FACEBOOK VS. HATE

AN ANALYSIS OF FACEBOOK'S WORK TO DISRUPT ONLINE HATE AND THE PATH TO FULLY PROTECTING USERS

BY CARMEN SCURATO
SENIOR POLICY COUNSEL, FREE PRESS

SEPTEMBER 2019

Free Press

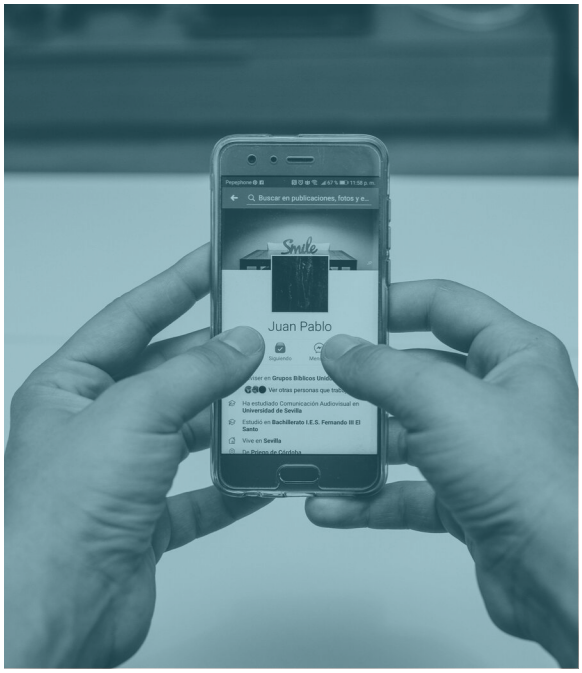
> We did not anticipate all of the risks from connecting so many people.
>
> We need to earn back trust.

SHERYL SANDBERG[1]
COO, FACEBOOK
January 2019

BACKGROUND & INTRODUCTION

White supremacists and related organizations are using platforms like Facebook to coordinate both online and offline attacks against women, people of color, immigrants, religious minorities, LGBTQIA people, and people with disabilities. And platform companies have long refused to acknowledge their responsibility to ensure the safety of their users against such attacks.

A coalition of racial-justice and civil-rights groups launched **Change the Terms**[2] on Oct. 25, 2018, with a set of recommended corporate policies and terms of service to reduce hateful activities.[3] The goals of the campaign are to crack down on hateful activities across online platforms and to ensure that the policies are enforced in a transparent, equitable and culturally relevant way.



"Internet companies must stop ignoring the racism and other forms of hate that are prevalent on their platforms and acknowledge that the hateful discourse of the few silences the speech of the marginalized many."

—Jessica J. González and Carmen Scurato, Colorlines, Oct. 25, 2018

The online platforms' failure to address these hateful activities silences the speech of the targeted groups, curbs democratic participation, and threatens people's real-life safety and freedom.

This paper examines the extent to which Facebook has changed its terms since October 2018 and makes pointed recommendations for further improvements. Over the past year, the company has taken several steps that have begun to align its policies with Change the Terms' recommendations. For example, the company has (1) prohibited content that explicitly praises white nationalism; (2) enforced its policies against dangerous individuals and organizations, resulting in the ban of several white supremacists; (3) implemented changes to its appeals process; and (4) worked to update the method its content reviewers use to analyze hateful content.

Facebook is still far from adopting the full set of recommended corporate policies specifically in the areas of enforcement, transparency, and evaluation and training. Yet we must acknowledge that it's marshaled more resources toward addressing hate on the platform than it has in previous years.

This report compares Change the Terms' recommended corporate policies with Facebook's Community Standards, terms of service, recent enforcement actions, transparency reports and civil-rights audits.[4] While considering both Facebook's progress and where it's falling short, this paper points out specific steps the company must take to better align with Change the Terms' recommendations and effectively address online hate.

More than 50 CIVIL-RIGHTS, HUMAN-RIGHTS, TECHNOLOGY-POLICY, AND CONSUMER-PROTECTION ORGANIZATIONS have signed on in support of the recommended policies for corporations to adopt and implement to reduce hateful activities on their platforms.

Learn more at changetheterms.org.

DEFINITION OF HATEFUL ACTIVITIES AND HATE SPEECH

Change the Terms defines hateful activities as follows—

> "activities that incite or engage in violence, intimidation, harassment, threats, or defamation targeting an individual or group based on their actual or perceived race, color, religion, national origin, ethnicity, immigration status, gender, gender identity, sexual orientation, or disability."[5]

Facebook's Community Standards define hate speech as—

> "a direct attack on people based on what we call protected characteristics — race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability. We also provide some protections for immigration status. We define attack as violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation." [6]

Though these two definitions might seem similar, Facebook's requirement that the posted content involve a "direct attack" is much narrower than Change the Terms' definition. The company characterizes this definition as "nuanced" and separates hate speech into different categories.

By choosing to act against only the most egregious attacks, Facebook is shielding itself from responsibility when it fails to remove hate speech from its platform.

CHANGE THE TERMS: MODEL CORPORATE POLICIES

Model Policy #1: Terms of service and acceptable-use policies

Change the Terms asserts that terms of service "should, at a minimum, make it clear that using the service to engage in hateful activities on the service or to facilitate hateful activities off the service shall be grounds for terminating the service for a user."[7] We recommend that platforms adopt model language stating that "users may not use these services to engage in hateful activities or use these services to facilitate hateful activities engaged in elsewhere, whether online or offline."[8]

HOW IS FACEBOOK DOING?

Facebook's Community Standards make it clear that the company removes content from the platform based on three "tiers of attacks" that are broken down by severity, with tier one as the most severe since it involves dehumanizing speech or calls for violence.[9]

Yet the exact penalty process for violating Facebook's hate-speech policy isn't noted in the company's Community Standards or elsewhere. It's well known that the company puts individuals in "Facebook jail" (blocking access to Facebook for a set number of hours or days) and has deplatformed both individuals and groups. This penalty process has drawn "criticism by the civil-rights community for its lack of transparency and seemingly disproportionate and/or unjustified penalties."[10]

Facebook's events policy,[11] which it updated in July 2019, now prohibits calls to action or statements of intent to bring weapons to an event or location to intimidate or harass vulnerable individuals or members of specific protected groups. However, enforcement of this policy has been lackluster.

March 2019

On March 27, 2019, Facebook implemented a policy barring white-nationalist and white-separatist content. The ban is on content[12] that expresses "praise, support and representation of white nationalism and white separatism."

This new policy shows the company beginning to align with the purpose and goals of Change the Terms. However, because Facebook's policy focuses solely on explicit representations of white nationalism and white separatism, there are many pieces of white-supremacist content that remain on the platform, including content that uses hateful slogans and symbols.[13]

RECOMMENDATIONS

This updated ban on white-nationalist and white-separatist content is present only in Facebook's Newsroom. It needs to be part of the company's Community Standards.

Beyond refining its Community Standards, Facebook must better define its penalty structure. The current provision — "If we determine that you have clearly, seriously or repeatedly breached our Terms or Policies, including in particular our Community Standards, we may suspend or permanently disable access to your account"[14] — fails to provide clarity to its users or those monitoring hateful activities. This provision also doesn't offer any real accountability to its users.

Change the Terms is also calling on Facebook to revise its definition of national origin, remove the "humor exception" to hate speech, and "prohibit calls for harassment or efforts to organize harassment campaigns through the platform."[15]

Model Policy #2: Enforcement

Change the Terms recommends that internet companies have enforcement strategies that adequately reflect the scope of hateful activities on their platforms. There are several specific recommendations,[16] including that users be allowed to flag hateful activities and that companies create "trusted flagger" programs.

In addition, Change the Terms recommends that content moderation involve a combination of technological solutions as well as human review, with regular audits of both the technology and human efforts. Finally, government actors should not be allowed to use these tools to flag content.

HOW IS FACEBOOK DOING?

Overall, Facebook's enforcement of its current policies remains lackluster, with the company making its decisions in an inconsistent and sporadic manner. Though Facebook conducts a trusted-flagger program, which fast tracks review and highlights content flags by vetted organizations and individuals, it's not clear how effective it is.

The following are examples of enforcement actions that Facebook has taken over the course of the last year, where it's deplatformed several individuals and removed pages under its "Dangerous Individuals and Organization Policy." [17]

Oct. 2018

On Oct. 31, 2018, Facebook "banned content linked to violent neo-fascist group Proud Boys,[18] citing the organization's promotion of hate speech. The Proud Boys and its founder Gavin McInnes were removed from Facebook and Instagram."

Feb. 2019

On Feb. 5, 2019, Facebook removed 89 pages [19] — including 22 connected to Infowars host Alex Jones — under its new policy aimed at cracking down on accounts that repeatedly violate the company's Terms of Service.

May 2019

On May 2, 2019, Facebook banned Alex Jones, Infowars, Milo Yiannopoulos, Paul Joseph Watson, Laura Loomer, Paul Nehlen and Louis Farrakhan. All were deplatformed from both Facebook and Instagram. [20]

June 2019

On June 30, 2019, Facebook released its second civil-rights audit report, which states that Facebook examines "online and offline activity" to identify dangerous individuals and organizations. This approach is not explicitly noted in the company's Community Standards. [21]

Facebook has struggled to enforce its own policy on white nationalism and white separatism that it put in place on March 27, 2019. In addition, *The Guardian* reported that an external audit that Facebook commissioned revealed the new policy has been enforced in a narrow fashion: It's "been undercut by the company's decision to ignore content that does not use the term 'white nationalism.'" [22]

Facebook does not link to or even mention its policy on white-supremacist content in its Community Standards. That policy is referenced only on the company's Newsroom page and isn't visible or obvious to most users.

Aug. 2019

On Aug. 3, a man intent on stopping the "invasion" of Latinx immigrants into Texas murdered 22 people at an El Paso Walmart.

Two days later, several news outlets reported that Trump's 2020 reelection campaign posted more than 2,000 Facebook ads in January and February 2019 claiming that there was an "invasion" at the southern border. There's an undeniable connection between the hateful rhetoric in these ads and the language in the shooter's manifesto.

Facebook clearly states that "ads must not violate our Community Standards." In the aftermath of the El Paso massacre, Free Press and Define American launched a campaign urging Facebook to remove these ads from the platform (note: These ads are no longer running) and noted that they fall within the scope of prohibited content. Facebook claimed that these ads are "nuanced," insisting that they're "on the line, but [don't cross] it" because they don't explicitly use the term "immigrants" and therefore don't constitute the kind of "direct attack" the company requires to remove content under its hate-speech policy.

On Aug. 27, 2019, Facebook removed an Islamophobic event page that the advocacy group Muslim Advocates flagged. Though the outcome was laudable, it took extensive coordinated pressure from members of the Change the Terms coalition over the course of more than 28 hours for Facebook to enforce its own policies restricting threat of armed protest on events pages.

RECOMMENDATIONS

Facebook's enforcement continues to be haphazard. For example, we've found that the company's trusted-flagger program hasn't produced results. In fact, organizations that were originally part of the program have sidestepped it and instead contacted company representatives directly when finding problematic hateful activity on the platform.

This presents a massive inequity between well-connected national organizatons and local groups in the United States. The inequity between organizations and groups in the Global South and those in the West is even more stark as many non-U.S. groups have pointed out their inability to have Facebook take their flags seriously in situations presaging serious threats and crimes.

Facebook must dedicate real resources to the trusted-flagger program and ensure that front-line groups can raise serious issues with the company without having to rely on individual personal relationships.

Model Policy #3: Right of appeal

Determining hateful activities can be complicated. That's why a user should have the right to appeal any material impairment, suspension or termination of service, whether that impairment, suspension or termination represents a permanent ban or a temporary one.

This right should allow a user to make an appeal to a neutral decision-maker — someone other than the person who made the initial determination. That decision-maker should have knowledge of the context and social, political and cultural history of the user's country or countries.[23] The user filing the appeal should be permitted to present information to advocate for their position.

HOW IS FACEBOOK DOING?

Facebook began implementing an appeals process for its users prior to the Change the Terms launch, with a commitment to improving the process over time. Facebook's hate-speech take-down policy has several exceptions, and the company should give users sufficient nuance when removing their content. Providing an adequate level of detail is necessary to facilitating the appeals process.

Facebook has also committed to creating a global independent Oversight Board of approximately 40 experts to decide "tough content decisions." On April 1, 2019, Facebook previewed this idea and asked the public for feedback.

Free Press filed comments stating its skepticism about the purpose and overall effectiveness of the Oversight Board, recommending that Facebook go back to the drawing board.[24] In its comments, Free Press urged Facebook to instead focus on tightening its internal content-moderation policies and using Change the Terms' recommendations to provide an "intersectional racial-justice lens to content moderation to ensure that those most marginalized in our societies have their speech protected, and that those engaged in hateful activities come under greater scrutiny."

Sept. 2019

On Sept. 17, 2019, Facebook published its Oversight Board Charter.[25] Facebook has said that users and the company will both be able to ask for Oversight Board review, though the board will decide which cases to adjudicate and will "consider cases that have the greatest potential to guide future decisions and policies." The board will focus exclusively on content decisions, and its decisions will be binding on Facebook. Decisions will also have "precedential value."

Under Article 4 of the charter, Facebook has preserved some of its discretion in content decisions: "In instances where Facebook identifies that identical content with parallel context – which the board has already decided upon – remains on Facebook, it will take action by analyzing whether it is technically and operationally feasible to apply the board's decision to that content as well."

Furthermore, the charter allows the Oversight Board to make policy recommendations, but those will simply be "advisory opinions" and not binding on Facebook.

RECOMMENDATIONS

Facebook has invested a lot of time and resources into engaging with groups and individuals across the globe to establish the plan for this Oversight Board. As structured, it's a step beyond the current appeals process.

However, the process for appealing ordinary content decisions remains unclear, as does the process for appealing major questions to the board. Facebook has yet to publish the bylaws detailing the Oversight Board's operation. We have yet to see Facebook's plans for "ordinary" appeals, and we remain skeptical that the Oversight Board is the correct mechanism to answer major content-moderation policy decisions. The board's success will also depend largely on who is chosen to fill the 40 seats for the initial three-year term.

We urge Facebook to clearly explain how appeals of ordinary content decisions will take place and to clarify how the board will receive and decide appeals for "major" content-moderation cases.

Model Policy #4: Transparency

Change the Terms' recommendations for increased transparency have several specific asks. We suggest additional data points to evaluate what hateful activities are occurring on the platform and how Facebook is addressing those activities.

For example, Facebook should provide regular updates about the number of hateful activities the company identifies. These updates should be broken down by protected characteristics, the types of victim targets, how and by whom the content was initially flagged, how many people have been denied services for hateful activities, and the success rate of appeals.[26]

Facebook should publish this information in a format that protects users' personally identifiable information, and make this content available in formats that both people and machines can read.

HOW IS FACEBOOK DOING?

On May 23, 2019, Facebook published a transparency-and-enforcement report[27] on how much content it had removed between the fourth quarter of 2018 and the first quarter of 2019 for violating its policies against adult nudity and sexual activity, fake accounts, hate speech, spam, terrorist propaganda, and violent and graphic content.

The report provides a broad overview of how Facebook addresses hate speech on its platform but fails to provide the kind of granular data Change the Terms has called for. The report falls short because it simply states broad percentages and trends. For example, "in six of the policy areas [Facebook includes] in this report, we proactively detected over 95% of the content we took action on before needing someone to report it. For hate speech, we now detect 65% of the content we remove, up from 24% just over a year ago when we first shared our efforts. In the first quarter of 2019, we took down 4 million hate speech posts and we continue to invest in technology to expand our abilities to detect this content across different languages and regions."

This report provides some information regarding the rate of appeals, stating that more than 1 million people appealed the decisions Facebook made and fewer than 25,000 of those decisions were overturned either through appeals or other means. These percentages give researchers and civil-society groups a broad overview of the work Facebook has done to enforce its Community Standards. But we need much more specific data to fully understand the scope of the hate-speech problem on the platform.

Facebook has also added a "Recent Updates" section to its Community Standards that enumerates the changes the company has made since May 2018, a proactive step that offers a better understanding of how often Facebook updates its policies.[28]

RECOMMENDATIONS

To align with Change the Terms' recommendations, Facebook must capture and make public specific details on its hate-speech content-moderation practices. Providing the level of granularity Change the Terms calls for would help the company and civil-society groups better identify the strengths and weaknesses in Facebook's content-moderation practices.

Model Policy #5: Evaluation and training

Change the Terms recommends that online platforms "establish a team of experts on hateful activities with the requisite authority to train and support programmers and assessors working to enforce anti-hateful activities programs of the terms of service, develop training materials and programs, as well as create a means of tracking the effectiveness of any actions taken to respond to hateful activities."

Change the Terms also urges each platform to assign a member of its executive team to serve as a senior manager focused on overseeing how the company addresses hateful activities. The senior manager would need to "approve all training materials, programs, and assessments."

Change the Terms recommends that platforms "routinely test any technology used to identify hateful activities to ensure that such technology is not biased against individuals or groups … make the training materials available to the public for review; locate assessment teams enforcing the hateful activities rules within affected communities to increase understanding of cultural, social, and political history and context."

HOW IS FACEBOOK DOING?

Facebook's second Civil-Rights Audit Report previewed several changes taking place at the company regarding evaluation and training for content moderators:

1. Hate-speech reviewer specialization
"Specialization will allow reviewers to build expertise in one specific area, with the ultimate goal of increasing accuracy and consistency in enforcement. This is especially important in the case of hate speech because expertise and context are critical to making decisions."

2. Removal of posts exposing hate speech
Context matters and Facebook's review tool doesn't always display captions "immediately adjacent to the post — making it more likely that important context [will be] overlooked." Facebook committed to:

- updating its review tool to add further context
- inserting additional prompts in its review tool to ask reviewers whether the speech condemns and doesn't condone hate speech
- updating training materials

3. Changing review tool
Right now reviewers must first make decisions about pieces of content and are only then asked questions about why the content does or doesn't violate Facebook's rules. The company initiated a pilot program that flips these steps to determine whether that shift improves accuracy.

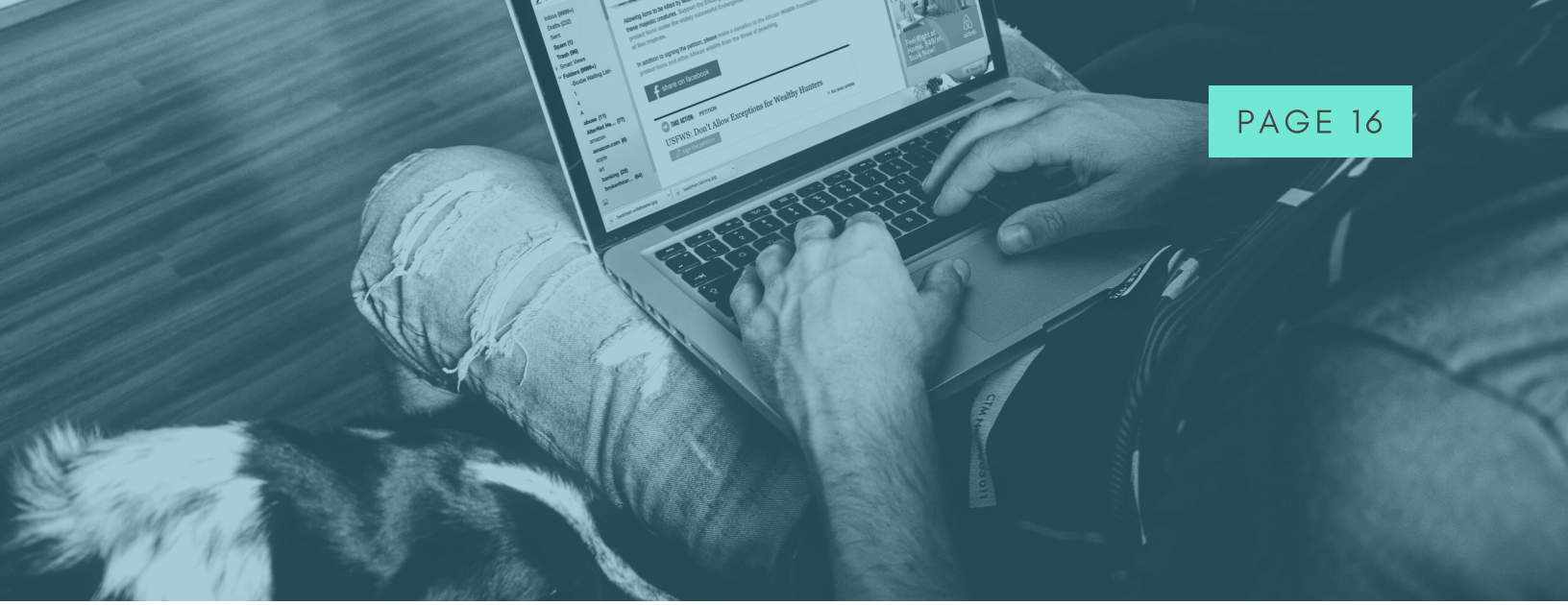
4. Bulk reporting or blocking
The audit team "recommends that Facebook develop mechanisms for bulk reporting of content and/or functionality that would enable a targeted user to block or report harassers en masse, rather than requiring individual reporting of each piece of content."

5. Protections for activists and journalists
The audit team "recommends that Facebook commit to working with journalists and activists who are routinely targeted on the platform to better understand the attacks and/or harassment — and to identify and test different solutions to improve Facebook's hate speech and harassment protections."

RECOMMENDATIONS

Facebook has yet to provide training materials for review by civil-society groups. Depending on how the global Oversight Board is implemented, its purpose and function could be a step closer to Change the Terms' recommendation that the company "locate assessment teams enforcing the hateful activities rules within affected communities to increase understanding of cultural, social, and political history and context."

For example, the board will be permitted to consult with outside experts when needed. It's unclear how the board's decisions will impact local content moderators or whether Facebook will commit additional local resources for content moderation. Facebook must deepen its ties to local communities and experts that can provide the company with needed perspective and contextual knowledge.



Model Policy #6: Governance and authority

Change the Terms recommends that a company "integrate addressing hateful activities into [their] corporate structures in three ways":

1. Assign a committee comprised of members from a platform's corporate board to assess management efforts to stop hateful activities on their services.

2. Assign an executive-team member to serve as a senior manager to oversee addressing hateful activities. Name that person publicly and ensure they have adequate resources and authority.

3. Create a committee of outside advisers with expertise in identifying and tracking hateful activities who will produce an annual report on the effectiveness of the steps the company has taken.

HOW IS FACEBOOK DOING?

The second Civil Rights Audit Report states that Facebook had created and institutionalized a "Civil Rights Task Force" within the company. COO Sheryl Sandberg will lead the task force, which will be composed of senior leaders from across the company. The task force will hire civil-rights experts to ensure effectiveness of its work, and Facebook will introduce civil-rights training for both senior leaders of the task force and key employees who work in product development.

The task force's proposed structure addresses Change the Terms' second recommendation under governance and authority. This task force is in its beginning stages.

Facebook hasn't assigned a board committee to assess management's efforts to stop hateful activities. The company relies on the individuals conducting the civil-rights audit to report on efforts to stop hateful activities. There will be a third and final report for the civil-rights audit released sometime in early 2020, and there's no sign that there will be a permanent effort to hold management accountable.

In addition, Facebook hasn't created a committee of outside advisers to identify and track hateful activities or create an annual report on such work. In light of the resources Facebook has already invested in the Oversight Board, we would like to see a similar effort to make this Change the Terms recommendation part of Facebook's future plans.

RECOMMENDATIONS

We remain concerned about Facebook's commitment to this process, including its willingness to implement the full recommendations of its own civil-rights audit. We recommend mirroring the process Facebook has taken in complying with the Federal Trade Commission's privacy settlement and compliance system.[29] That is, the company should appoint a committee at the board level to raise questions and monitor Facebook's progress on addressing civil rights, anti-discrimination, hate speech and disinformation issues.

Model Policy #7: State actors, bots and troll campaigns

Change the Terms recommends that platforms ban the use of bots or teams of individuals for coordinated campaigns that engage in hateful activities.

HOW IS FACEBOOK DOING?

Facebook's Community Standards don't explicitly align with the Change the Terms recommendations. Under the "Violence and Criminal Behavior" section is a policy that bars "people from facilitating or coordinating future activity, criminal or otherwise, that is intended or likely to cause harm to people, businesses, or animals."

RECOMMENDATIONS

There's an urgent need to combat coordinated disinformation and hateful campaigns in the run-up to the 2020 U.S. elections. Facebook should release its plans to combat state actors, bots and troll campaigns and commit resources to ensuring the platform is an asset to democracy — not a detriment as it was in 2016.

CONCLUSION

Over the past year, Facebook has made changes to its internal policies and governance that have the potential to better address hateful activities on its platform. Change the Terms will continue to monitor how consistent and effective these changes are.

Facebook still struggles to enforce its policies, and it has intentionally built flexibility into its Community Standards on hate speech: This flexibility allows hateful activities to thrive when they don't technically "cross the line."

Facebook must also improve its transparency and invest resources in capturing and sharing the granular data Change the Terms recommends.

Facebook has committed extensive resources to improving its appeals process. It's providing users with additional information regarding content it removes, and it has solicited comments and held meetings to establish an independent global Oversight Board to address those content decisions that fall right on the line. But we're skeptical about how effective the board will actually be given its current structure and the absence of details on many issues — including the identities of the 40 board members.

Change the Terms welcomes the progress Facebook has made — but the company has a long way to go to combat hateful activities on its platform and keep our communities safe.

ABOUT THE AUTHOR

Carmen works to protect the open internet, prevent media and telecom-industry concentration, promote affordable internet access and foster media diversity. She also coordinates responses to regulatory proposals that threaten to widen the digital divide and has authored several notable pieces on the importance of Net Neutrality and the Lifeline program for communities of color. Before joining Free Press, Carmen led the policy team at the National Hispanic Media Coalition, where she drafted dozens of filings and Op-Eds, participated in speaking engagements across the country and testified before federal agencies on behalf of the Latinx community. She was the architect of a series of prominent federal-records requests that compelled the FCC to release more than 50,000 previously undisclosed consumer complaints about Net Neutrality violations. Earlier in her career, she worked at the Department of Justice. Carmen, a native of Puerto Rico, earned her J.D. from Villanova University School of Law and her B.A. from New York University.

CARMEN SCURATO

ENDNOTES

1. Kanter, Jake, "Here's What Sheryl Sandberg's Been Telling the Global Elite About the Radical Change Coming Facebook's Way," Business Insider, Jan. 25, 2019, https://www.ctpost.com/technology/businessinsider/article/Facebook-s-Sheryl-Sandberg-reveals-the-two-quotes-4707829.php.
2. Change the Terms is a coalition of more than 50 racial-justice and civil-rights groups. The core contributors are the Center for American Progress, Color Of Change, Free Press, the Lawyers' Committee for Civil Rights Under Law, the National Hispanic Media Coalition and the Southern Poverty Law Center, https://www.changetheterms.org/coalition (last visited on Sept. 23, 2019).
3. *See* generally "Change the Terms Recommended Internet Company Corporate Policies and Terms of Service to Reduce Hateful Activities," http://bit.ly/2lmfeUO ("Change the Terms Model Policies").
4. Facebook initiated a civil-rights audit in May 2018 in response to months of public pressure from civil-rights groups. *See* Fischer, Sara, "Exclusive: Facebook Commits to Civil Rights Audit, Political Bias Review," Axios, May 2, 2018, https://www.axios.com/scoop-facebook-committing-to-internal-pobias-audit-1525187977-160aaa3a-3d10-4b28-a4bb-b81947bd03e4.html. Later that year, Facebook released its first report on the progress of its audit. *See* "Update on Facebook's Civil Rights Audit," Dec. 18, 2018, https://fbnewsroomus.files.wordpress.com/2018/12/Civil-Rights-Audit-Update.pdf. This first report focused on election-protection issues, advertising practices and transparency, among other issues. Facebook released its second report in mid-2019; it focused on issues such as content moderation, enforcement and creating a civil-rights accountability structure. *See* "Facebook's Civil Rights Audit — Progress Report," June 30, 2019, https://fbnewsroomus.files.wordpress.com/2019/06/civilrightaudit_final.pdf.
5. Change the Terms Model Policies, p. 2, http://bit.ly/2lmfeUO
6. Facebook Community Standards, https://www.facebook.com/communitystandards/hate_speech (last visited on Sept. 23, 2019). Facebook states that immigration status has "some protections" under its hate-speech policy because it allows for "criticism of immigration policies and arguments for restricting those policies." Id.
7. Change the Terms Model Policies, p. 3.
8. Change the Terms Model Policies, p. 4 (emphasis added).
9. Facebook Community Standards: Additional Information breaks down the severity of each tier of attack under its hate-speech policy. "Tier 1, the most severe, involves calls to violence or dehumanizing speech against other people based on their race, ethnicity, nationality, gender or other protected characteristic ("Kill the Christians"). Tier 2 attacks consist of statements of inferiority or expressions of contempt or disgust ("Mexicans are lazy"). And Tier 3 covers calls to exclude or segregate ("No women allowed")." https://www.facebook.com/communitystandards/additional_information (last visited on Sept. 23, 2019).
10. "Facebook's Civil Rights Audit — Progress Report," p. 14, https://fbnewsroomus.files.wordpress.com/2019/06/civilrightaudit_final.pdf (June 30, 2019).
11. *See* "Facebook Community Standards, Violence and Incitement," https://m.facebook.com/communitystandards/credible_violence (last visited on Sept. 23, 2019).
12. *See* "Standing Against Hate," Facebook Newsroom, https://newsroom.fb.com/news/2019/03/standing-against-hate/
13. Dickey, Megan Rose, "Facebook Civil Rights Audit Says White Supremacy Policy Is 'Too Narrow,'" TechCrunch, June 30, 2019, https://techcrunch.com/2019/06/30/facebook-civil-rights-audit-says-white-supremacy-policy-is-too-narrow/
14. Facebook Terms of Service, https://www.facebook.com/terms.php (last visited on Sept. 23, 2019).
15. "Facebook's Civil Rights Audit — Progress Report," June 30, 2019, pp. 12-13, https://fbnewsroomus.files.wordpress.com/2019/06/civilrightaudit_final.pdf.
16. *See* Change the Terms, Enforcement, p. 4 (full list of enforcement recommendations).
17. Facebook Community Standards, Dangerous Individuals and Organizations, https://www.facebook.com/communitystandards/dangerous_individuals_organizations (last visited on Sept. 23, 2019).
18. Frej, Willa, "Facebook Bans Content Links to Proud Boys, Gavin McInnes," HuffPost, Oct. 31, 2018, https://www.huffpost.com/entry/facebook-bans-proud-boys_n_5bd97f2fe4b01abe6a19964c?guccounter=1
19. Wagner, Kurt, "Facebook Is Taking Down 22 More Pages Tied to Infowars Founder Alex Jones," Recode, Feb. 5, 2019, https://www.vox.com/2019/2/5/18212439/facebook-alex-jones-remove-pages-infowars-again
20. Lorenz, Taylor, "Instagram and Facebook Ban Far-Right Extremists," The Atlantic, May 2, 2019, https://www.theatlantic.com/technology/archive/2019/05/instagram-and-facebook-ban-far-right-extremists/588607/.
21. *See* Facebook Community Standards, Dangerous Individuals and Organizations, https://www.facebook.com/communitystandards/dangerous_individuals_organizations (last visited on Sept. 23, 2019) stating that Facebook does not "allow any organizations or individuals that proclaim a violent mission or are engaged in violence [to have] a presence on Facebook." By contrast, Twitter states under its "Terrorism and Violent Extremism Policy" that it "examine[s] a group's activities both on and off Twitter to determine whether they engage in and/or promote violence against civilians to advance a political, religious and/or social cause." *See* Twitter, General Guidelines and Policies, Terrorism and Violent Extremism Policy, https://help.twitter.com/en/rules-and-policies/violent-groups (last visited Sept. 23, 2019).
22. Hern, Alex, "Facebook Ban on White Nationalism Too Narrow, Say Auditors," The Guardian, July 1, 2019, https://www.theguardian.com/technology/2019/jul/01/facebook-ban-on-white-nationalism-too-narrow-say-auditors.
23. *See* Change the Terms, Right of Appeal, p. 6.
24. *See* Carmen Scurato and Gaurav Laroia, "Free Press Comments on Facebook Oversight Board," Free Press, May 13, 2019, https://www.freepress.net/sites/default/files/2019-05/Facebook_Oversight_Board_Comment_0.pdf.

ENDNOTES

25. *See* "Oversight Board Charter," Facebook, Sept. 17, 2019, https://fbnewsroomus.files.wordpress.com/2019/09/oversight_board_charter.pdf.
26. *See* Change the Terms, Transparency, pp. 6-7 (recommendation that platforms collect detailed information and make that data easily accessible to the general public)
27. *See* Facebook Transparency, "Community Standards Enforcement Report," May 23, 2019, https://transparency.facebook.com/community-standards-enforcement.
28. *See* Facebook, "Recent Updates," https://www.facebook.com/communitystandards/recentupdates/ (last visited on Sept. 23, 2019).
29. *See* Federal Trade Commission, "FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook," July 24, 2019, https://www.ftc.gov/news-events/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions "The order creates greater accountability at the board of directors level. It establishes an independent privacy committee of Facebook's board of directors, removing unfettered control by Facebook's CEO Mark Zuckerberg over decisions affecting user privacy."