

- As the labeling rules apply to Threads, we are operating under the assumption that Meta's Manipulated Media policy applies to Threads as well.

YouTube: Partial

- On November 14, Google announced that YouTube creators will 'soon' be required to disclose if they use synthetic manipulation or AI in a video. However, there is nothing specific about the false depiction of election irregularities in their new policy.

TikTok: Yes

- TikTok's Election Integrity policies explicitly state that content including "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; content that misrepresents the date of an election; attempts to intimidate voters or suppress voting; and more" would be removed. Further, their Synthetic and Manipulated media policy explicitly disallows material that may mislead a person about real-world events and specifies that synthetic media of public figures that violates any other policy is not allowed. They do not allow for deceptive manipulated media, including "AI-generated content that contains the likeness of a public figure if the content is used for endorsements or violates any other policy."

Snapchat: Yes

- Snapchat's 2024 Elections Plan says: "Our Community Guidelines, which apply equally to all Snapchat accounts, have always prohibited the spread of misinformation and purposefully misleading content, like deepfakes — including content that undermines the integrity of elections."

Discord: No

- Discord does not have any specific policies related to generative AI and election information, although the platform's policy does state that they will "remove harmful misinformation about civic processes when we become aware of it."

LinkedIn: Partial

- LinkedIn does not explicitly prohibit the use of generative AI or manipulated media to falsely depict election irregularities or undermine faith in the process or results. However, a 2023 blog post on their "Responsible AI Principles" underscores the platform is dedicated to upholding trust, providing transparency, and embracing

accountability when it comes to AI. LinkedIn's "False or misleading content" policy does state that LinkedIn will remove content that is "demonstrably false or substantially misleading and likely to cause harm."

Nextdoor: Partial

- Nextdoor's electoral misinformation policy prohibits "False or misleading claims about the results of an election that could lead to interference with the election process." Presumably this extends to generative AI posts that falsely depict election irregularities, but the stipulation that the claims could lead to interference with the election process lacks clarity.

X: Partial

- If X can "reliably" determine that media has been "substantially and deceptively" edited, posts containing such media that are likely to cause serious harm – including widespread civil unrest and voter suppression or intimidation – should be removed under their policy, but the platform's caveats around widespread harm leave room for interpretation on how manipulated media falsely depicting election irregularities might be handled.

2b. Prohibit - in policy and practice - the use of generative AI or manipulated media to: fraudulently misrepresent the speech or actions of public figures in video, audio, or images.

Facebook: Partial

- Meta states that it will remove any content on Facebook that violates their Community Standards, whether or not it was created by AI or manipulated media. However, Meta's Manipulated Media policy notably does not include false depictions of public figures – only their speech. Ahead of the 2022 elections, Meta announced that "videos that are manipulated in ways that would not be apparent to an average person" would be subject to removal.

Instagram: Partial

- Instagram does not have their own policies that are specific to misrepresentation of speech or actions of public figures, whether or not it is generated by AI or manipulated media, we are operating under the assumption that Meta's Manipulated Media policy applies to Threads as well.

Threads: Partial

- As the labeling rules apply to Threads, we are operating under the assumption that Meta's Manipulated Media policy applies to Threads as well.

YouTube: Partial

- On November 14, Google announced that YouTube creators will 'soon' be required to disclose if they use synthetic manipulation or AI in a video. However, there is nothing specific to prohibit falsely depicting public figures so long as there is disclosure the content is synthetic.

TikTok: Yes

- TikTok explicitly prohibits AI-generated content featuring public figures used for political purposes.

Snapchat: Yes

- Snapchat's Approach to Preventing the Spread of False Information includes a reference to the 2020 policy specifically prohibiting deepfakes.

Discord: No

- Discord does not have any policies related to generative AI and election information. While the platform does "prohibit users from misrepresenting their identity on our platform in a deceptive or harmful way" this policy appears to be directed at fake accounts rather than fake content.

LinkedIn: Partial

- LinkedIn does not explicitly prohibit the use of generative AI or manipulated media to fraudulently misrepresent the speech or actions of public figures in video, audio, or images. However, a 2023 blog post on its "Responsible AI Principles" underscores the platform is dedicated to upholding trust, providing transparency, and embracing

accountability when it comes to AI. LinkedIn's "False or misleading content" policy does state that LinkedIn will remove content that is "demonstrably false or substantially misleading and likely to cause harm."

Nextdoor: Partial

- Nextdoor's misinformation policy specifies that it will remove disinformation about candidates if it falls into one of the following two categories: "Doctored or fake tweets, quotes, images, or other material from a candidate that is designed to make that candidate seem unfit for office or "Debunked claims about a candidate's citizenship, criminal history or activity, or other personal history designed to make that candidate seem unfit or ineligible for office." Presumably doctored material "from" a candidate would include doctored material falsely depicting a candidate, but again the stipulation about the intent of its design leaves wiggle room that could be clearer in this policy.

X: Partial

- X's manipulated media policy states that content that includes "media depicting a real person [that] have been fabricated or simulated, especially through use of artificial intelligence algorithms" may be labeled or removed. However, the platform will not label potentially manipulated media that it cannot reliably determine has been altered, and there are no specific carve outs for public figures.

2c. Prohibit - in policy and practice - the use of generative AI or manipulated media to: Create personalized political or issue ads, microtargeting voters with distinct content generated by using their personal data

Facebook: No

- Meta's election integrity policies state that advertisers are now required to disclose whether "they use AI or digital methods to create or alter a political or social issue ad in certain cases." This however, does not mean that Meta will be prohibiting political ads that use personal data to microtarget voters with distinct content generated by AI or manipulated media. As an enforcement mechanism, Meta says they will remove political ads that use but do not disclose use of AI.

Instagram: No

- Instagram does not have policies regarding AI or manipulated media. It can be assumed that Meta's political ad policies would apply to Instagram, as they share the same ad management platform and policies.

Threads: Yes

- Threads does not host advertisers at the moment.

YouTube: Yes

- Google only allows the following criteria for targeting election ads: "Geographic location (except radius around a location), allowed age, gender, contextual targeting options such as: ad placements, topics, keywords against sites, apps, pages and videos." Since microtargeting isn't allowed, manipulated media in ads cannot reach very small audiences.

TikTok: Yes

- TikTok does not allow paid political ads.

Snapchat: Partial

- Snapchat's political advertising policies do not allow the deceptive use of AI, and while they require a paid for disclosure, they do not appear to require advertisers to disclose AI generated ads. Snap does require human approval of all political ads so deceptive ads would not be placed, but AI-generated ads that otherwise meet Snap's advertising guidelines would be. It's unclear what targeting parameters are in place for political ads, but the platform does say that "Ads for certain products or services may not be targeted on the basis of gender, age, or location."

Discord: Yes

- Discord's profit model is based on subscriptions and they do not host ads.

LinkedIn: Yes

- LinkedIn prohibits the use of political ads: "Political ads are prohibited, including ads advocating for or against a particular candidate, party, or ballot proposition or otherwise intended to influence an election outcome; ads fundraising for or by political candidates, parties, political action committees or similar organizations, or ballot propositions; and ads exploiting a sensitive political issue even if the advertiser has no explicit political agenda" – but this does not specifically apply to generative AI.

Nextdoor: Yes

- Nextdoor's ad guidance appears to allow only geolocated ads (by radius or zip code). So while they do not appear to have a policy around AI generated ads, they do not allow personalized targeting of any ad.

X: No

- X prohibits political campaign ads from promoting "false or misleading content," but it does not explicitly include deceptively altered content in this policy.

2d. Embrace industry-wide provenance standards and require clear disclosures within any political or issue ad that features AI-generated images, video, or audio.

Facebook: Yes

- Meta's election integrity policies state that advertisers are now required to disclose whether "they use AI or digital methods to create or alter a political or social issue ad in certain cases." On February 6, 2024 Meta announced work to better identify AI-generated content, based on industry standards. This is an important step in the right direction, but it remains to be seen how it will apply to election ads particularly as global elections are currently underway.

Instagram: Yes

- Instagram does not have policies regarding AI or manipulated media. It can be assumed that Meta's political ad policies would apply to Instagram, as they share the same ad management platform and policies.

Threads: N/A

- Threads does not host advertisers at this time.

YouTube: Yes

- YouTube has a strong policy requiring disclosure for election ads. The advertiser themselves must be verified, and must disclose synthetic content that falsely depicts people or events, including image, video and audio. The disclosure must be "clear and conspicuous be placed in a location where it is likely to be noticed by users."

TikTok: N/A

- Since TikTok does not allow paid political ads, this recommendation is not applicable. Because so many influencers on the platform do work with electoral campaigns, it is important to note that TikTok asks creators to voluntarily label AI-generated content and further states that it may automatically apply a label to content that it detects was created or edited by AI. The policy goes on to say that labels are required for "all AI-generated content where it contains realistic images, audio, and video, as explained in our Community Guidelines."

Snapchat: No

- Snap's political ad policy states that "paid for by" disclosures are required, but it does not require that AI-generated ads be disclosed.

Discord: N/A

- This does not apply to Discord because they do not host ads.

LinkedIn: No

- LinkedIn does not have a publicly available policy on this specific topic yet.

Nextdoor: No

- Nextdoor's ad policy only allows political ads to be placed by managed service clients and requires they comply with FEC guidelines. There is nothing specific about provenance standards in their policy.

X: No

- X only mentions disclosure on political ads as applicable to local law.

2e. Prompt users as election season starts, outlining the main parameters of core algorithmic systems and giving users autonomy to adjust ranking criteria so their recommendations are not based on profiling or optimized for engagement.

Facebook: Partial

- While Meta does not prompt users as election season starts, they do give users meaningful control over their recommendations. Facebook's recommendation engine is now less likely to take engagement signals as a meaningful indicator to recommend political content, and they outline for users what this means for their Feeds.

Instagram: Partial

- On February 9, 2024 Instagram and Threads announced a policy to limit by default political content recommendations, but users have the option not to limit political content recommendations.

Threads: Partial

- On February 9, 2024 Instagram and Threads announced a policy to limit by default political content recommendations, but users have the option not to limit political content recommendations.

YouTube: No

- While YouTube says users have control over their recommendations, the instructions are not easy to find (there is no user prompt), they are not connected to election season, and they offer little transparency on the main parameters of core algorithmic systems.

TikTok: No

- Nothing in TikTok's election integrity policy or elsewhere in this research indicates there are prompts for users explaining its For You algorithm, but it does give users insight into the recommendation engine. While TikTok launched an Election Center in 2022, it did not give users the option to temporarily alter how recommendations are served to them.

Snapchat: No

- Snapchat does not appear to give users meaningful control over their algorithmic recommendations. There is no mention of how to change the recommendations in Spotlight or Discover in their Safety Center. They do inform content creators that they "reward creativity," which is the only indicator available of how their recommendations work.

Discord: N/A

- Discord does not have any specific policies related to election season, but the nature of the platform does not have an algorithmically recommended content.

LinkedIn: No

- LinkedIn does not have a publicly available policy on this topic.

Nextdoor: Partial

- Nextdoor allows users to sort their feeds by "Top posts, Recent posts, or Recent activity" and explains how the order of "Top Posts" is determined, which does appear to be optimized for engagement. Nextdoor does not appear to prompt users to sort their feeds, but it is accessible.

X: Partial

- X does not prompt users, but it does give some autonomy over recommendations.

PAPER TRAILS

Fostering Trust through Meaningful Transparency

3a. Clearly and accessibly detail in one place all election-related (or applicable platform-wide) policies and approaches

Facebook: Partial

- While Meta has a "Preparing for Elections" page, it is more of a public relations effort than a meaningful collection of election-integrity practices. Meta's November 28 policy update about the 2024 elections links to Facebook's Community Standards "policies on election and voter interference, hate speech, coordinating harm and publicizing crime, and bullying and harassment" as well as "Preparing for Elections page." We referenced numerous other policy pages for this research including "Our Approach to Political Content," "Our Approach to Facebook Feed Ranking," "Our approach to misinformation," "Our Approach to Facebook Feed Ranking," "Helping People Understand When AI Or Digital Methods Are Used In Political or Social Issue Ads," "Labeling AI-Generated Images on Facebook, Instagram and Threads," "How fact-checking works," "Request a verified badge on Facebook," Facebook's Community Standards related to Inauthentic Behavior and the Manipulated Media policy, and Facebook's 2020 policies including "Labeling State-Controlled Media On Facebook." We also reviewed policies around enforcement including "Meta's enforcement of fact-checker ratings," "Providing context on sensitive or misleading content," and "Restricting Accounts."
- In short, Meta provides incredibly detailed and nuanced election-related policies and approaches, but they are difficult to navigate, and in many cases unclear if they apply to just Facebook, or to Instagram and Threads as well. We agree with the declaration in the Oversight Board's recent ruling that Meta's manipulated media policy "is incoherent and confusing to users," but we find that extends to much of the company's election-related approach.

Instagram: No

- Outside of their ad policies Instagram largely does not host its own election related policies. Certain Meta policies specifically mention Instagram and Threads alongside Facebook, but it is far from clear and in one place. In conducting this research, we also consulted Instagram's "Community Guidelines," "Reducing the Spread of False Information on Instagram," "Verified Badges on Instagram," and the recent update "Continuing our Approach to Political Content on Instagram and Threads."

Threads: No

- Threads does not have their own election-integrity policy.

YouTube: Partial

- Google put out an update on Dec 19, 2023 that outlines policies related to the U.S. elections and includes relevant links to numerous policies referenced in this research. YouTube's Elections misinformation policies page also links to a June 2023 update rolling back the policy against false claims of widespread fraud in past presidential elections.

TikTok: Partial

- TikTok has robust and thorough policies, but they do appear in different parts of the site. In the course of doing this research, we accessed seven distinct pages from TikTok's Newsroom to its Safety Center and Transparency Center. Their blog post from January 2024 is concise, and while it links out to relevant pages, is not comprehensive.

Snapchat: Partial

- While Snap's "Planning for the 2024 Elections" post is discoverable in Snapchat's Privacy and Safety Hub, it is not easy to find and is not inclusive of all details about Snap's election-integrity policies. It is, however, clear and concise and includes relevant links.

Discord: No

- Discord only has a small section of its site dedicated to how it handles "Civic Disruptions" under its Misinformation Policy Explainer page, which states that the platform "will remove harmful misinformation about civic processes when we become aware of it. We will determine the truthfulness of claims by looking at independent, third-party sites like PolitiFact, FactCheck.org, and Snopes."

LinkedIn: No

- LinkedIn has vague election-related policies in 2022 about protecting users from election-related information, including "removing fake profiles, hate speech, posts and comments that incite violence, and partnering with fact-checking services to remove content confirmed to be false," but it's far from comprehensive, nor current.

Nextdoor: No

- The information required to assess the indicators in this framework required reviewing eleven different pages on Nextdoors site, three of which seemed to directly pertain to election integrity, but with very sparse information – open to different interpretations. It is easy to read, but it still lacks clarity.

X: No

- X has a clear Civic Integrity Policy page which appears to have been updated in August of 2023, but does not include all election-related policies nor granular detail about enforcement actions, which the platform previously shared.

3b. Release regular transparency reports during election season - broken down across widely spoken languages - detailing high-performing and violative content, enforcement, and resource allocation

Facebook: Partial

- Facebook does not release regular transparency reports during election season with these details. Facebook does provide limited information about high performing content, focusing on reach rather than engagement metrics, some Community Standards Enforcement Reports about content removal, and enforcement actions, which are not broken down by language spoken. As noted above, we could not find any reporting about specific resource allocation.

Instagram: Partial

- Instagram has their own Community Standards Enforcement reports, but nothing additional to the reports for Facebook.

Threads: No

- Threads does not appear to have their own transparency reports.

YouTube: Partial

- Google has a transparency report hub, and YouTube releases community guidelines enforcement reports on a quarterly basis, which detail how many videos are removed, for what reason and in which reasons. However, these reports are not timed to election season nor do they detail resource allocation or the types of violative content within the categories of "promotion of violence and violent extremism, other and misinformation" which are the three least removed categories in the most recent report (July-Sept 2023). Furthermore, they are not – at least as far as we could tell – broken down by language.

TikTok: Partial

- TikTok's Transparency Center includes granular detail about enforcement and in which regions. It is accessible in multiple languages. However it is released quarterly and does not appear to be published on a more frequent basis during election season.

Snapchat: Partial

- Snapchat releases bi-annual transparency reports, which is broken down by country and includes violative content and enforcement, but not resource allocation.

Discord: Partial

- Discord does publish transparency reports, but it is unclear whether this information specifically applies to election information.

LinkedIn: Partial

- LinkedIn has a transparency center where they "provide members with regular transparency updates on the actions we take to protect members" but this is not specific to elections.

Nextdoor: No

- There is a Transparency Report on Nextdoor's site from 2022, but it appears to only be available in English and there is no indication of frequent reporting.

X: No

- Before Elon Musk's takeover, X had a robust transparency center. While the site is still live, the most recent report about rules enforcement is dated December 2021.

3c. A snapshot of the week's highest-performing election-related content - the posts, advertisements, accounts, URLs, and groups that generated the most engagement

Facebook: Partial

- While Facebook does not release weekly reports, they publish a quarterly widely-viewed content report, which includes the number of impressions and clicks a post receives – although it is not broken down by language or specifically focused on election-related content. Facebook does not publish information about the specific accounts, URLs, or groups that are performing well.

Instagram: No

- Facebook's widely-viewed content report appears to only apply to Facebook.

Threads: No

- Threads is not included in Meta's transparency reports.

YouTube: No

- YouTube does not release weekly reports.

TikTok: No

- TikTok does not release weekly transparency reports.

Snapchat: No

- Snapchat does not release transparency reports specific to elections, nor do they share reports about high-performing content that isn't violative.

Discord: No

- Discord provides a page dedicated to "Resources for Exercising Your Right to Vote" and a "Misinformation Policy Explainer," but does not include weekly transparency reports specific to elections.

LinkedIn: No

- Linkedin does not release weekly transparency reports specific to elections.

Nextdoor: No

- Nextdoor appears to only release annual reports.

X: No

- X does not release regular transparency reports.

3d. Aggregate data about enforcement of all election-related policies, including specifics about the measures detailed in previous sections and their effectiveness

Facebook: Partial

- While Facebook publishes quarterly Community Standards Enforcement Reports, they don't comprehensively cover election-specific policies. These reports lack details about enforcement actions like removals, warnings, or account suspensions. Effectiveness data is also not included, although Meta takes steps to measure content actioned.

Instagram: Partial

- Instagram is included in the quarterly Community Standards Enforcement Reports, but the same concerns as above apply.

Threads: No

- Threads is not included in Meta's transparency reports.

YouTube: Partial

- YouTube tracks a 'Violative View Rate' for aggregate removal. The categories detailed in YouTube's transparency reports include: "promotion of violence and violent extremism, other and misinformation" which are the three least removed categories in the most recent report (July-Sept 2023 as of Feb 20, 2024).

TikTok: Yes

- TikTok publishes granular data including its enforcement of all of its policies including the percentage removed for violation of its integrity and authenticity policies. In January of 2024, they also stated that the platform intends to "introduce dedicated covert influence operations reports" in the coming months. They go much further than other platforms to include specific disinformation networks in their enforcement reports.

Snapchat: No

- Snapchat does not release data about its enforcement for election misinformation violations.

Discord: Partial

- Discord released their most recent "Transparency Report" on March 30, 2023, and shared how many accounts and servers were removed for misinformation or violent extremism, but falls short of sharing details of all election-related content.

LinkedIn: No

- Linkedin does not release data about its enforcement for election misinformation violations.

Nextdoor: No

- Nextdoor does not detail its enforcement mechanisms for election misinformation violations.

X: No

- X has not published data about enforcement in more than a year.

3e. Details about the number of full-time staff and contractors working on election integrity, broken down by department role and primary languages spoken

Facebook: Partial

- In a fact sheet, Meta says there are 40,000 people working on safety and security for elections globally. That number links to a 2023 report, which on page 74 cites the same number but without specific reference to elections. They do not share any details about the breakdown of its safety and security staff.

Instagram: Partial

- Instagram does not publish reports or release elections specific content details, but presumably Meta's overall report applies.

Threads: Partial

- Threads also does not publish their own reports, but presumably Meta's overall report applies.

YouTube: No

- YouTube does not publish these numbers.

TikTok: Partial

- TikTok details that there are 40,000 trust and safety professionals working to enforce their community guidelines, and publishes transparency reports in multiple languages, but it does not disclose the regions or primary languages of those 40,000 workers.

Snapchat: No

- Snapchat does not release details about their staffing breakdown.

Discord: No

- Discord does not release details about their staffing breakdown.

LinkedIn: No

- LinkedIn does not release details about their staffing breakdown, although a 2022 blog post outlines that a "dedicated team of data scientists, software engineers, machine learning engineers, and investigators" are working to "constantly [analyze] abusive behavior on the platform."

Nextdoor: No

- Nextdoor publishes no details about their staffing breakdown.

X: No

- X does not publish data about existing employees.

3f. Provide independent researchers with direct access to platform data to inform studies, threat analysis, and systemic impact assessments.

Facebook: Partial

- Facebook hosts an "Open Research and Transparency (FORT)" initiative, aimed at sharing data with independent researchers. Researchers based at academic institutions are eligible for access, with others considered individually. However, it is worth noting that Meta came under scrutiny for cutting off access for researchers at NYU who were investigating ad-targeting.

Instagram: Partial

- Instagram was included in the 2020 partnership with academic researchers to study the 2020 elections, however that program no longer appears to be in effect.

Threads: No

- There is no indication that Threads is included in FORT or other academic research initiatives.

YouTube: No

- According to a 2020 report from Mozilla Foundation, "Google periodically releases datasets for use by researchers; the current offerings include three YouTube datasets that may be of interest to researchers." However, that link is now broken and https://research.google appears to only show internal research products. Whatsmore, YouTube's terms of service disallow "access [of] the Service using any automated means (such as robots, botnets or scrapers)" which would be necessary for most independent research. If an exception exists for researchers, it's not immediately clear.

TikTok: Yes

- TikTok has a dedicated research program with its own Terms of Service for researchers.

Snapchat: No

- Snapchat does not appear to provide access to independent researchers.

Discord: No

- Although Discord has a page dedicated to its privacy policy, there is no information about access for independent researchers.

LinkedIn: Partial

- As of August 2023, LinkedIn offers a program to give independent researchers access to some of its data, but requires a list of criteria to be met via an application process for access to be granted.

Nextdoor: No

- A search for "Nextdoor independent researcher access" yielded no results. The company points to its research partnerships examining societal issues such as loneliness, but it does not appear to grant researchers access to examine the platform itself.

X: No

- While X used to be a leader in researcher access, since Elon Musk's takeover, the platform has rolled back programs such as its API for researchers. That page is now broken. The platform does provide endpoint access to public posts and replies to researchers and NGOs to "identify, understand and counter misinformation around public health initiatives," but started charging for API access in 2023.