+accountabletech

ELECTION INTEGRITY ROADMAP FOR SOCIAL MEDIA PLATFORMS

The United States is entering an election season fraught with unprecedented risks: a pandemic, a deluge of disinformation, leaders sowing distrust in the integrity of the election, new – and old – methods of voter suppression, and the potential for online communities to foment real-world violence after polls close. While every institution and citizen has a role to play, social media giants have an outsized influence in our modern information ecosystem and an added responsibility to mitigate these threats.

Across Silicon Valley, companies are working with each other and state and local governments to prepare, and they have already taken important steps – from Facebook's Voting Information Center, to Twitter's tightening enforcement of their civic integrity policy, to YouTube's crackdown on videos using hacked materials.[1] But significant gaps and vulnerabilities remain.

First and foremost, platforms must – at the bare minimum – immediately start enforcing their own civic integrity rules fully and consistently. To date they have been interpreting them so narrowly as to render them ineffectual.[2] Blatant lies about mail-in ballots or baseless attacks on legitimate, fair elections cannot be dismissed as standard political debate; they pose the same threat to our democracy as COVID-19 lies pose to our public health. And Facebook's deeply unpopular fact-checking exemption for politicians must be reversed for posts about voting, election integrity, and election results.[3] As Facebook's own Civil Rights Audit noted while condemning the exemption, "politicians have historically been some of the greatest perpetrators of voter suppression in this country."[4]

But those steps are not enough. And social media companies proactively outline clear, nonpartisan, enforceable policies now, rather than reacting haphazardly as crises arise. To that end, this Election Integrity Roadmap was designed in consultation with technologists, civil rights leaders, and disinformation experts to help social media companies responsibly navigate the 2020 election season, from early voting through the official certification of results.[5] Recognizing the difficulty of implementing changes at scale, the Roadmap is grounded in platforms' existing policies, tools, and technologies – and is limited in scope to address election-related threats only, not political misinformation issues more broadly. As Americans begin to cast their ballots in this unprecedented election, the integrity of democracy is at stake.

---

[1] Facebook Newsroom, tweet, (August 12, 2020), available at: https://twitter.com/fbnewsroom/status/1293631278953029632

[2] Daniel Kreiss and Yael Eisenstat, "Additional Steps Platforms Can Take to Protect the Vote," Center for Information, Technology, and Public Life, available at: https://citap.unc.edu/files/2020/08/CITAP-Election-Misinfo-Workshop-position-paper.pdf.

[3] An Accountable Tech/GQR poll found that more than 90% of both Democrats and Republicans believe politicians should be subject to the same fact-checking policies as everyone else, available at: https://accountabletech.org/wp-content/uploads/2020/07/Accountable-Tech-071920-FQ.pdf.

[4] Facebook Civil Rights Audit - Final Report (Jul. 8, 2020) ("The civil rights community expressed deep concern because politicians have historically been some of the greatest perpetrators of voter suppression in this country. By continuing to exempt them from fact-checking at a time when politicians appear to be increasingly relying on using misinformation, and giving them a presumption of newsworthiness in favor of allowing their speech to remain up, the civil rights community felt like Facebook was inviting opportunities for increased voter suppression. The Auditors shared the civil rights community's concerns and repeatedly (and vigorously) expressed those concerns directly to Facebook."), available at: https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf.

[5] Since 2016, platforms have made significant progress addressing foreign threats. This roadmap focuses on domestic threats.

+accountabletech

Social media companies have a responsibility to act and meet this moment, including by:

> **As early voting begins: Implementing a clearly defined Election Integrity Strike System.** To deter the worst vectors of disinformation at the Account, Page, Group and website level, the Election Integrity Strike System offers specific tiered restrictions and penalties. It would progressively limit the reach of serial disinformers and provide context to help inform and inoculate other users, defanging the worst actors before the most volatile period of the election season.

> **As Election Day nears: Implementing a Platform Poll Watchers program.** Just as election observers are deployed to ensure the legitimacy of democratic processes and report problems, Platform Poll Watchers would help ensure the integrity of the online information sphere. Platforms should create specialized verification labels for state Election Directors and nonpartisan civil society groups with expertise in voting rights, election law, or online disinformation, allowing them to promote credible information, flag specific pieces of misleading content, and counter-message false narratives in real-time. As in the context of COVID-19, these trusted accounts would serve as the first line of defense against misinformation.

> **As polls close: Implementing Violence Prevention Preclearance.** Just as the Voting Rights Act required certain states to preclear new voting laws, social media platforms should implement Violence Prevention Preclearance for election-related posts from highly influential accounts. These posts would be subject to proactive detection and rapid human review, preempting content that would violate incitement of violence or civic integrity policies. Posts that haven't been checked within 10 minutes would be published with a flag noting review is pending.

TIMELINE OF RECOMMENDATIONS

Below is a full timeline of recommendations, with three flagship proposals highlighted and expanded upon in supplemental sections. Each of the proposals would be **phased in at the start date and continue until results are certified**. Heading off major vectors of disinformation early and heightening protections as Election Season progresses is a cumulative process.

NOW

> **Begin fully and consistently enforcing existing civic integrity policies.**

> **Don't let politicians lie about elections.** Eliminate fact-checking exemptions for politicians on posts or ads about voting, election integrity, or results.

SEPTEMBER 14

> ## KEY PROPOSAL — ELECTION INTEGRITY STRIKE SYSTEM

> **Correct the record**. There is little doubt about the effectiveness of well-designed fact-checking treatments – both in real-time and retroactively. Avaaz has found that properly correcting the record for users that have been exposed to lies would reduce belief in misinformation by *half*[6] and would not trigger the once-feared 'backfire effect'.[7] According to Facebook, their warning labels have deterred users from viewing the underlying content 95%[8] of the time. Facebook recently began alerting users who interacted with coronavirus misinformation[9] – but unfortunately, these are too generic to have any meaningful impact. Platforms should implement warning labels and clear explanations for all voting misinformation posts, and push similarly tailored notifications to users who have previously been exposed.

> **Begin expanding context and friction.** Context and friction are key mechanisms for better educating voters and slowing the spread of viral misinformation, while minimizing restrictions on speech. Platforms can adapt and expand upon existing tools to bolster election integrity.

   + Append authoritative information to all voting-related posts. Facebook has already begun labeling all voting-related posts that point users to their authoritative Voting Information Center. Other platforms should follow suit.

      + All platforms should make clear these neutral labels are distinct from, and not a substitute for, fact-checking treatments. Facebook's efforts have already drawn consternation for this ambiguity,[10] and Twitter's "get the facts" label – which is meant to flag misinformation – has been criticized for its similar lack of clarity.[11]

   + Enhance transparency on Posts, Pages, Accounts, and Groups. Providing more information about the origins of content and actors can be valuable for users evaluating their trustworthiness. Facebook's "Page Transparency" function discloses when Pages were created, whether their name has changed, who is behind them, ads they have run, and – for high-reach pages – where they are posting from. This should be expanded with additional background, extended to Groups, adopted by other platforms, and featured prominently, rather than forcing users to search for the transparency features.

---

[6] Avaaz, "Correcting the Record," (Apr. 16, 2020), available at: https://secure.avaaz.org/campaign/en/correct_the_record_study/.

[7] Brendan Nyhan, Ethan Porter, et. al., "Taking Fact-checks Literally But Not Seriously? The Effects of Journalistic Fact-checking on Factual Beliefs and Candidate Favorability," Political Behavior (Jan.7, 2019), available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2995128.

[8] Facebook, "Using AI to detect COVID-19 misinformation and exploitative content," (May 12, 2020), available at: https://ai.facebook.com/blog/using-ai-to-detect-covid-19-misinformation-and-exploitative-content.

[9] Guy Rosen, "An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19," (Apr. 16, 2020), available at: https://about.fb.com/news/2020/04/covid-19-misinfo-update/.

[10] Shirin Ghaffary, "Facebook's new label on a Trump post is an 'abject failure,' says a Biden campaign spokesperson," Vox (Jul. 21, 2020), available at: https://www.vox.com/recode/2020/7/21/21333001/facebook-label-trump-fact-check-biden-democrats-republicans-2020-elections-policy.

[11] Gilad Edelman, "Twitter Finally Fact-Checked Trump. It's a Bit of a Mess," Wired, (May 27, 2020), available at: https://www.wired.com/story/twitter-fact-checked-trump-tweets-mail-in-ballots/.

+ Begin certifying *Platform Poll Watchers* and appending special verification icons to their accounts. (More on this below.)

+ Work with government partners to ensure all official county election board accounts are verified and any imposter accounts are flagged for removal or other appropriate recourse.

+ Deploy "informative friction" interstitials for voting-related posts.[12] Introducing front-end friction is an easy way to better inform users and curtail the spread of false content. Even simple accuracy prompts that ask people to pause and think critically before proceeding have been proven to reduce the likelihood of sharing false news without having the same effect on legitimate headlines.[13] Twitter has been testing a feature along these lines that prompts users to open links before retweeting them.[14] And Facebook recently introduced pop-ups when users try to share coronavirus-related links that add context about the article's origin and recency.[15] An expanded version of these interstitials should be instituted across all platforms for all voting-related content – including repeat offender flags for any posts linked to Accounts, Pages, Groups, or websites with strikes against them. (More on this below.)

> **Roll out weekly transparency reports on enforcement of civic integrity policies**. Facebook, Twitter, and YouTube all release quarterly or biannual transparency reports on detection and enforcement of policy violations. All platforms should begin providing reports for election integrity strikes and related violations – broken down by categories of actor and infraction – on a weekly basis to give the public greater visibility into the extent and accuracy of their enforcement. This should include data on successful and unsuccessful appeals, and a full accounting of all 'Superspreader' designations.

OCTOBER 1

> **KEY PROPOSAL — PLATFORM POLL WATCHERS**

> **Escalate proactive information distribution efforts.**

+ Promote voter assistance tools from the nonpartisan Election Protection coalition, including the 866-Our-Vote hotline and ProtectTheVote.net volunteer program.

+ Provide free advertising credits to state Election Directors.

+ Send users push alerts with timely information about how to vote early or by mail.

+ Begin educating voters about what to expect after ballots are cast. Platforms have a responsibility to help users understand the unique dynamics of this election season, emphasizing that official results might not be known for days after polls close and

---

[12]David Gillis, tweet, (Aug. 12, 2020), available at: https://twitter.com/davegillis/status/1293585791822643200

[13] Lisa Fazio, "Pausing to consider why a headline is true or false can help reduce the sharing of false news," Harvard Kennedy School's Misinformation Review (Feb. 10, 2020), available at: https://misinforeview.hks.harvard.edu/article/pausing-reduce-false-news/

[14]Twitter Support, tweet, (June 10, 2020), available at: https://twitter.com/TwitterSupport/status/1270783537667551233

[15]John Hegeman, "Providing People with Additional Context About Content They Share," Facebook (Jun. 25, 2020), available at: https://about.fb.com/news/2020/06/more-context-for-news-articles-and-other-content

neither delays nor partisan swings will indicate anything untoward. This step is critical to inoculating users from inevitable misinformation about the election's legitimacy.

> **Introduce additional context and friction.**

+ Begin appending flags to all new Pages/Accounts/Groups and those that have changed their name to prompt users to be more alert to potential deception.

+ Implement virality circuit breakers if possible.[16] As proposed by information policy expert Ellen P. Goodman[17] and expanded on by the Center for American Progress, this mechanism would function like a stock market circuit breaker, temporarily throttling unverified fast-spreading content and prioritizing it for immediate human review.

> **Surge capacity for election integrity content moderation.**

+ Detail steps being taken to train all content moderators on civic integrity policies, working with *Platform Poll Watchers* to ensure best practices are incorporated.

+ Outline plans to boost on-duty content moderator presence in the lead-up to and immediate aftermath of November 3.

OCTOBER 20

> **Tighten advertising standards to prevent foul play.**

+ Apply the highest standard of existing transparency requirements to all ads, not just "political" advertisers. This will eliminate the possibility of partisan "news publishers" or bad actors engaging in electioneering while avoiding scrutiny.[18]

+ Place a freeze on new advertisers until post-election to prevent bad actors and other mysterious groups from swooping in at the last minute with significant ad buys.

  + Facebook recently announced a freeze on all new political ads beginning October 27.[19] While this will effectively freeze new "political" advertisers, it will also block campaigns and even officials from running new ads at a critical juncture, while failing to address the issue that the "political" ad classification relies on self-reporting. Additionally, the freeze ends right after 'Election Day', green-lighting new political ads without enhanced restrictions in the volatile period between when polls close and when results are certified.[20]

---

[16] Erin Simpson and Adam Conner, "Fighting Coronavirus Misinformation and Disinformation," Center for American Progress (Aug. 2020), available at: https://cdn.americanprogress.org/content/uploads/2020/08/10102935/Coronavirus-Disinformation.pdf.

[17] Ellen P. Goodman, "Digital Information Fidelity and Friction," Knight First Amendment Institute (Feb. 26, 2020), available at: https://knightcolumbia.org/content/digital-fidelity-and-friction.

[18] Facebook, "Updates to Our Ad Transparency and Authorization Efforts," (Nov. 29, 2018), available at: https://www.facebook.com/business/news/updates-to-our-ad-transparency-and-authorization-efforts.

[19] Mark Zuckerberg, "New Steps to Protect the U.S. Election," (Sept. 3, 2020), available at: https://about.fb.com/news/2020/09/additional-steps-to-protect-the-us-elections/.

[20] Katie Paul, "Facebook to limit last-minute political ads before U.S. elections, label premature victory claims," Reuters (Sept. 3, 2020), available at: https://www.reuters.com/article/us-usa-election-facebook/facebook-to-limit-last-minute-political-ads-before-u-s-elections-label-premature-victory-claims-idUSKBN25U1N3.

> **Limit algorithmic amplification.** The largest social media platforms all employ content curation and/or recommendation algorithms designed to maximize engagement, thereby boosting whatever is most outrageous.[21] For the final week of voting in this unprecedented election season, platforms must elevate authoritative information and disincentivize false, divisive, and sensationalist posts that often travel the fastest.[22]

+ <u>Temporarily turn off algorithmic curation of Twitter timeline and Facebook News Feed in default settings.</u> These feeds would therefore be reverse-chronological (with authoritative election updates pushed to the top) and recommendation tools disabled. Users could still opt to turn such tools back on. This stopgap would significantly limit the preponderance of – and the incentive to produce – toxic content, without depriving anyone of functionality.[23]

+ <u>Temporarily turn off YouTube autoplay and "Up Next" engine</u>. Instead, direct users to authoritative news sources or human-curated content. Critics have long warned that these tools push users to increasingly extreme and conspiratorial content to keep them watching – a concern validated by YouTube's own senior personnel[24] and engineers.[25]

+ <u>Temporarily turn off Facebook Group recommendations.</u> In 2016, Facebook's own internal research showed that "64% of all extremist group joins are due to our recommendation tools."[26] Experts have repeatedly warned that Group recommendations help grow dangerous disinformation networks.[27] Facebook must take every precaution to ensure they're not pushing users into these rabbit holes in the peak weeks of election season.

+ <u>Temporarily turn off Trending Topics on Twitter</u>. Although Twitter recently began adding context to Trending Topics amid widespread concerns about the tool's propensity for being gamed by bad actors, insiders warned the half-measure wasn't an adequate remedy.[28]

---

[21]Mark Zuckerberg, "A Blueprint for Content Governance and Enforcement," (Nov. 15, 2018), available at: https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/?comment_id=224239918468772.

[22]Peter Dizikes, "On Twitter, false news travels faster than true stories," MIT News (Mar. 8, 2018), available at: http://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308.

[22] See e.g., https://accountabletech.org/campaign/facebook-turn-off-the-algorithms/

[23]Mark Bergen, "YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant," Bloomberg (Apr. 2, 2019), available at: www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant.

[24]Paul Lewis and Ericn McCormick, "How an ex-YouTube insider investigated its secret algorithm," The Guardian, (February 2, 2018), available at: https://www.theguardian.com/technology/2018/feb/02/youtube-algorithm-election-clinton-trump-guillaume-chaslot

[25]Salvador Rodriguez, "Mark Zuckerberg shifted Facebook's focus to groups after the 2016 election, and it'sc changed how people use the site," CNBC, (February 16, 2020), available at: https://www.cnbc.com/2020/02/16/zuckerbergs-focus-on-facebook-groups-increases-facebook-engagement.html

[26]Craig Silverman, Jane Lytvynenko, and Lam Thuy Vo, "How Facebook Groups Are Being Exploited To Spread Misinformation, Plan Harassment, And Radicalize People," Buzzfeed News, (March 19, 2018), available at: https://www.buzzfeednews.com/article/craigsilverman/how-facebook-groups-are-being-exploited-to-spread

[27]Associated Press, "Calling Election Winners," available at: https://www.ap.org/en-us/topics/politics/calling-election-winners

[28]Avaaz, "Facebook's Algorithm: A Major Threat to Public Health", (Aug. 19, 2020), available at: https://secure.avaaz.org/campaign/en/facebook_threat_health/.

NOVEMBER 3 | AFTER POLLS CLOSE

> **KEY PROPOSAL — VIOLENCE PREVENTION PRECLEARANCE**

> **Strictly adhere to AP or NEP official results and standards for partial reporting.** Exercising restraint and providing more context than usual will be key to staving off widespread distrust in the legitimacy of the results. The Associated Press and the National Election Pool – a consortium of ABC News, CBS News, CNN, and NBC News – are the most authoritative sources when it comes to responsibly reporting election data and ultimately calling races.[29,30] Facebook recently announced a partnership with Reuters and NEP that they'll rely on for surfacing credible updates and data.[31] Other platforms should follow suit.

+ Platforms should continue appending labels to all election-related posts directing users to authoritative information hubs, and should engage in ongoing education efforts – in partnership with AP or NEP and Platform Poll Watchers – to contextualize results and related developments as votes are tabulated.

---

[29] Niam Yaraghi, "How should social media platforms combat misinformation and hate speech?" Brookings Institute, (Apr. 9, 2019), available at: https://www.brookings.edu/blog/techtank/2019/04/09/how-should-social-media-platforms-combat-misinformation-and-hate-speech/.

[30] Facebook, "Enforcing Our Community Standards," (Aug. 6, 2018) ("We don't want people to game the system, so we do not share the specific number of strikes that leads to a temporary block or permanent suspension."), available at: https://about.fb.com/news/2018/08/enforcing-our-community-standards/.

[31] Mark Zuckerberg, "New Steps to Protect the U.S. Election," (Sept. 3, 2020), available at: https://about.fb.com/news/2020/09/additional-steps-to-protect-the-us-elections/.

ELECTION INTEGRITY STRIKE SYSTEM

Phase-In Date: September 14 | Beginning of Voting

A disproportionate amount of harmful misinformation on social media platforms can be tied back to a relatively small number of accounts, groups, and websites that effectively serve as superspreaders.[32] Moderators will never be able to eliminate every problematic piece of content on their massive platforms, nor should they try. Instead, as experts have noted, the goal should be mitigation.[33] By imposing a series of escalating penalties for repeat offenders during election season, platforms can head off the most insidious vectors of voting disinformation.

The Election Integrity Strike System offers specific recommendations for tiered restrictions to deter this behavior. It would limit the reach of serial disinformers and provide context to help inform and inoculate other users. These proposals are rooted in **existing guidance** major platforms like Facebook, Twitter, and YouTube already have around repeat offenders and treatments they already apply to certain content.

> While platforms have fairly noted that publishing their thresholds might help bad actors game the system,[34] they should outline their rules as clearly as possible to preempt claims of bias and effectively deter bad behavior, including the fact that the most serious offenses could immediately lead to high-level punishments.

> This is a cumulative penalty structure that should disincentivize deceptive behavior meant to undermine our democracy. If implemented successfully, the system will drastically limit the reach of the worst actors before the information ecosystem will be most vulnerable to exploitation.

+ Strikes that relate solely to voting misinformation (i.e. did not also violate policies against incitement of violence, hate speech, etc.) could be removed after results are certified, as this effort is narrowly designed to protect the election.

+ Elected officials and candidates for office would be generally exempt from Account-level punishments. All other users should be guaranteed the right to rapidly appeal strikes. Any escalation to the "Superspreader" level should require confirmation from multiple moderators and be documented in weekly transparency reports.

+ Platforms should consider partnering with established credibility evaluators like NewsGuard and the Oxford Internet Institute's Junk News Aggregator to identify prolific disinformation websites and determine proper Election Integrity Strike ratings.

---

[32] Avaaz, "Facebook's Algorithm: A Major Threat to Public Health", (Aug. 19, 2020), available at: https://secure.avaaz.org/campaign/en/facebook_threat_health/.

[33] Niam Yaraghi, "How should social media platforms combat misinformation and hate speech?" Brookings Institute, (Apr. 9, 2019), available at: https://www.brookings.edu/blog/techtank/2019/04/09/how-should-social-media-platforms-combat-misinformation-and-hate-speech/.

[34] Facebook, "Enforcing Our Community Standards," (Aug. 6, 2018) ("We don't want people to game the system, so we do not share the specific number of strikes that leads to a temporary block or permanent suspension."), available at: https://about.fb.com/news/2018/08/enforcing-our-community-standards/.

+accountabletech

ELECTION INTEGRITY STRIKE SYSTEM

Tiered penalties for repeat voting misinformation spreaders will limit the reach and impact of bad actors throughout election season.

	ACCOUNTS / PAGES	WEBSITES	FACEBOOK GROUPS
ON NOTICE	• Alert account that they earned a strike for sharing voting misinformation • Include fact-check and warning about future penalties	• Alert website that they earned a strike for sharing voting misinformation • Include fact-check and warning about future penalties	• Facebook should proactively scan for high-reach groups with an excess of posts from flagged accounts or linking to flagged websites, and subject those groups to higher scrutiny
REPEAT OFFENDER	• Temporarily append a flag to the account noting it has repeatedly spread voting misinformation • Ask users who try to share voting-related posts from this account if they're sure they want to proceed	• Temporarily append a flag to posts with links to the site noting it has repeatedly spread voting misinformation • Notify users who try to share posts about voting with links to the site and ask if they're sure they want to proceed	• Temporarily append a flag to the group about elevated voting misinformation risk; warn users with pop-up before they join • Notify members, ask if they wish to leave, and warn that continued violations will lead to further restrictions
FREQUENT MISINFORMER	• Append misinformer flag until post-election and warn users who try to share content • Revoke ability to advertise or serve as a group admin/moderator until post-election • Temporarily down-rank all posts by the account, remove from recommendations, and reduce in search results	• Append misinformer flags to posts linking to the site until post-election and warn users who try to share them • Prohibit ads from linking to site and remove any 'news' designation until post-election • Temporarily down-rank all posts about voting linking to the site	• Make repeat offender flag permanent until post-election • Temporarily remove posts within group from main News Feed, remove group from recommendations, reduce in search results • Require admins verification before lifting any restrictions and alert group members
SUPER-SPREADER	• Place all posts behind warning label, down-rank, and remove from recommendations until post-election • Alert other major platforms • Warn users who try to share voting-related posts from this account that it may count as a strike against them	• Place all posts linking to the site behind warning label, down-rank, and remove from recommendations until post-election • Alert other major platforms • Warn users who try to share voting-related posts linking to the site that it may count as a strike against their account	• Suspend the group until post-election • Subject admin/moderators' accounts to orange-level restrictions • Find other high-reach groups with overlap in membership and scam for violations

PLATFORM POLL WATCHERS

Phase-In Date: October 1

Just as election observers are deployed to ensure the legitimacy of democatic processes and report problems, Platform Poll Watchers should help ensure the integrity of the online information sphere. Platforms should create specialized verification labels for state Election Directors, as well as nonpartisan civil society organizations with expertise in voting rights, election law, or online disinformation.[35] These trusted accounts would push out proactive updates in real-time and serve as the first line of defense against misinformation. In the context of COVID-19, elevating trusted and reliable sources has been critical to combating misinformation and ensuring user safety.[36] Platforms should similarly empower nonpartisan experts to protect the integrity of our elections. Platforms should provide these Poll Watchers with clear guidance on best practices and proper resources to carry out the following tasks:

> **Promoting authoritative information**. Platform Poll Watchers should have their posts elevated across platforms and included in any authoritative information hubs where geographically relevant. Facebook, Twitter, and YouTube already do this in the context of COVID-19. Among other things, Platform Poll Watchers should help provide context about what to expect in terms of timelines for results. State Election Directors should further be enabled to push notifications with any critical updates to their constituents that may arise.

> **Flagging misinformation**. Platform Poll Watchers should be empowered to flag Unconfirmed or Disputed posts and links pertaining to their jurisdiction. This would alert platforms' election teams that potential voting misinformation had been detected for further review. Similar to how Facebook works with independent fact-checking partners to flag false content, users that engage with the post in the intervening period would see a pop-up notifying them that the content has been flagged by an independent partner and provide context, such as:

  + This post was marked as Disputed by Lawyers' Committee at 9:45am

  + This post was marked as Unconfirmed by the [State] Election Director at 10:05am

  + [Platform] integrity team will review it as soon as possible to provide a conclusive ruling; in the meantime, please exercise caution

> **Counter-messaging.** Platforms should design a format for Poll Watchers to provide clear counter-messaging to false or misleading election narratives that does not amplify the original misinformation itself. Those vectors should also be flagged immediately for platforms' election teams and the Election Integrity Partnership for further analysis,[37] including whether platforms should boost Poll Watchers' counter-messaging to relevant audiences.

---

[35]NASED, "Our Members: NASED Members," available at https://www.nased.org/members; FSI Stanford, "Fighting Election-Related Disinformation with the Election Integrity Partnership," YouTube, (August 12, 2020), available at: https://www.youtube.com/watch?v=JjpxevSKLGA&feature=youtu.be&t=1112
[36]Center for American Progress, "Fighting Coronavirus Misinformation and Disinformation", (Aug. 18, 2020), available at: https://www.americanprogress.org/issues/technology-policy/reports/2020/08/18/488714/fighting-coronavirus-misinformation-disinformation/.
[37] Election Integrity Partnership, webpage, available at: https://www.eipartnership.net/

VIOLENCE PREVENTION PRECLEARANCE

Phase-In Date: November 3 | After Polls Close

Experts are already warning of an "unprecedented post-election crisis" if the validity of the results is in doubt.[38] The spread of misinformation as ballots are being counted has the potential to cause chaos and even real-world violence. A recent report found that 1-in-5 Democrats and Republicans say that "violence would be at least a little justified" if the other party won, with experts warning that social media giants and news organizations may "hold the state of the republic in their hands" once polls close.[39] Platforms must take steps to stem violence and election misinformation before it has the chance to take hold.

Facebook, Twitter, and YouTube already use proactive detection and enforcement to protect children, identify nudity, and target hate speech. As Mark Zuckerberg has said, "proactive enforcement simply helps us remove more harmful content, faster."[40] Pre-emptive action to prevent election-related harms also has strong precedent in American democracy. Just as the Voting Rights Act required certain states to preclear new voting laws to ensure they did not harm voters, social media platforms should implement a rapid preclearance process for high-reach accounts to screen posts for violations of violence incitement and civic integrity policies in the volatile period after polls close.

> **Accounts Subject to Preclearance.** After all polling places are closed, accounts with more than 250,000 followers and those with Election Integrity Strikes against them should be subject to preclearance for election-related posts. Utilizing existing proactive detection technologies, these posts would be flagged before publication for expedited human review against violence and civic integrity policies. *Within ten minutes*, all posts not in clear violation of those policies would publish with the context outlined below. News organizations (except those with Election Integrity Strikes against them) and Platform Poll Watchers would be exempt.

> **Content Subject to Moderation.** All election-related posts from these accounts would receive expedited review to identify potential violations in two areas:

   + Incitement of Violence – Facebook, Twitter, and YouTube already prohibit content that encourages or celebrates violence. Under this proposal, posts determined to violate these existing policies in the context of the election would be blocked before they have the opportunity to do damage during this volatile period, rather than being removed reactively. The offending accounts would be notified of the action and subject to potential strikes.

---

[38]Marshall Cohen, "'Things could get very ugly': Experts fear post-election crisis as Trump sets the stage to dispute the results in November," CNN, (July 21, 2020), available at: https://www.cnn.com/2020/07/20/politics/disputed-election-crisis-trump/index.html
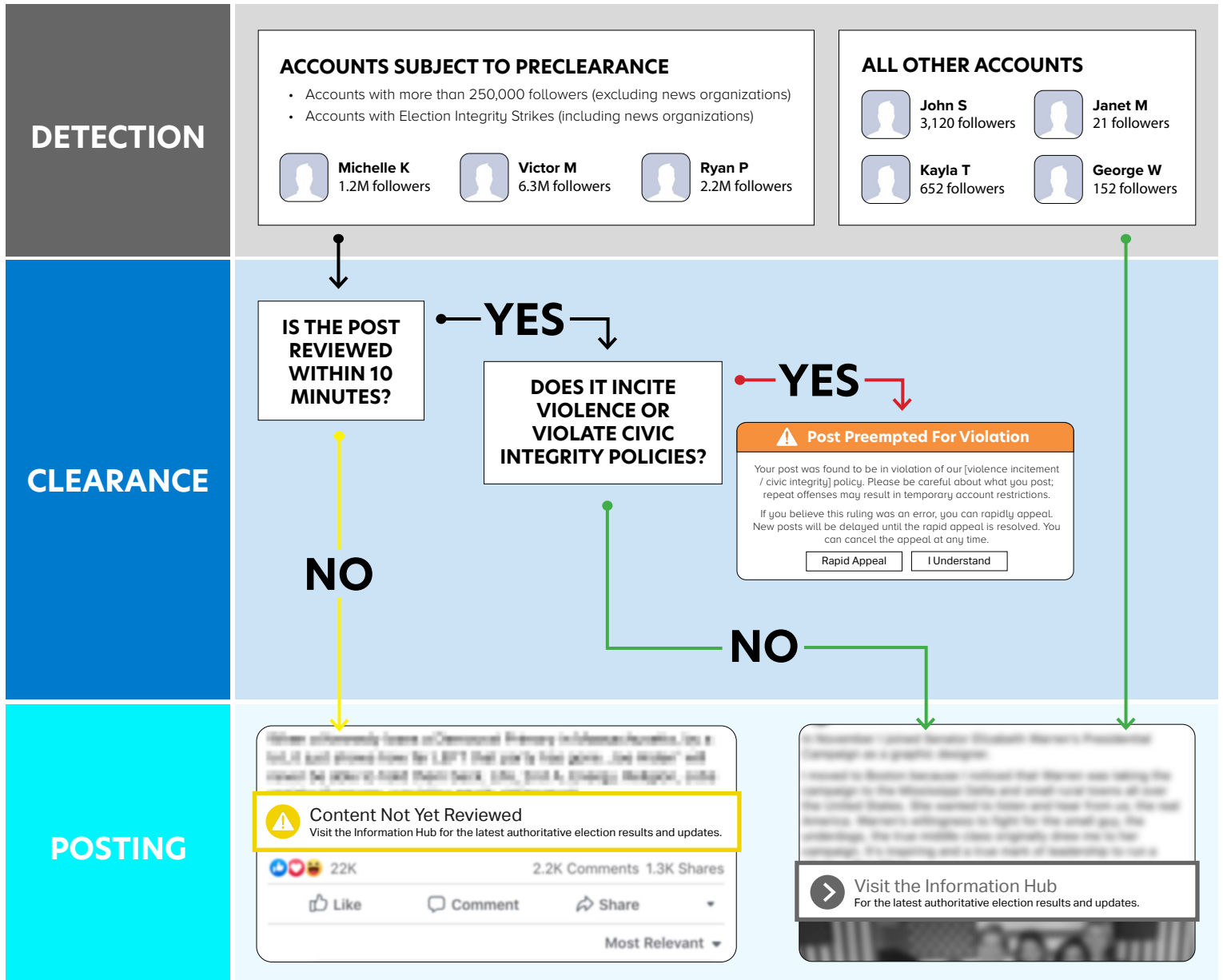
[39]Democracy Fund Voter Study Group, "Democracy Maybe: Attitudes on Authoritarianism in America", (Jun. 2020), available at: https://www.voterstudygroup.org/publication/democracy-maybe; Transition Integrity Project, "Preventing a Disrupted Presidential Election and Transition," (August 3, 2020), available at: https://paxsims.files.wordpress.com/2020/08/preventing-a-disrupted-presidential-election-and-transition-8-3-20.pdf; Zack Beauchamp, "How to avert a post-election nightmare," Vox, (August 18, 2020), available at: https://www.vox.com/policy-and-politics/2020/8/18/21371964/2020-transition-integrity-project-simulation-trump

[40]Mark Zuckerberg, "A Blueprint for Content Governance and Enforcement," (Nov. 15, 2018), available at: https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/?comment_id=224239918468772.

+ Civic Integrity – [Facebook](), [Twitter](), and [YouTube]() have already committed to cracking down on election misinformation, meaning they have the ability to detect election-related posts. But this content will inherently be harder to evaluate in real-time than content encouraging or inciting violence, and platforms should not be expected to act with a heavy hand. Under this proposal, only election-related posts that clearly violate existing civic integrity policies would be blocked. All other posts that can't be immediately verified would be appended with a label noting the content has not yet been verified and directing users to the platform's election.

VIOLENCE PREVENTION PRECLEARANCE FLOWCHART

<u>After polls close</u>, use proactive detection of election-related posts from high-reach accounts and expedited human review to preempt posts that violate violence incitement or civic integrity policies.

DETECTION

ACCOUNTS SUBJECT TO PRECLEARANCE
- Accounts with more than 250,000 followers (excluding news organizations)
- Accounts with Election Integrity Strikes (including news organizations)

Michelle K 1.2M followers
Victor M 6.3M followers
Ryan P 2.2M followers

ALL OTHER ACCOUNTS

John S 3,120 followers
Janet M 21 followers
Kayla T 652 followers
George W 152 followers

CLEARANCE

IS THE POST REVIEWED WITHIN 10 MINUTES?

YES

DOES IT INCITE VIOLENCE OR VIOLATE CIVIC INTEGRITY POLICIES?

YES

⚠ **Post Preempted For Violation**

Your post was found to be in violation of our [violence incitement / civic integrity] policy. Please be careful about what you post; repeat offenses may result in temporary account restrictions.

If you believe this ruling was an error, you can rapidly appeal. New posts will be delayed until the rapid appeal is resolved. You can cancel the appeal at any time.

Rapid Appeal | I Understand

NO

NO

POSTING

⚠ Content Not Yet Reviewed
Visit the Information Hub for the latest authoritative election results and updates.

😊😍😠 22K   2.2K Comments 1.3K Shares

👍 Like   💬 Comment   ↗ Share   ▾

Most Relevant ▾

→ Visit the Information Hub
For the latest authoritative election results and updates.

ELECTION INTEGRITY ROADMAP PROPOSED TIMELINE

All proposals would be phased in at the listed dates and remain in place until election results are certified.

AS EARLY VOTING BEGINS

> NOW - Eliminate fact-checking exemptions for politicians on posts about voting, election integrity, or results.

> **SEPTEMBER 14 – Implement a clearly defined *Election Integrity Strike System* aimed at deterring and penalizing the worst vectors of disinformation at the account, page, group, and website level.**

> SEPTEMBER 14 – Start correcting the record with specific fact-checking treatments both in users' feeds and via retroactive alerts to users previously exposed.

> SEPTEMBER 14 – Begin expanding context and friction to educate voters and slow the spread of viral lies, including by appending authoritative information to all voting-related posts, proactively surfacing key information about the origins of content and actors disseminating it, and prompting users to think critically about the accuracy of election-related information before they share content.

> SEPTEMBER 14 – Release weekly transparency reports on civic integrity enforcement.

AS ELECTION DAY NEARS

> **OCTOBER 1 – Implement a *Platform Poll Watchers* program, empowering state Election Directors and NGOs specializing in voting rights, election law or disinformation to promote credible information, flag specific pieces of misleading content, and counter-message false narratives in real-time.**

> OCTOBER 1 – Promote voter assistance tools from the nonpartisan Election Protection coalition, including the 866-Our-Vote hotline and ProtectTheVote.net volunteer program.

> OCTOBER 1 - Introduce additional context, friction, and transparency, including by appending flags to all new accounts/groups and those that change names and implementing virality circuit breakers if possible.

> OCTOBER 1 - Escalate proactive information distribution, including by beginning voter education about the likely timeline for election results, providing free advertising credits to state election officials, and sending users push alerts about how to vote early or by mail.

> OCTOBER 1 – Detail steps being taken to train all content moderators on election-related policies and all plans to surge moderation capacity in the lead-up to November 3 and following days.

> OCTOBER 20 - Tighten advertising standards to prevent foul play, including by applying the highest transparency standards to all ads, not just "political" ones, and placing a freeze on new advertisers until post-election.

> OCTOBER 20 - Limit algorithmic amplification, including by temporarily turning off algorithmic curation in default settings, turning off YouTube's autoplay and "Up Next" engine, and halting Facebook Group recommendations and Twitter Trending Topics.

AFTER POLLS CLOSE

> **NOVEMBER 3 – Implement Violence Prevention Preclearance for election-related posts from the highest-reach accounts for proactive detection and rapid human review, preempting content that would violate incitement of violence or civic integrity policies before they can do harm.**

> NOVEMBER 3 – Strictly adhere to Associated Press or National Election Pool results and standards for partial reporting, providing as much context to users as possible about the current circumstances and what to expect moving forward.