



**FP** Free  
Press

# EMPTY PROMISES:

Inside Big Tech's Weak Effort to  
**FIGHT HATE AND LIES** in 2022

# Table of Contents

Executive Summary.....	3
Defining the Problem.....	6
What Platforms Claim They're Doing.....	7
Grading the Companies.....	8
Gaps in Platform Performance to Protect Users.....	10
Fixing the Algorithm & Downranking Violative Content .....	10
Bolstering Civic-Integrity Teams & Protecting Users Across Languages.....	13
Transparency & Disclosures .....	15
Recommendations for How Big Tech Can Disrupt Hate and Lies.....	16
Methodology and Acknowledgements.....	17
Endnotes.....	18

# Executive Summary

What happens on social media can have a powerful influence on the offline world. Major social-media companies that have a responsibility to address the spread of toxic content have instead exhibited reckless disregard for measures needed to curb hate, disinformation and extremism on their networks.

This is especially true during elections, as bad actors mislead voters by spreading disinformation and bigotry across platforms like Facebook, Instagram, TikTok, Twitter and YouTube. The results are felt not just in the polling booth but on the streets, where — as we saw on Jan. 6, 2021 — disinformation about election results can lead to real-world violence.

Change the Terms, a coalition of more than 60 civil- and consumer-rights organizations, developed a set of 15 priority reforms for social-media companies to implement ahead of the midterm elections that would fight algorithmic amplification of hate and lies, protect users across all languages and increase company transparency. Our coalition, of which Free Press is a founding member, then met with Meta, TikTok, Twitter and YouTube throughout the summer of 2022, calling on each company to implement these 15 priority reforms as soon as possible and to share more data about their enforcement practices around election integrity. Over the last several months, we continued to follow up with the companies in writing, yet we have received no substantive response from Meta, Twitter or YouTube.<sup>1</sup> *The Columbia Journalism Review*, Reuters and *The Washington Post* have all covered our efforts.<sup>2</sup>

Now, Free Press has reviewed the policies of these platforms against the Change the Terms demands to consider how prepared, both in writing and practice, the companies are for what lies ahead. **We have found that the problem is just as dire in advance of the 2022 U.S. midterms as it was during the nation's 2020 elections.**

## Topline Findings

- 1.** Meta, TikTok, Twitter and YouTube fail to provide sufficient data to show whether there are significant gaps in their policies' applications and enforcement, and have created a labyrinth of company commitments, announcements and policies that make it difficult to assess performance.
- 2.** Going by stated policies alone, Meta and TikTok appear to make the most convincing and comprehensive commitments. Yet Meta policies fully meet only two of the Change the Terms coalition's 15 demands and TikTok policies fully meet only one of 15 demands. Both Twitter and YouTube fail to fully meet any of the 15 demands. YouTube makes the fewest commitments in its policies.
- 3.** All four companies fail to close what they call "newsworthiness" or "public interest" exceptions that give prominent users a "get out of jail free" card and allow them to post as they choose. Every promising protective policy seems as though it could be circumvented with each platform's arbitrary "newsworthiness" or "public interest" exception.
- 4.** Neither of the video platforms — TikTok and YouTube — report denominators on violative videos to illustrate the prevalence, the user impressions or length of time videos were kept up before moderation action was taken to remove them.

Meta (the corporate parent of Facebook and Instagram), TikTok, Twitter and YouTube have each made public announcements about their U.S. election-integrity efforts this year.<sup>3</sup> Each company rolled out these announcements between Aug. 11 and Sept. 1, 2022. TikTok claimed that its Aug. 17 launch of its in-app center, which features videos encouraging viewers "to think critically about content they see online," would help "capture more of the midterms conversation over the summer months."<sup>4</sup>

These announcements — along with promises to connect users to credible information, stem the spread of mis- and disinformation and fully enforce their policies — came way too late. Election misinformation and disinformation are not anecdotal or seasonal. Lies — particularly the brand of election-denialism rhetoric that rose in 2020 — have been ubiquitous online for years, and this crisis has no end in sight. To treat "election-related" disinformation in particular as episodic ignores that it is present year-round and shapes beliefs and opinions that lead to harassment of election officials, and election-related hoaxes and violence.

Social-media companies make it nearly impossible to understand how they deal with this array of problematic content, whether through labeling and other friction, downranking or removal, and beyond. Attempting to gain clarity is like trying to find one's way through a forest of ever-changing policy updates, contradictory community standards, newsroom announcements, blog posts, Terms of Service, business centers, advertising centers, help or customer-support centers, and more. All of these companies' efforts are seemingly designed to distract the public, advertisers and regulators about the role each platform plays in amplifying hate and disinformation year-round.

In particular, there's an excess of redundant internal linking between company policies and announcements. This creates a loop back to the same sources where companies cite themselves as evidence of accountability, directing readers to a maze of hyperlinks that lead back to policies that have not been updated in years. This internal link-loop makes it impossible for readers, and researchers of this report, to find a consistent answer or get a clear sense of certain policies and which ones are applicable now.

**This tangle of conflicting policies and webpages — in addition to the patchwork of non-public internal guidelines given to content moderators — enables these companies to systematically break their commitments to keep people safe, leading researchers here to conclude that these policies are nothing more than empty promises. Meanwhile, white supremacists and hate groups continue to use these platforms to spread racist ideologies, fundraise, recruit and organize events that incite violence.**

Symptoms of these empty promises abound. While each of the companies has taken recent actions to remove some proliferators of hate, their policies are not equitably and consistently applied to all users, allowing other high-profile accounts to promote hate and lies without penalty.<sup>5</sup> Company audit reports typically lack context and fail to use denominators to understand the scale of reach or engagement of violative content, making these audits difficult for external researchers to assess. And when civil-society organizations request concrete information, the companies show an unwillingness to share details of their enforcement practices and the prevalence of various kinds of potentially violative content.

**Snapshot:** YouTube touts that it removed 4,496,933 violative videos from April to June 2022.<sup>6</sup> Yet the platform does not report what the ratio is to all videos that existed on the platform during that period. The closest relevant statistic YouTube provides is that “every day, people watch over a billion hours of video and generate billions of views.”<sup>7</sup> Using this data point, it’s unclear what fraction of videos or hours watched that YouTube actually finds to be violative and worthy of removal from the platform. YouTube provides no relevant data.

**Despite the extensive patchwork of policies and announcements from each company — purportedly pointing to civic-integrity safeguards — there’s little evidence that these policies are applied in an equitable and robust manner.**

Change the Terms engaged in months of direct conversations with the four major platforms and pressured the companies to enact 15 specific measures to reduce hate and disinformation.<sup>8</sup> (These demands, developed with the entire Change the Terms coalition as priority reforms that social-media companies should adopt, can be found in detail on page 8.) Yet when the coalition’s leadership asked for details and evidence of company action, we received no substantive responses from any of the platforms. Therefore, our primary method to identify gaps in policy application has been to identify one-off failures where violative content remains on their platforms.

**When social-media companies force civil society to employ “whack-a-mole” accountability, they dodge true commitment to holistic reforms, which would involve investment, reflection and acknowledgement of their role in shaping discourse and engagement.**

**The platforms’ promises of improvement amount to little more than public-relations exercises in a dysfunctional system where changes to policies can’t be checked comprehensively for accuracy and real-world impact.** The companies provide just enough information to seem credible but their reporting lacks the context needed to give external stakeholders a full picture.

# Defining the Problem

In recent years, white supremacists and conspiracy theorists have harnessed social media to sow chaos and bigotry online and in the real world. Meanwhile, online platforms give high-profile users a free pass to post election-denialism content and threaten election workers and voters. While making public displays of concern about these problems, tech companies put their profits above all else — and have failed to sufficiently protect human rights, safety and democracy.

In 2020, for instance, targeted online ads and other messaging discouraged users of color, particularly Black, Indigenous and Latinx people, from voting and participating in the U.S. census.<sup>9</sup> Meta and Twitter have consistently failed to remove or label as false content that discourages people from voting.<sup>10</sup> This content includes deception (lying about the time, place and manner of voting); calls for electoral boycotts from individuals with alleged sponsorship ties to foreign state actors; and voter intimidation or threats, such as claims that people will show up to polling locations with guns.<sup>11</sup> Meanwhile, social-media platforms have permitted white-supremacist militia groups to use their sites to organize armed demonstrations, including the Jan. 6 insurrection at the U.S. Capitol.<sup>12</sup>

Civil society — including Change the Terms — has urged major social-media companies to implement safety mechanisms ahead of the midterm elections.<sup>13</sup> Change the Terms launched the *Fix the Feed* campaign in April 2022, urging companies like Meta, TikTok, Twitter and YouTube to mitigate the harms caused by hate and lies on their platforms.<sup>14</sup> **Change the Terms has called for Big Tech to: 1) fix their algorithms to stem hate and disinformation; 2) bolster civic-**

**integrity efforts year-round to apply policies equitably and across languages; and 3) provide transparency and disclosures about the companies' practices.**

**Nearly every day we learn about another tech company creating or exacerbating deep societal harms. Meta, TikTok, Twitter and YouTube have a moral duty to address this and the harms inflicted on their users.**

Social-media platforms have the resources to invest at scale to disrupt the spread of hate and anti-democratic disinformation — especially for marginalized users who are part of various protected classes. Yet these companies aren't just failing to remove or moderate this content. In fact, platforms like Meta and YouTube are amplifying hate and disinformation.<sup>15</sup>

Although tech companies had promised to fight disinformation and hate on their platforms this fall, **there is a notable gap between what the companies say they want to do and what they actually do in practice. In sum, platforms do not have sufficient policies, practices, AI or human capital in place to materially mitigate harm ahead of and during the November midterms.**

We cannot take these companies at their word. We need transparent records of their implementation of safety mechanisms and application of their own policies. We've launched this analysis for civil-society groups, researchers and the public to better understand the gap between what companies promise and what they do in practice.

# What Platforms Claim They're Doing

Free Press, which is a founding Change the Terms member, conducted a landscape analysis of Meta, TikTok, Twitter and YouTube. We compared how they perform — according to their stated policies, blog posts and Terms of Service — against Change the Terms' demands. Our full analysis, by platform and with citations to platform-specific policies, appears [here](#).

## We found:

- **Meta's** policies and announcements — on their surface — seem promising. But they are just that: promises. We've found ample evidence of the company's failures to apply its policies equitably or at scale. At best, Meta fully meets only two of the 15 demands. Meta claims (without evidence) that it's now emphasizing content moderation across languages. There is also no mention of Facebook Protect, a security program the company launched and touted during the 2020 election cycle.
- **TikTok** has a shorter track record, only coming onto the scene in force as a major social-media platform in the last few years. Nevertheless, there is evidence of misinformation on the platform as well as a track record of failures in regional-election contexts.<sup>16</sup> TikTok claims that it does not recommend content that is questionably violative and under review.
- **Twitter** fails to fully meet any of Change the Terms' demands. Twitter policies lack detail, and there are discrepancies between Twitter election-related blog posts and Twitter policies in the Terms of Service. It's not clear if users are expected to agree to the terms that are listed only in the election-policy post but not in the general Terms of Service that users consent to when joining Twitter. This begs the question of what policies Twitter is actually enforcing.
- **YouTube** has the largest gaps in policy protections. The company lacks transparency on its approach to violative content. There are also few specifics on moderation and enforcement practices (such as the existence of civic-integrity teams, moderation across languages, etc.).

The table on the following pages shows how each company performs across the 15 Change the Terms demands. Company scores range as follows:











- 😊 = the company meets the demand in a stated policy
- 😐 = the company insufficiently or incompletely references the demand in a stated policy
- 😞 = the company fails to meet the demand
- \* = instances when it was impossible to assess a company's performance given insufficient transparency

# Grading the Companies

				
Discontinue amplification of, and actively downrank or remove hateful content and disinformation	 INSUFFICIENT*	 INSUFFICIENT	 INSUFFICIENT	 INSUFFICIENT
Develop temporary virality circuit breakers	 INSUFFICIENT*	 INSUFFICIENT*	 INSUFFICIENT*	 INSUFFICIENT*
Remove high-volume producers of election disinformation from recommendations	 INSUFFICIENT*	 INSUFFICIENT	 INSUFFICIENT*	 INSUFFICIENT*
Increase hiring of human moderators across languages	 INSUFFICIENT*	 INSUFFICIENT*	 INSUFFICIENT	 FAILS
Civic-integrity teams active 365 days/year across language	 INSUFFICIENT	 INSUFFICIENT	 FAILS	 INSUFFICIENT
Remove content that calls for or incites violence at events	 INSUFFICIENT	 INSUFFICIENT	 INSUFFICIENT	 INSUFFICIENT
Ban calls to arms	 YES	 YES	 INSUFFICIENT	 INSUFFICIENT
Remove content that could inspire violence such as attacks on election workers	 INSUFFICIENT	 INSUFFICIENT	 INSUFFICIENT	 INSUFFICIENT



# Grading the Companies

				
Apply third-party fact checkers for political ads	 YES	N/A TikTok does not run political ads	N/A Twitter does not run political ads	 INSUFFICIENT
Remove newsworthiness/ public-interest exceptions	 FAILS	 FAILS	 FAILS	 FAILS
More efficient labeling of misleading content about elections, which includes increased resourcing for fact-checking consultants	 INSUFFICIENT	 INSUFFICIENT	 INSUFFICIENT	 FAILS
Provide quarterly transparency and audit reports on enforcement practices	 INSUFFICIENT	 INSUFFICIENT	 INSUFFICIENT	 FAILS
Disclosures on algorithmic business models and insights into content which platforms themselves amplify	 FAILS	 FAILS	 FAILS	 FAILS
Disclosures on life cycle of problematic content before action is taken to remove or downrank content	 FAILS	 FAILS	 FAILS	 INSUFFICIENT
Researcher and watchdog journalist access	 INSUFFICIENT	 INSUFFICIENT	 INSUFFICIENT	 INSUFFICIENT

# Gaps in Platform Performance to Protect Users

Below, we provide examples, drawn from our own research or external reports, of gaps or systemic failures across the core demands Change the Terms has made of the companies.

## Fixing the Algorithm & Downranking Violative Content

Change the Terms has found that all four companies consistently fail to remove content that violates their existing Community Guidelines and Terms of Service. In some instances, companies are actually amplifying hateful, harassing and misleading content to maximize engagement and profit.<sup>17</sup> **We've seen this in previous election cycles: The companies have the ability to turn safety functions on yet fail to do so early enough, long enough, in non-English languages, or at all.**

**Snapshot:** Recent research from the NYU Stern School of Business found that YouTube often amplifies disinformation and hate with algorithms designed to show users content that provokes negative emotions — thus boosting incendiary and misleading video content.<sup>18</sup> In an Aug. 26 meeting with YouTube, company executives assured Change the Terms: “We were the first and we continue to be the platform that enforces against ‘Big Lie’ content. We remove content even under extreme political pressure. We aren’t perfect and sometimes people need to flag it for us, but we are the only platform that continues to ban and take down ‘Big Lie’ content because of how it seeps into the political discourse. That is a policy decision that

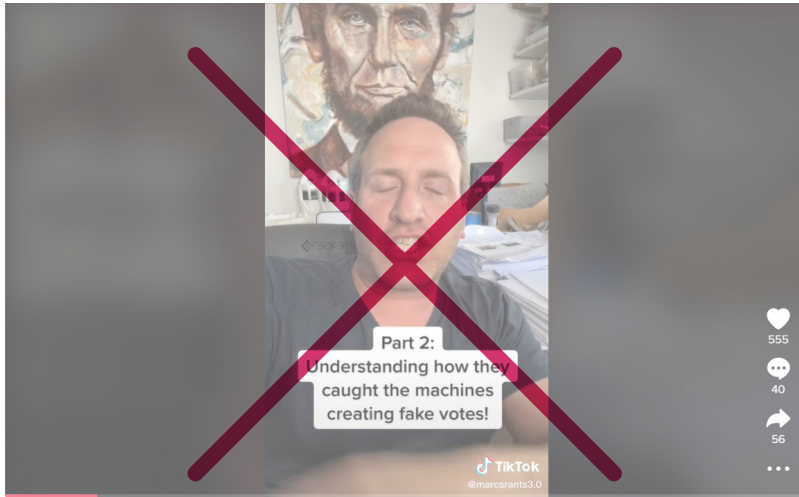
we made that remains intact, unlike some of the other platforms.” **Findings from NYU Stern and YouTube’s stated commitment appear in conflict: How is YouTube enforcing against “Big Lie” content while also amplifying hate and lies?**

The “Big Lie” spreads across platforms; examples also abound on Meta and Twitter, where hateful and misleading posts pack a one-two punch: encouraging violence against election workers because of demonstrably false claims about stealing the 2020 election from Donald Trump.<sup>19</sup>

Violative Election-Related Content Is Slipping Through the Cracks: Content that violates platforms’ Terms of Service — particularly content about the electoral process and democracy writ large — does not have an “end” time after an election occurs. Therefore, Change the Terms has urged Big Tech to keep their election-integrity efforts in place through at least the first quarter of 2023. Change the Terms is also calling on social-media companies to announce their timeline and plans for the 2024 election without delay.

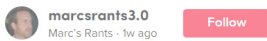
Ahead of the election this November, we have noted that the four companies fail to consistently remove violative content.

## Snapshots



#trump #trumptok #trump2024 #trumpwontheelection

original sound - Marc's Rants



On TikTok, [this post claimed](#) that Trump actually won the 2020 election, and that voting machines were rigged. Although the post has been removed, it received more than 6,000 views while it was up. The creator of this post had more than 12,000 followers at the time and has shared countless videos featuring demonstrably false allegations of election fraud as well as COVID-19 vaccine disinformation. The creator account, marcsrants3.0, was eventually removed by TikTok. Since TikTok's removal of the marcsrants3.0 account, the creator has iterated on his account profile, making marcsrants version 4 and marcsrants version 5. He now operates marcsrantsv6, where he continues to spread similar disinformation.




Meta should prioritize enforcement review of possible harassment of election workers — as seen in the following [example](#) — by expediting human review of posts that mention specific election workers and/or specific election offices in text or imagery. Despite Meta's assurance that this post had been taken down due to violations of its bullying and harassment policies — as well as its disinformation policies under voter or census interference — versions of this content remain on the platform. We have asked Meta why this post continues to fall through the cracks and whether this relates to a decrease in election-integrity staff, but have received no response.<sup>20</sup>



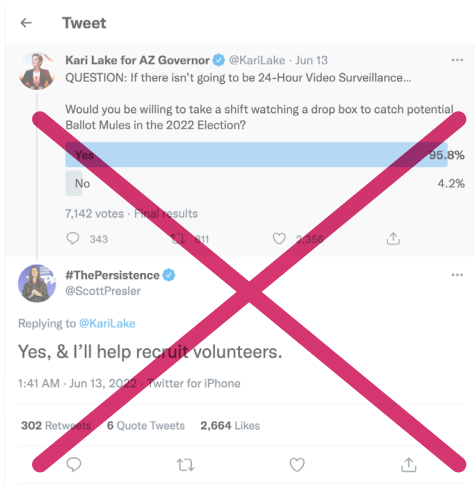
## Snapshots

Meta should continue to remove and downrank content promoting the “Big Lie” that the 2020 election was stolen, as such claims are likely to discourage voters from participating in the upcoming election cycle. In some instances, content appears likely to violate various platforms’ policies, and we believe it is essential for companies to explain clearly why content remains on their platforms. Two examples appear below:



← **This post** on Facebook, by the Philadelphia Trumpet, spreads the false stolen-election narrative and claims the Jan. 6 insurrection was a hoax, and that the House’s Jan. 6 committee hearings are theater. This post remains up and unlabeled on Facebook, despite the company’s policy calling for removal of “misinformation that is likely to directly contribute to a risk of interference with people’s ability to participate in [voting or the census].” Did Meta overlook this post or did it review and find that the post was non-violative?

On Twitter: In the example below, Arizona gubernatorial candidate Kari Lake is actively recruiting her followers to monitor ballot drop-box locations in an attempt to intimidate election officials and voters who may wish to vote by mail.



→ **This tweet** remains up with over 2,200 likes, though it appears to violate Twitter’s policy against using its services “for the purpose of manipulating or interfering in elections or other civic processes,” which “includes posting or sharing content that may suppress participation or mislead people about when, where, or how to participate in a civic process.” Clarifying responses from Twitter would help outside reviewers understand how the company applies its policies. As it is, outside reviewers can only guess at company practices.

### VIP & Candidate Accounts Receive Special Treatment:

Social-media companies selectively and sparingly enforce their policies on accounts for celebrities, politicians and other high-profile users. As a result, violative content reaches high volumes of people. Every single platform fails in this category.

**Snapshots:** Twitter’s policy states that it’s often in the “public-interest” to “leave up a tweet from an elected or government official that would otherwise be taken down.” We believe it is not in the public interest to allow content to remain up when that content enables and incites violence or spreads lies, even if that is

somehow “newsworthy”. Powerful people should have to abide by the same rules as any other user, especially when posts threaten, endorse or legitimize violence.

Twitter also claims to ensure such tweets aren’t “algorithmically recommended by Twitter.” Without further disclosures from Twitter, it’s difficult to ascertain whether this is true. Regardless, it’s crucial that Twitter hold influential users accountable: Even without the amplification, such users are using Twitter’s platform to reach wide audiences.



According to Twitter policy, “the public-interest exception does not mean that any eligible public official can Tweet whatever they want, even if it violates the Twitter Rules.” A clear guiding principle in this policy affirms this understanding of how to treat “content that threatens or glorifies violence,” stating that in such cases “public-interest exceptions are unlikely.”

YouTube has several policies, including “harmful or dangerous content policies” and “election misinformation” policies, that limit related postings. However, the platform also has a carve-out for content that is educational, documentary, scientific or artistic. YouTube CEO Susan Wojcicki has said publicly that this

exception applies to politicians who are not subject to the platform’s rules.<sup>21</sup>

### **Bolstering Civic-Integrity Teams & Protecting Users Across Languages**

The companies do not provide adequate resources to live up to their commitments around civic integrity and moderation across languages. At best, the companies simply state their commitment to civic integrity across languages and to mitigating election lies — but not much more.<sup>22</sup>

**Meta cuts its civic-integrity & human-rights teams:** Meta eliminated its Responsible Innovation team this fall.<sup>23</sup> According to the company’s internal 2020 Civil Rights Audit, “Responsible Innovation is a positive development [and] is worth mentioning here because the trainings, tools, and processes that team is looking to build may help surface civil rights considerations. The Auditors also recommend that Facebook add personnel with civil rights expertise to this team.”<sup>24</sup>

In the same month, Meta combined multiple civic-integrity teams tasked with moderating different kinds of content, a move the company says will help improve efficiency but which two sources told Axios were motivated by cost-cutting considerations.<sup>25</sup> With elections happening this year in countries including Brazil, Kenya and the United States, these cuts underscore how Meta fails to prioritize civic integrity across languages and to mitigating election lies.

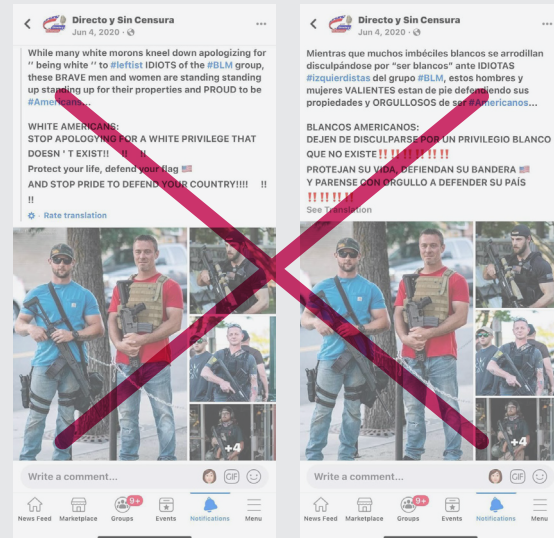
In the last several years, Change the Terms, members of Congress and other civil-society groups and researchers have asked for details about the four major companies’ non-English-moderation practices. Stakeholders have sought information on the number of people monitoring non-English languages, the training they receive, their working conditions and

employment statuses. The companies have declined to answer these questions. We have long urged the platforms to equitably moderate non-English content with a speed and accuracy that are at least equivalent to their English-language content-moderation efforts. Language disparities in moderation abound.

## Snapshots



Meta’s response to Spanish-language disinformation and other violative posts has been sluggish at best. In one instance, it took 11 months for Facebook to action a [Spanish-language post](#) that glorified white supremacists and referenced a nonexistent threat to the United States — a post it had already actioned in English.



A YouTube election blog post stated: “As our systems connect viewers to authoritative content and limit the spread of harmful misinformation, we’re also removing election content that violates our policies. This includes misleading voters on how to vote, encouraging interference in the democratic process, inciting violence, or advancing certain types of elections misinformation.” Yet Media Matters for America released several reports that found dozens of “Spanish-language YouTube videos spreading misinformation about U.S. elections [..despite the fact that] YouTube has explicit election misinformation policies and has committed to combating Spanish-language election misinformation ahead of the midterms.”<sup>26</sup>

**Empty promises on political ad fact-checking:** Research by Global Witness and NYU Tandon reveals that Meta and TikTok approve ads containing misleading and false election disinformation. This runs contrary to Meta’s stated commitment to fact-checking these ads and, in TikTok’s case, its stated commitment to refusing to run political ads altogether.<sup>27</sup>

## Transparency & Disclosures

**To be clear:** Much of our comparative research was inconclusive due to the companies’ failures to offer genuine transparency. We were provided little meaningful insight into: 1) the companies’ moderation and enforcement practices; 2) the results of their algorithmic decision-making when it comes to preferential or downranked content; and 3) details around human moderators across languages. This lack of transparency calls into question the companies’ commitments to curbing the harms that flow from their platforms, business models and their own negligence in protecting users around the globe.

**Snapshots:** The YouTube webpage headlined “Building a more responsible platform over the years” sets out to share the platform’s progress and impact.<sup>28</sup> The page reads: “At YouTube, we’re committed to building a responsible platform our users, creators, and artists can rely on, and over the years, we’ve made huge progress,

making our community safer. But we know that work is never finished - and we’ll continue to invest in the teams, technology, and product features it takes to make sure YouTube continues to be a place where people come to be informed, inspired, or just plain delighted.”

But the overall timeline of YouTube’s impact across all categories ends on Sept. 29, 2021. If one clicks on subcategories, those timelines end at different points: e.g., the “Enforcing our policies” category timeline ends in February 2019; the “Standing up to hate and harassment” timeline ends in April 2021; the “Fighting misinformation” timeline ends in September 2021; and so on, with haphazard and inconsistent endpoints falling no later than 2021.

Likewise, when we investigated whether Twitter removes content that could inspire violence such as doxxing or attacks on election workers, the company first directed us to the “Our approach to the 2022 elections” blog post. However, there was no information about doxxing against election workers. We attempted to dig further across other links from Twitter and were directed to a “violent threats” policy last updated in March 2019.

From there, we were directed to Twitter’s “private information” policy, which explicitly prohibits doxxing but does not name election workers as a group protected under this policy. When we tried to look further into this policy via a hyperlink to the Twitter Rules, we were simply rerouted back to the original private-information media-policy page, in what became a frustratingly repetitive loop.

**TikTok & YouTube make promises about researcher access:** TikTok announced in July that it will allow researchers to delve into its data, evaluate its content and test its moderation system.<sup>29</sup> And YouTube invited researchers to apply for access to its global data. However, promising access is different than providing access.<sup>30</sup> Change the Terms will continue to monitor how both companies open up their data to researchers.

# Recommendations for How Big Tech Can Disrupt Hate and Lies

These companies have failed to take proactive or even mitigating action to stem the spread and engagement of hate, disinformation and extremism on their platforms.

While they claim to have crafted and enforced new policies addressing the spread of such toxic content, these claims are difficult for independent auditors to verify. The companies' websites are tangles of contradictory policies and standards that are difficult to unravel. Reporters covering the technology sector should take nothing from the platforms at face value. Every claim must be backed by empirical evidence and a full-field view of its impact.

With the U.S. midterm elections imminent, online disinformation has continued to proliferate at an alarming pace. The need for the platforms to do more is growing more urgent by the day. Their failure to take our recommendations seriously will be felt most acutely by voters, and could lead to more of the kinds of violence seen at the U.S. Capitol on Jan. 6, 2021.

The platforms must stop making empty promises — and show the courage and commitment necessary to fix their feeds.

## **We continue to call on platforms to:**

- 1.** Stop amplifying hate and disinformation content and implement algorithms without discrimination.
- 2.** Protect people equally around the world and across languages through increased resourcing for civic-integrity teams year-round.
- 3.** Boost transparency about company business models and moderation-and-enforcement practices, ensuring access to data for external researchers and journalists.



# Methodology and Acknowledgements

To develop this research report, Free Press researched 2022 election blog posts, Terms of Service, and policies published by Meta, TikTok, Twitter and YouTube. We conducted comparative analysis between public company documents and the Change the Terms demands through the coalition's Fix the Feed campaign. We then pulled external examples of instances where companies did not take action on violative content.

Where there was no evidence to convey that a company met one of the demands, we conducted lateral research beyond blog-post announcements, Terms of Service and policies for each company. This lateral research included other company-specific websites and even external reporting and research.

Free Press wrote and researched this report. The report was led by Nora Benavidez, with significant research support and editing from Rose Lang-Maso, both of Free Press. The development and production of this report could not have happened without support from Free Press colleagues Craig Aaron, Jessica J. González, Timothy Karr and S. Derek Turner. This report was edited by Amy Kroin and designed by Erika Rose of Limelight Creative with support from Dutch Cosmian of Free Press.

Change the Terms leadership, including co-chairs from other organizations that are part of the coalition, reviewed the report in advance of publication. Free Press thanks the Change the Terms partners for their continued collaboration and support.

Free Press is a U.S.-based nonpartisan organization that is a co-founder of the Change the Terms coalition. Free Press believes that positive social change, racial justice and meaningful engagement in public life require equitable access to technology, diverse and independent ownership of media platforms, and journalism that holds leaders accountable and tells people what's actually happening in their communities. Free Press is one of the leaders of the nationwide fight against hate and disinformation. Learn more at [freepress.net](https://freepress.net).

# Endnotes

- 1 TikTok responded in writing to Change the Terms' demands; however, the response was lackluster at best, reiterating much of the company's civic-integrity blog post; see Han, *infra* n. 3.
- 2 Naomi Nix, "Inside the Civil Rights Campaign to Get Big Tech to Fight the 'Big Lie,'" *The Washington Post*, Sept. 22, 2022, <https://www.washingtonpost.com/technology/2022/09/22/midterms-elections-social-media-civil-rights/>; Mathew Ingram, "The Social Media Platforms, the 'Big Lie,' and the Coming Elections," *Columbia Journalism Review*, Sept. 27, 2022, [https://www.cjr.org/the\\_media\\_today/the-social-media-platforms-the-big-lie-and-the-coming-elections.php](https://www.cjr.org/the_media_today/the-social-media-platforms-the-big-lie-and-the-coming-elections.php); Avi Asher-Shapiro, "Big Tech Accused of Failing to Tackle U.S. Midterm 'Lies,'" *Reuters*, Oct. 6, 2022, <https://www.reuters.com/article/usa-election-tech/big-tech-accused-of-failing-to-tackle-u-s-midterm-lies-idUKL8N30Z6SJ>
- 3 Nick Clegg, "How Meta Is Planning for the 2022 US Midterms," *Meta*, Aug. 16, 2022, <https://about.fb.com/news/2022/08/meta-plans-for-2022-us-midterms/>; "Our Approach to the 2022 US Midterms," *Twitter*, Aug. 11, 2022, [https://blog.twitter.com/en\\_us/topics/company/2022/-our-approach-to-the-2022-us-midterms](https://blog.twitter.com/en_us/topics/company/2022/-our-approach-to-the-2022-us-midterms); Leslie Miller, "The 2022 U.S. Midterm Elections on YouTube," *YouTube*, Sept. 1, 2022, <https://blog.youtube/news-and-events/the-2022-us-midterm-elections-on-youtube/>; Eric Han, "Our Commitment to Election Integrity," *TikTok*, Aug. 17, 2022, <https://newsroom.tiktok.com/en-us/our-commitment-to-election-integrity>
- 4 See Han, *supra* n. 3
- 5 Jeff Horwitz, "Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That's Exempt," *The Wall Street Journal*, Sept. 13, 2021, <https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353>
- 6 YouTube, "YouTube Community Guidelines Enforcement," <https://transparencyreport.google.com/youtube-policy/removals?hl=en>
- 7 YouTube, "YouTube for Press," <https://blog.youtube/press/>
- 8 See *infra*, n. 14
- 9 Samuel Woolley and Mark Kumleben, "At the Epicenter: Electoral Propaganda in Targeted Communities of Color," *Protect Democracy*, Nov. 2021, <https://protectdemocracy.org/project/understanding-disinformation-targeting-communities-of-color/#section-1>; Brian Contreras and Maloy Moore, "What Facebook Knew About its Latino-Aimed Disinformation Problem," *The Los Angeles Times*, Nov. 16, 2021, <https://www.latimes.com/business/technology/story/2021-11-16/facebook-struggled-with-disinformation-targeted-at-latinos-leaked-documents-show>
- 10 Ian Vandewalker, "Digital Disinformation and Vote Suppression," *The Brennan Center for Justice*, Sept. 2, 2020, <https://www.brennancenter.org/our-work/research-reports/digital-disinformation-and-vote-suppression>
- 11 *Id*
- 12 Cristiano Lima, "Jan. 6 Panel Spotlights Twitter's Role in Insurrection," *The Washington Post*, July 13, 2022, <https://www.washingtonpost.com/politics/2022/07/13/jan-6-panel-spotlights-twitter-role-insurrection/>
- 13 Free Press is an anchoring co-chair organization for Change the Terms, a coalition of over 60 civil- and consumer-rights organizations. Change the Terms launched in 2017 with the goal of moving technology companies to adopt its model policies to blunt the spread of hate and extremism on their platforms.
- 14 Change the Terms, "Fix the Feed," <https://www.changethetterms.org/fix-the-feed>
- 15 Scott Pelley, "Whistleblower: Facebook Is Misleading the Public on Progress Against Hate Speech, Violence, Misinformation," *CBS News*, Oct. 4, 2021, <https://www.cbsnews.com/news/facebook-whistleblower-frances-haugen-misinformation-public-60-minutes-2021-10-03/>; Paul Barrett and Justin Hendrix, "A Platform 'Weaponized': How YouTube Spreads Harmful Content – And What Can Be Done About It," *NYU Stern Center for Business and Human Rights*, June 10, 2022, [https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/62a38fc022745a7274601da0/1654886337000/NYU+CBHR+YouTube\\_Final\\_June10.pdf](https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/62a38fc022745a7274601da0/1654886337000/NYU+CBHR+YouTube_Final_June10.pdf)
- 16 Tiffany Hsu, "On TikTok, Election Misinformation Thrives Ahead of Midterms," *The New York Times*, Aug. 14, 2022, <https://www.nytimes.com/2022/08/14/business/media/on-tiktok-election-misinformation.html>

- 17 Paul Barrett and Justin Hendrix, "A Platform 'Weaponized': How YouTube Spreads Harmful Content – And What Can Be Done About It," NYU Stern Center for Business and Human Rights, June 10, 2022, [https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/62a38fc022745a7274601da0/1654886337000/NYU+CBHR+YouTube\\_Final\\_June10.pdf](https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/62a38fc022745a7274601da0/1654886337000/NYU+CBHR+YouTube_Final_June10.pdf) ("The platforms collect the lion's share of their considerable revenue from advertisers seeking the attention of consumers. To demonstrate that users remain glued to their screens, platforms prioritize 'engagement,' a metric reflecting user watch time and the volume of 'likes,' shares, retweets, and comments. But here's the problem: Content that tends to heighten engagement often does so because it provokes emotions like anger, fear, and resentment. So, when software engineers design algorithms to rank and recommend content, the automated systems favor posts that stir these negative emotions.")
- 18 Id
- 19 Brennan Center, "Election Officials Under Attack," June 16, 2021, <https://www.brennancenter.org/our-work/policy-solutions/election-officials-under-attack>. ("78 percent of election officials surveyed by the Brennan Center said that social media, where mis- and disinformation about elections both took root and spread, has made their job more difficult; 54 percent said they believe that it has made their jobs more dangerous.")
- 20 "Meta Quieter on Election Misinformation as Midterms Loom," Associated Press, Aug. 5, 2022: <https://apnews.com/article/2022-midterm-elections-technology-social-media-voting-0ab5375951df71093d240a6631edb9da>
- 21 Michael Grosack, "A Look at How We Treat Educational, Documentary, Scientific and Artistic Content," YouTube, Sept. 17, 2020: <https://blog.youtube/inside-youtube/look-how-we-treat-educational-documentary-scientific-and-artistic-content>; Nina Gologowski, "YouTube CEO Says Politicians Are Exempt from Content Rules," HuffPost, Sept. 25, 2019 [https://www.huffpost.com/entry/politicians-exempt-from-youtube-rules-ceo-says\\_n\\_5d8ba3c4e4b01c02ca627f81](https://www.huffpost.com/entry/politicians-exempt-from-youtube-rules-ceo-says_n_5d8ba3c4e4b01c02ca627f81)
- 22 Kari Paul, "Disinformation in Spanish Is Prolific on Facebook, Twitter and YouTube Despite Vows to Act," *The Guardian*, Oct. 6, 2022, <https://www.theguardian.com/media/2022/oct/06/disinformation-in-spanish-facebook-twitter-youtube>
- 23 Jeff Horwitz, "Facebook Parent Meta Platforms Cuts Responsible Innovation Team," *The Wall Street Journal*, Sept. 8, 2022, <https://www.wsj.com/articles/facebook-parent-meta-platforms-cuts-responsible-innovation-team-11662658423>
- 24 Laura Murphy, "Facebook's Civil Rights Audit — Final Report," July 8, 2020, <https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf>
- 25 Mack DeGeurin, "Meta's Combining Teams Responsible for Moderating Ads and User Content," Gizmodo, Sept. 13, 2022, <https://gizmodo.com/meta-facebook-ads-user-content-1849531750>; Sara Fisher, "Scoop: Meta Merging Content Moderation Teams for Ads and User Posts," Axios, Sept. 13, 2022, <https://www.axios.com/2022/09/13/meta-merging-content-moderation-teams-ads-posts>
- 26 M. Estrada, "Dozens of Spanish-Language Election Fraud YouTube Videos Demonstrate the Platform's Toothless Policies," Media Matters for America, Oct. 10, 2022, <https://www.mediamatters.org/spanish-language-media/dozens-spanish-language-election-fraud-youtube-videos-demonstrate-platforms>; see also YouTube, "Violent Or Graphic Content Policies," <https://support.google.com/youtube/answer/2802008?hl=en>; YouTube, "Election Misinformation Policies," <https://support.google.com/youtube/answer/10835034?hl=en>; Cristiano Lima, "Hispanic Caucus Pushes YouTube Chief for Data on Misinformation in Spanish," *The Washington Post*, April 29, 2022, <https://www.washingtonpost.com/politics/2022/04/29/hispanic-caucus-pushes-youtube-chief-data-spanish-language-misinformation/>; Brian Fung, "Google Releases Details on How It Will Combat Misinformation in Advance of Midterm Elections," CNN, Sept. 1, 2022, <https://www.cnn.com/2022/09/01/tech/google-youtube-election-misinformation/index.html>
- 27 "TikTok and Facebook Fail to Detect Election Disinformation in the US, While YouTube Succeeds," Global Witness, Oct. 21, 2022, <https://www.globalwitness.org/en/campaigns/digital-threats/tiktok-and-facebook-fail-detect-election-disinformation-us-while-youtube-succeeds/>
- 28 YouTube, "A History of YouTube's Responsibility Efforts: Building a More Responsible Platform Over the Years," <https://www.youtube.com/howyoutubeworks/progress-impact/timeline/>
- 29 Emma Roth, "TikTok to Provide Researchers with More Transparency as Damaging Reports Mount," *The Verge*, July 27, 2022, <https://www.theverge.com/2022/7/27/23280406/tiktok-researchers-api-transparency-damaging-reports-china>
- 30 YouTube, "YouTube Researcher Program," <https://research.youtube/>