

VOTER PROTECTION PRECLEARANCE

As Americans have begun casting their ballots, we've seen increasingly dangerous efforts to intimidate, suppress, and deceive voters — many times via social media posts that clearly violate platforms' policies. When that content comes from high-reach accounts, by the time platforms take action, the posts have often already been seen by millions. The damage has been done.

Platforms must do more to address these harms before users are exposed, and they have the tools to do it. [Facebook](#) and [Twitter](#) already use proactive detection and enforcement, combining AI and human review to preempt other violations. As Mark Zuckerberg has said, "proactive enforcement simply helps us remove more harmful content, faster."¹ Similarly, Twitter has touted their "increased focus on proactively surfacing potentially violative content for human review."²

There are strong precedents for preemptive action to prevent harms, both in traditional media—where live TV events are broadcast on a short delay—and in American democracy. Just as the Voting Rights Act required preclearance for new voting laws from states with a history of discrimination, social media platforms should immediately implement Voter Protection Preclearance (VPP) for high-reach accounts with civic integrity violations on all election-related posts.

- > **Accounts Subject to Preclearance.** Only high-reach accounts with a history of election integrity violations would be subject to preclearance. "High-reach" includes candidates for federal office or accounts with more than 100,000 followers. "Election integrity violations" refers to posts subject to Facebook's [more specific labeling](#), Twitter's [civic integrity](#) labels, or removal on these grounds from either platform. Those actions have been taken on both [Democrats](#) and [Republicans](#).
- > **Content Subject to Moderation.** Voting-related content from accounts subject to VPP would be flagged for rapid human review before they are published. Content that appears to violate civic integrity or violence policies would be escalated internally, while all other posts would be published within 10 minutes.
- > **Preclearance Process.** Any posts with voting misinformation, attempts to delegitimize the election, or misrepresentations of the results would—at a minimum—be published behind strong warning labels that require click-through and provide context, similar to Twitter's [public interest notice](#) or Facebook's [fact-checking warning labels](#). More egregious violations—including any attempts to intimidate voters or incite violence—would be preempted. Repeat offenses should lead to additional account-level sanctions.

¹ 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.

² Twitter, Inc. blog post, "15th Transparency Report: Increase in proactive enforcement on accounts," (Oct. 31, 2019), available at: https://blog.twitter.com/en_us/topics/company/2019/twitter-transparency-report-2019.html.