

# **Facebook's Civil Rights Audit – Progress Report**

---

June 30, 2019

## Contents

---

<b>I. A Message from Laura Murphy</b>	<b>4</b>
<b>II. Content Moderation &amp; Enforcement</b>	<b>7</b>
<b>III. Advertising Targeting Practices</b>	<b>15</b>
<b>IV. Elections &amp; 2020 Census</b>	<b>18</b>
<b>V. Civil Rights Accountability Structure</b>	<b>26</b>

## About the Civil Rights Audit

---

Facebook's civil rights audit began in 2018 at the encouragement of the civil rights community and as part of the company's commitment to advance civil rights on the platform. The audit is intended to create a forum for dialogue between the civil rights community and Facebook with the goal of informing the company's examination of critical areas of concern. Through this engagement, recommended changes and improvements will be identified, prioritized, and implemented. This process aims to focus the company's attention on its civil rights footprint in a sustained and comprehensive way.

The civil rights audit is led by Laura Murphy, a widely respected civil rights and civil liberties leader. The work of the audit—the analyses, reviews, and recommendations—is carried out by the Audit Team, which includes Laura Murphy and a team from noted civil rights law firm Relman, Dane & Colfax led by Megan Cacace, one of the firm's partners.

During the first six months of the audit, Laura Murphy conducted interviews with over 90 civil rights organizations to understand their concerns and identify key issues to analyze. The focus areas for the audit, which were informed by those interviews, were described in [the first audit report](#), released in December 2018.

In an effort to ensure transparency in the audit's progress, this second audit report describes developments since December 2018. A third and final report will be issued in the first half of 2020. Each report describes the audit work during that period and identifies changes implemented through the audit, along with commitments to further improvements, studies, or steps to be taken.

## I. A Message from Laura Murphy

---

I was honored to accept Facebook's invitation to lead their civil rights audit in May 2018 because the work is so timely and significant to both the company and the civil rights community. This is the second civil rights audit update and it builds upon the first update that was issued on December 18, 2018.

This work has implications for Facebook's users—over 2 billion people—and the broader tech community and society. Facebook aims to be a platform to connect, build communities, and bring people together. In an increasingly diverse nation and world, that vision for the company requires vigilance against actions on the platform that could result in hate or discrimination.

Given the significance of civil rights issues to our society, the goal of this audit is to make sure important civil rights laws and principles are respected, embraced, and inform the work at Facebook. These are a multifaceted and complex set of challenges. Our nation's history includes over a century of long and painful debates that have produced laws to make our society more just and fair, yet discrimination based on race, religion, national origin, ethnicity, disability, sex, gender identity, sexual orientation, age, and other categories still exists and continues to plague U.S. society. This persistent discrimination in our society—coupled with today's reality that there is an unprecedented volume of digital content and speech on social media—is what makes Facebook's commitment to the civil rights audit so critical.

Without an active commitment by tech companies to understand these impacts, patterns of discrimination and bias can spread online and become more entrenched. When you consider all of the groups that are subject to discrimination, it is clear that civil rights issues don't concern one set of people or one community but they affect everyone and speak to our shared values. As such, being responsive to civil rights concerns is not just a moral imperative but a business imperative for all online platforms.

In my view, the work outlined in this report represents meaningful progress on several issues and a concrete commitment to further improvement. This report covers substantive changes to Facebook's policies and operations—and makes note of the building blocks that the company has put in place to establish civil rights engagement that involves oversight from the highest levels of the company.

Instead of dealing with issues like election interference, hate speech, civil rights lawsuits, and algorithmic bias in silos, Facebook understands that it needs a strong, systematic, cross-functional framework to address these issues over time. This report announces important steps Facebook is putting into place to address that need, while also recognizing that more work is ahead.

I believe that the civil rights audit has helped effect meaningful change at Facebook and will affect millions of people's lives for the better. As a result of this audit, which has benefited from invaluable input from civil rights advocates, Facebook committed to the following over the last six months:

- **Enforcing against hate and other harmful content—and strengthening policies.** In March 2019, Facebook built upon its policies against white supremacy and hate to explicitly ban praise, support and representation

of white nationalism and white separatism. The company also updated its violence and incitement policies to prohibit posts from people who intend to bring weapons anywhere to intimidate or harass others, or who encourage people to do the same. Facebook has recognized that new policies must be complemented by stronger and wiser enforcement. In this report, I walk through an investigation into the company's enforcement mistakes, which has prompted a series of pilots the company is running to get better at enforcement—particularly with content that is mistakenly removed when it is condemning, not promoting, hate speech.

- **Fighting discrimination in Facebook ads.** As a result of the historic civil rights settlement in March 2019, Facebook is implementing a new advertising system so advertisers running U.S. housing, employment, and credit ads will no longer be allowed to target by age, gender, or zip code—and they will have a much smaller set of targeting categories overall. These changes go a long way toward protecting social and economic opportunities for all users. The report outlines the specific changes the company has agreed to make—and announces that Facebook is going above and beyond the settlement terms to create searchable databases for credit and employment in addition to housing ads nationwide.
- **Protecting against misinformation and suppression with the upcoming U.S. election and 2020 Census.** Facebook is treating the Census as a top priority. In the report, I describe the policies and protections the company is putting in place to help ensure that the constitutionally mandated census count isn't tainted by malicious actors spreading false information or engaging in campaigns of intimidation designed to discourage participation. I also outline the investments and changes the company is making to address the learnings from 2016 leading up to the 2020 election cycle—including developing a new policy that prohibits ads telling people not to vote (a tactic that was used in 2016).

In addition, this report introduces the steps Facebook is taking to build greater civil rights awareness and accountability across the company on a long-term basis. In practice, this means:

- Facebook has made a long-term commitment to institutionalize and continue the Civil Rights Task Force that is composed of senior leaders across the company.
- Facebook's COO Sheryl Sandberg is leading this work and overseeing the Task Force on a long-term basis.
- The Civil Rights Task Force will onboard civil rights expertise to ensure the effectiveness of its work.
- The company will introduce civil rights training for all senior leaders on the Task Force and key groups of employees who work in the early stages of product and policy development.

These are important first steps to building structure that will be a foundation for long-term success.

In the next phase of the audit, I want to continue to advise the company on improvements with content policy and enforcement, and with elections. Diversity and inclusion is another area I want to better understand, and privacy and algorithmic fairness are other issues of concern to the civil rights community. The civil rights auditors and Facebook will develop a focused roadmap for the next phase of the audit in the coming weeks.

Since I initiated my audit of Facebook, I've observed a greater willingness for Facebook to listen, adapt, and intervene when it comes to civil rights. I do believe the company is in a different place today than it was a year ago on these issues—I've seen employees asking the right questions, thinking about things through a civil rights lens, and poking holes in things on the front end of products and policies.

While this audit report shows Facebook is making progress, there is still much more work to do—and Facebook acknowledges this. As the company's work in this area continues, I want this effort to be even more robust and deeply embedded in the company's DNA.

I'm grateful to all the leaders and advocates who have contributed to the audit so far, and to my colleague Megan Cacace of Relman, Dane & Colfax. I look forward to continuing to work with Facebook leadership to effect meaningful and long-lasting change that long outlives the completion of the audit.

## II. Content Moderation & Enforcement

---

### Introduction

One of the primary focus areas of this update is content moderation—that is, the lines Facebook has drawn about what is and isn’t allowed on the platform. Facebook’s content moderation policies are generally reflected in its Community Standards, but the civil rights groups have questioned the adequacy and the consistent enforcement of those standards.

Within the civil rights community, both the advocates of free speech and the advocates for removal of more hate speech complain about the opaque nature of Facebook’s content moderation and enforcement practices. Specifically, the definition and policing of hate speech and harassment on the platform has long been an area of concern. The civil rights community also claims that a lack of civil rights expertise informing content decisions leads to vastly different outcomes for users from marginalized communities.

Given this backdrop, the Audit Team worked with Facebook to focus on the three categories of content moderation concerns that have been most frequently raised by civil rights groups:

- 1. Over-enforcement of hate speech policies that takes the form of removing content where users condemn or speak out against hate speech**
- 2. Under-enforcement of hate speech policies where hateful content is left on the platform**
- 3. Harassment of users on the platform**

While the audit’s work on content moderation remains ongoing, this section outlines the recent changes that have been made, provides an update on continuing efforts as well as recommendations for further improvements, and previews work that is planned for the next phase of the audit.

### Content Moderation Process Overview

To start, it is important to provide some context on how Facebook’s content review and enforcement system works based on the Audit Team’s review of the system.

Facebook defines hate speech as a direct attack on people based on “protected characteristics” — race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability. Facebook also provides some, but not all, of the same protections for immigration status. An “attack” for hate speech purposes is defined as violent or dehumanizing speech, statements of inferiority, expressions of contempt or disgust, or calls for exclusion or segregation. To unpack this definition, Facebook separates prohibited hate speech into three tiers based on the severity of the content.

Enforcement of Facebook’s content policies requires that hate speech appearing on the platform is **identified** or **flagged**, **reviewed** against Facebook’s policies, and then **removed** when it violates policy.

**Flagging Hate Speech for Review:** The Auditors have been told that Facebook identifies hate speech on its platform in two ways: (1) user reporting and (2) proactive detection using technology. Both are important. But advances in technology, including in artificial intelligence, mean that Facebook has gotten better at identifying hate speech before anyone reports it: Facebook reports that in December 2017, 24% of removed hate speech was detected proactively, without having to wait for a user to report it. According to Facebook, that number has more than doubled to 65% as of March 2019.

**The Review & Removal Process:** It is important to note that we have been told that content identified by technology as potential hate speech is not immediately removed. Instead, it is flagged and routed to human content reviewers who can understand context and make a decision accordingly. Further, to ensure that reviewers have appropriate knowledge of local terminology and nuance, over 90% of U.S.-based hate speech is reviewed by teams based in the U.S.

Because of the scale at which Facebook operates and the billions of pieces of content posted to the platform each day, Facebook insists that it must write policies that are operable globally. Facebook also expresses a desire for consistency in review decisions, such that different reviewers applying Facebook policy to the same content should reach the same conclusion. In the Audit Team's view, the scale of Facebook's review operations coupled with the stated goal of achieving consistency in review decisions translates to a blunter set of policies that are less able to account for nuance.

At the same time, we've learned an important point: to promote consistency and mitigate against the potential for bias, reviewers also lack discretion to stray from the policy language. Reviewers lack the authority to override policy language for content that falls within gray areas (e.g., content that seems to fit within the intended spirit of the policy but is technically not within the scope of the policy language). Reviewers have the option to flag content for more senior internal review.

The Audit Team has observed that when the flagging, review, and removal process identifies gray areas or potential policy gaps, Facebook has a process for re-evaluating existing policy lines to assess whether they should be moved. Over the last year, with input from civil rights groups and others, Facebook has gone through that process on aspects of its hate speech policy resulting in both small and substantive changes to its policies and enforcement.

### Content Moderation Progress Update

Based on feedback from the civil rights community on areas where Facebook's content moderation practices fall short and are still lacking, the Audit Team has identified and advocated for a series of policy and enforcement improvements, outlined in Sections A and B.

In addition, the Audit Team has identified areas where additional changes and improvements were recommended, which are discussed in Section B. While these changes and recommendations are a step forward, the issue of content moderation was simply too vast to comprehensively examine in the six-month period since the December report. Therefore, the last section of this chapter, Section C, describes both ongoing efforts that will continue and upcoming areas of focus for the next phase.



## A. Policy & Enforcement Improvements

As a result of the Audit, Facebook made two substantive policy changes that aim to counter hate on the platform. We also mention here a related initiative that Facebook advanced on its own.

**White Nationalism Policy:** In March 2019, Facebook announced a new policy banning the explicit praise, support, or representation of white nationalism and white separatism on Facebook. This policy was the culmination of months of engagement with global experts, civil rights groups, and the Audit Team. While Facebook has long prohibited statements of white supremacy and white supremacist organizations as well as other hateful content that calls for the exclusion of people based on race, ethnicity or religion (among other protected characteristics), Facebook previously did not prohibit white nationalism or white separatism. Civil rights groups criticized Facebook's treatment of these ideologies as less harmful or otherwise distinct from white supremacy and highlighted the connection between white nationalism and white separatism and hate organizations, symbols, and figures, as well as specific acts of violence. Civil rights groups pointed to connections between these groups and specific acts of violence.

While (as detailed below) the policy can and should be improved, Facebook's adoption of the white nationalism policy is a noteworthy step forward and a result of the audit.

The Auditors believe that Facebook's current white nationalism policy is too narrow because it prohibits only explicit praise, support, or representation of the terms "white nationalism" or "white separatism." The narrow scope of the policy leaves up content that expressly espouses white nationalist ideology without using the term "white nationalist." As a result, content that would cause the same harm is permitted to remain on the platform.

The Audit Team recommends that Facebook expand the white nationalism policy to prohibit content which expressly praises, supports, or represents white nationalist ideology even if it does not explicitly use the terms "white nationalism" or "white separatism."

**Events Policy:** Facebook also recently adopted a policy designed to combat attempts to use the platform to organize events aimed at intimidating or harassing targeted minority groups or vulnerable people. More specifically, Facebook's new events policy, which is included as part of its policies against [violence and incitement](#), prohibits calls to action or statements of intent to bring weapons to an event or location when the stated intent is to intimidate or harass vulnerable individuals or members of a specific racial, religious, or other protected group. The civil rights community had expressed significant concern that the platform was being used to organize armed gatherings outside mosques, synagogues, or other houses of worship for the express purpose of intimidating targeted groups. Through ongoing dialogue and engagement with groups through the civil rights audit, these concerns were translated into a policy aimed at preventing this kind of organized, targeted intimidation. Facebook will continue to engage with groups to determine how effective this policy is at preventing use of the platform to organize harmful events. The Auditors recommend that Facebook continue to examine other ways that event pages may be abused to spread hate.

**Designated Hate Figures Enforcement:** In recent months, Facebook has continued to apply its "Dangerous Individuals and Organizations" policy, which bans organizations or individuals from the platform when they meet

certain hate or violence criteria. Under that policy, individuals or organizations that amplify or traffic in hate are banned from the platform, as was the case with Facebook's recent bans of Alex Jones, Milo Yiannopolous, Laura Loomer, Minister Louis Farrakhan and others. For individuals or organizations that go beyond amplifying hate to proclaiming a violent or hateful mission or engaging in acts of hate or violence, Facebook not only bans them from the platform but also prohibits other users from praising or supporting them.

In determining whether an individual or organization should be banned under the policy and to what extent, Facebook examines their online and offline activity, and evaluates signals such as:

- Whether the individual or organization had called or advocated for, supported, or carried out acts of violence based on race, ethnicity, religion, or similar protected characteristics;
- Whether they are a self-described or identified follower of a hateful ideology;
- Whether they use hate speech or slurs in the sections of their Facebook, Instagram or other social media profiles describing themselves or their organization (e.g., the "About" section);
- Whether they have had pages or groups removed from Facebook or accounts removed from Instagram for content that violates the platform's hate speech policies.

The Auditors are aware that civil rights groups have raised concerns about the consistency and comprehensiveness of Facebook's enforcement of its Dangerous Individuals and Organizations policy. They have found it both over-inclusive and under-inclusive. They have also identified examples where individuals or organizations are able to circumvent those bans and persist on the platform with apparent ease and visibility.

## **B. Content Review Improvements**

In addition to the aforementioned policy changes, Facebook has initiated a set of pilots and internal case studies to identify ways that Facebook's content review tools and practices could be improved to reduce hate speech under- and over-enforcement.

**Hate Speech Reviewer Specialization:** Facebook is exploring a move that allows some of its content reviewers to specialize in and focus on hate speech rather than having to review content across all of Facebook's content policies. 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. Facebook reports that initial results suggest this type of specialization improves accuracy.

Commitment: As the civil rights audit proceeds, the Auditors are encouraged that Facebook commits to continuing this pilot to assess how hate speech specialization affects the accuracy of content reviewer decisions. If the pilot continues to yield improved accuracy, Facebook will look to implement hate speech specialization responsibly with limits on the number of hours that a person should spend reviewing hate speech in a given work day. Facebook and the Auditors are very aware of the concerns that content reviewers have access to sufficient psychological support and wellness resources.

**Analysis of Errors:** One of the most frequent content moderation complaints raised in this audit process has been persistent enforcement errors. Of special concern are those errors that result in the removal of posts that condemn or expose hate speech (also referred to as “hate speech false positives”). Given the significance of those concerns, an investigation was conducted examining sample instances where this kind of content was incorrectly removed and later restored to the platform. The investigation revealed several ways Facebook could make changes to review tools and training to help guard against these errors.

Specifically, the investigation found that context matters. For example, sometimes users post photos that would otherwise violate Facebook’s hate speech policies, but accompany those photos with captions to indicate they are not embracing the hateful content but instead are calling attention to racism or hate. To accurately apply Facebook’s hate speech rules, content reviewers must look at the entire post, inclusive of comments or captions that accompany the content. Facebook’s investigation revealed that its content review system does not always place sufficient emphasis on captions and context. Specifically, the tool that content reviewers use to review posts sometimes does not display captions immediately adjacent to the post—making it more likely that important context is overlooked. More explicitly prompting reviewers to consider whether the user was condemning or discussing hate speech, rather than espousing it, may reduce errors.

Finally, the investigation suggested that providing training on how reviewers should consider context condemning or exposing hate speech could help prevent errors.

Commitment: Through its own investigation, as well as the audit, Facebook commits to:

1. Update its review tools to better highlight important context critical to determining whether a post is condemning hate speech and therefore should not be removed.
2. Insert into the review tool an additional prompt directing those assessing hate speech violations to consider whether the condemnation exception applies.
3. Update reviewer training materials to make explicit that hate speech that would otherwise violate policy does not violate (and should not be removed) where accompanied by context denouncing the hateful content.

**Review Tool Changes:** Facebook is considering a change to the order and manner in which hate speech content reviewers analyze content for action. Facebook’s current review process leads with a decision. First, content reviewers decide whether the content should be removed as a violation of Community Standards or be left up. Then, the reviewer follows prompts to answer questions about the basis for the decision.

To improve overall accuracy and help reduce the incorrect removal of posts condemning hate speech, Facebook is testing a new review procedure that would reverse this order by starting with a series of questions about the content that then lead up to a decision on whether or not to remove it. Facebook has reported positive feedback from content reviewers based on initial testing of the new, information-first guided review process, indicating that reviewers find the process easier and it lessens the burden of applying policy to content decisions.

**Commitment:** In this report, Facebook commits to completing the pilot and analyzing how this information-first review process affects U.S. hate speech enforcement. If the results show substantial improvement in accuracy, Facebook will change its content review tool to implement the information-first approach for hate speech. Facebook will also assess whether the approach would improve accuracy for other policies. If, on the other hand, accuracy does not improve through the information-first process, Facebook commits to exploring other mechanisms for increasing the accuracy of hate speech enforcement.

In addition to improvements mentioned above, the Audit Team also has suggestions to make progress on content policy and enforcement. The Audit Team conducted an analysis of Facebook's hate speech and harassment content policies and spent time observing content reviewers while they assessed content for violations to identify additional recommended policy and enforcement changes.

## Hate Speech

- **National Origin:** Facebook defines hate speech to include attacks against people based on their national origin. Facebook's current conception of national origin is focused on attacks based on one's country of origin. However, hate speech based on one's ancestry or location of origin often does not track country borders; instead it can include attacks based on larger regions or continents. Dehumanizing attacks based on regional origins (e.g., people from Central America or the Middle East) or continental ancestry should not be treated differently simply because the geographical location at issue extends beyond a single country.

The Audit Team recommends that Facebook revise the definition of national origin in Facebook's Community Standards to include continents and regions larger than a single country where used to attack people from that region or continent.

- **Humor:** Facebook's hate speech policy currently contains an exception for humor. However, what qualifies as humor is not standard, objective, or clearly defined; what one user (or content reviewer) may find humorous may be perceived as a hateful, personal attack by another. Nor does labeling something as a joke or accompanying it with an emoji make it any less offensive or harmful. Because of the inability to precisely define and consistently identify a subjective and amorphous category like humor, identifying humor as an exception to Facebook's hate speech policy likely contributes to enforcement inconsistencies and runs the risk of allowing the exception to swallow the rule.

The Audit Team recommends that Facebook remove humor as an exception to the hate speech policy (and ensure that any future humor-related carve-outs are limited and precisely and objectively defined).

## Harassment

- **Calls for Harassment:** The civil rights community has identified harassment as a significant problem on the platform. Facebook's current bullying and harassment policy does not expressly prohibit users from encouraging or orchestrating the harassment of others. Such a prohibition would help ensure that users do not use the platform as a tool for organizing and unleashing harassment campaigns on targeted users.

The Audit Team recommends that Facebook prohibit calls for harassment or efforts to organize harassment campaigns through the platform.

- **Bulk Reporting or Blocking:** One particularly concerning form of harassment identified by the civil rights community involves an individual user being flooded with multiple harassing messages or posts from a large number of users. Facebook's current reporting mechanisms do not allow for the bulk reporting of messages or blocking of user contact. Individualized reporting is time-consuming, requires the targeted user to accept and/or review each individual hateful message or post to report it (magnifying exposure to the content) and is often ineffective against the influx of messages.

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.

- **Protections for Activists and Journalists:** Activists and journalists are often particularly vulnerable to harassment campaigns. Many civil rights advocates have been vocal about their experiences being targeted by such campaigns. Facebook's current hate speech and harassment protections have not adequately addressed these kinds of attacks in a comprehensive or consistent way.

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.

### C. Ongoing Efforts

While the above changes, pilots, and recommendations represent the Audit's Team's progress on content moderation issues, much of the audit's (and Facebook's) work in this area remains ongoing and will be the subject of further reporting. Two issues in particular will be focus areas for the next phase of the audit: (1) harassment and (2) Facebook's content moderation appeals and penalty systems.

**Harassment:** Harassment will remain a significant focus area for the audit and the company. In the last six months, Facebook has conducted a number of its own studies on harassment and bullying on the platform to better understand the scope of the issue and the instances in which current enforcement measures have been ineffective. The Audit Team has conducted additional analysis and has flagged gaps in policy and enforcement.

As a result of this work and its illumination of the size and complexity of the harassment problem, Facebook has invested in additional product, engineering, and data science resources dedicated to studying and tackling the issue. This full-time team is tasked with building products to help protect people from bullying and harassment on the platform. The team's approach is guided by a focus on the perspective of the person experiencing the abuse and understanding how various enforcement approaches would feel to the targeted user.

Facebook's immediate priorities are to (1) complete the analysis of the nature and severity of harassment and bullying occurring on the platform, and (2) use those findings to begin building a suite of detection and enforcement tools designed to address the behavior. On a substantive level, addressing organized harassment, sexual harassment, and harassment of vulnerable groups like activists and journalists will be top priorities for the next phase.

In the next audit report, the Audit Team will report on developments and progress on these issues and specific priority areas.

**Appeals & Penalties:** The second focus area for the next phase is Facebook's content moderation appeals process and penalty system. In April 2018, Facebook gave users the option to appeal decisions where their photos, videos, and posts were removed for violating the platform's Community Standards. The civil rights community has criticized both the opaqueness of the appeals process, and its limited effectiveness at correcting reviewer errors. Facebook has informed the Auditors that it seeks to improve the transparency and consistency of its appeals decisions.

Facebook's penalty system, whereby users are temporarily or permanently barred from using the platform (sometimes referred to by users as "Facebook jail") as punishment for posting content found to be violating, is similarly the subject of criticism by the civil rights community for its lack of transparency and seemingly disproportionate and/or unjustified penalties.

The Audit Team will be focusing on both of these systems and will include an update on its work in the next audit report.

In addition to Facebook's work on the internal appeals process, the company will also create an Oversight Board so that Facebook users can appeal content decisions to an external body. The goal of the company is to establish a body with independent judgment whose decisions are transparent and binding to review some of the most challenging content decisions that Facebook makes. During the last six months, Facebook sought public input on how the board should function and be designed, a consultation process that involved civil rights leaders and organizations. Facebook plans to release a final charter that will govern the work of the board in August. The Auditors are encouraged that consultation with civil rights groups is ongoing.

### III. Advertising Targeting Practices

---

The way Facebook allows businesses to buy targeted advertising has been a focus of the civil rights audit since its inception. As Facebook Chief Operating Officer Sheryl Sandberg recognized in a [blog post](#), “housing, employment and credit ads are crucial to helping people buy new homes, start great careers, and gain access to credit. They should never be used to exclude or harm people. Getting this right is deeply important to me and all of us at Facebook because inclusivity is a core value for our company.”

Civil rights advocates have long maintained that Facebook's ad targeting features enabled unlawful discrimination by allowing housing, employment, and credit advertisers to include (or exclude) people from seeing ads based on gender, race, age, and other protected classes. Because advertising is increasingly moving to web-based systems like Facebook, civil rights groups are focusing on platforms as key forums to help eliminate bias, ensure fairness and create equal access to crucial economic and social opportunities.

In recent years, several groups sued Facebook, alleging violations of federal, state, and local civil rights laws including the Fair Housing Act, the Equal Credit Opportunity Act, and the Civil Rights Act among other statutes. As the audit progressed, Laura Murphy continued to convey to Facebook the vigorous objections of civil rights groups to advertising practices that were seen as inviting housing, employment, and credit advertisers to target ads in discriminatory ways. At the same time, Relman Dane & Colfax reviewed Facebook's ad tools to help the company understand what more it could do to prevent advertising discrimination.

The December 2018 Audit Update identified some changes that Facebook had made to its advertising system, including (1) removing thousands of targeting terms that could be perceived as describing users' race, color, national origin, sexual orientation, veteran or military status, or disability status; (2) requiring all advertisers to affirmatively commit to adhering to Facebook's non-discrimination policy; and (3) creating a library of active ads so people may see ads even if they weren't targeted to them. However, the December report was clear that more needed to be done.

Commitment: Since the December report, Facebook has committed to five major changes to its advertising targeting system designed to prevent discrimination. These changes were announced in March 2019 as part of the historic settlement of multiple discrimination lawsuits filed by the National Fair Housing Alliance, Communications Workers of America, the American Civil Liberties Union, and other private parties. Through these changes, Facebook hopes to set a new standard for non-discrimination in online advertising, and move toward a civil rights-forward advertising system.

First, Facebook has committed to creating a new process where advertisers offering housing, employment, or credit opportunities will be diverted to a system that only offers a limited set of targeting options. Recognizing the difference between targeting ads for cosmetics or concert tickets and targeting ads for jobs, housing, or credit opportunities, this new system will restrict ad targeting options to help guard against discrimination or misuse.

Specifically, the new system will not offer targeting by gender, age, or any other categories that appear to describe people of a certain race, religion, ethnicity, sexual orientation, disability status, or other protected class. For example, housing, employment, and credit advertisers will not be able to target ads based on interests like Spanish-

language TV channels (appears associated with ethnicity or race), disabled parking permits (appears associated with disability status), or whether a user identifies as a “mom” (appears associated with gender and family status). Multicultural affinity targeting categories, such as “Multicultural Affinity (African American),” will remain unavailable to advertisers diverted to the new housing, employment, and credit system. When the new system launches, it will contain a limited list of ad targeting options, and Facebook will not make any new terms available without first notifying the plaintiffs in the lawsuits and giving them an opportunity to object.

In addition, the new housing, employment, and credit ad system will place restrictions on location targeting. Because of residential segregation patterns, certain ZIP codes and neighborhoods are known to have higher concentrations of one racial, ethnic, or religious group or another. Accordingly, targeting ads by ZIP code or precisely defined geographic areas can result in ads being targeted along racial, ethnic, or religious lines. In the new advertising system, Facebook will no longer allow targeting by ZIP code. Additionally, location targeting by city, town, address, or other specific location will automatically cover a 15-mile radius around the targeted location. These changes will help ensure that any target area will be more inclusive and less precise in tracking specific ZIP code, neighborhood, or city lines (and therefore less likely to track residential segregation patterns).

To implement the new, restricted advertising system, Facebook will ask advertisers to indicate to Facebook before attempting to run an ad whether their ad involves housing, employment, or credit opportunities. Facebook will also use technology (i.e. algorithms and machine learning) as a backstop to identify and redirect housing, employment, or credit ads to the restricted advertising system whenever advertisers fail to self-identify before attempting to run their ad. Facebook will take steps to ensure that housing, employment, or credit advertisers that repeatedly fail to self-identify their ads are brought into compliance with the new system. To educate advertisers on their non-discrimination obligations, the new housing, employment, and credit advertising system will also provide resources for advertisers to better understand their obligations under federal, state, and local laws prohibiting discrimination.

The agreements settling the lawsuits required that all housing, employment, and credit ads created through Facebook's primary advertising program, Ads Manager, be redirected to the new system by September 30, 2019. Facebook is starting to roll out these updates ahead of schedule – and the full implementation in Ads Manager will be completed in September. Advertisers who attempt to create housing, employment, or credit ads through any other company system will be redirected to the new restricted advertising system by December 31, 2019.

**Second**, Facebook is building a tool that will allow users to search and view active housing ads by advertiser and by location targeted, regardless of whether they were in the targeted audience. This new, searchable housing ads tool will increase transparency and ensure that housing opportunities are visible to all users regardless of an advertiser's targeting criteria. Being able to see and take advantage of available housing opportunities advertised on Facebook will therefore no longer depend on whether a user is within the audience an advertiser decides to target. Facebook is actively developing this tool and commits that it will be available to users by the end of this year.

**Third**, Facebook has agreed to continue to require all housing, employment, and credit advertisers, as well as all advertisers over time, to certify that they will comply with Facebook's policies prohibiting discrimination and all applicable anti-discrimination laws. This certification requirement ensures advertisers are aware of their non-



discrimination obligations and highlights the significance of those obligations. Facebook began instituting this certification in 2018 and has committed to maintain it.

**Fourth**, key Facebook employees with responsibilities relating to advertising will undergo training on fair housing and fair lending laws. The training will be developed by the National Fair Housing Alliance in partnership with Facebook. Facebook has committed that its Ads Leadership Team and other teams who work on products that touch these issues, including policy, legal, and engineering teams, will participate in the training within the next year.

**Fifth**, Facebook has agreed to and has already begun engagements with academics, researchers, civil rights and privacy advocates, and civil society experts to study algorithmic modeling by social media platforms, and specifically the potential for unintended bias in algorithms and algorithmic systems. Algorithms are becoming ubiquitous on internet platforms and in social media. Facebook recognizes that the responsible deployment of algorithms requires an understanding of their potential for bias and a willingness to attempt to identify and prevent discriminatory outcomes. While more needs to be done, by engaging with groups to study unintended bias in algorithms on social media platforms, Facebook is taking a much needed step forward.

**Additional Changes:** These five changes represent a significant shift in how Facebook handles housing, employment, and credit ads. But Facebook's advertising changes extend beyond those agreed to as part of settlement. After rolling out the searchable housing ads tool, Facebook plans to develop similar tools for employment and credit ads too. Creating search tools for employment and credit ads will ensure that those opportunities are visible to interested users regardless of how advertisers target their ads. Given how critical access to housing, employment, and credit opportunities are, this could have a significant impact on people's lives.

**More to Come:** While this progress report includes multiple positive changes to Facebook's advertising system, it is not the end of the conversation. For example, civil rights groups have surfaced concerns about Facebook's ad delivery system. They have highlighted a [study](#) that argues that Facebook's ad delivery system may steer ads to certain groups of people even when the advertisers do not select any targeting options.

There is a pending Department of Housing and Urban Development Fair Housing Act complaint against Facebook focused on ad targeting and delivery that is of great concern to the civil rights community. The Auditors are eager to see the parties reach a meaningful resolution.

Facebook has asserted its commitment to continue to exploring ways to prevent advertising discrimination on its platform. This will include ongoing work to review, evaluate, and iterate on its advertising processes, including ad delivery, and seek feedback from the civil rights community.

## IV. Elections & 2020 Census

---

When the civil rights audit began in 2018, the upcoming 2018 midterm elections and specifically the prevention of voter suppression and intimidation, were a primary area of focus. The December 2018 Audit Update discussed new policies and operations changes that were introduced through the audit's work in advance of the midterm elections. But, the December 2018 report also acknowledged that more work needed to be done and the Audit Team has continued to focus on these issues. This chapter outlines the policy improvements, enforcement efforts, and resource commitments that have been made through this audit as we enter the 2020 election season. Facebook has taken steps to combat voter suppression, build defenses against suppressive tactics previously used on the platform, and onboard voting rights expertise. While the Audit Team acknowledges these steps forward, there is more to be done.

Additionally, recognizing the importance of the 2020 Census, the company has made this a priority issue and is building out policies, operations, and preparations for the census. This chapter will describe Facebook's 2020 Census-related efforts and commitments. Facebook is taking important steps to ensure that the census count mandated by the Constitution is not manipulated or interfered with through misrepresentation, misinformation, or suppressive tactics online.

### **A. Understanding & Acknowledging Voter Suppression and Intimidation Efforts**

Before detailing elections-related improvements, the Auditors believe that it is important to understand three points of context for Facebook's and the audit's work in this area.

First, voter suppression and intimidation have a long history in this country. The tactics used in such campaigns have evolved and will likely continue to do so. As a result, voter suppression and intimidation will require ongoing focus and attention.

Second, harmful manipulation and suppression tactics have been used on Facebook and other online platforms. In 2016, foreign actors used a network of fake Facebook accounts to sow discord and spread misinformation. The Senate Select Committee on Intelligence and the Mueller investigation found that the Russian Internet Research Agency's (IRA) efforts were specifically targeted at communities of color. Those findings are consistent with concerns expressed by the civil rights community: that foreign actors exploited the platform by using fake accounts to run ads that encouraged racial and religious animus, attempted to suppress minority voter turnout, and targeted communities of color and racial and ethnic minorities.

Third, Facebook acknowledges the harm to communities of color caused by these activities on the platform. Facebook says it was alarmed by the manipulative tactics deployed by the IRA and how its advertising platform was used to foster racial discord, and is committed to preventing such harms in the future.

In reaction to concerns raised through the civil rights audit, Facebook is making the following statement:

“Facebook was created to give people voice and provide a space to connect, engage, and build community. Voting is one of the most fundamental forms of constitutionally protected engagement and expression. Efforts to intimidate voters and suppress voting undermine those core values.

Yet in the United States, there has been a long and painful history of efforts to suppress the voting rights of certain citizens, particularly African-Americans, Latinos, Asian Americans, Native Americans, and language minorities that unfortunately persists even today. As such, the promise of the 15th Amendment to the Constitution and the 1965 Voting Rights Act has not been fully realized and voter suppression continues to affect our elections and our democracy. More recently, in the 2016 election foreign actors used a series of tactics, including networks of fake Facebook accounts, to spread misinformation, sow racial division and depress the vote, disproportionately targeting African Americans and other communities of color.

These are incredibly complex challenges, and the threats will continue to evolve even as platforms strengthen their defenses. We have unfortunately seen such efforts to suppress the right to vote surface on Facebook, and we are committed to combatting these tactics. We expressly ban attempts to suppress the vote, including by misleading people on when or how to vote and we remove this content from our platform. Learning from past efforts to misuse our platform for voter suppression, we're also now working on a policy to prohibit targeted ads encouraging people not to vote. In addition, we have set up dedicated reporting channels for state election authorities so that they can let us know if they see voting information on the platform that may be incorrect.

We will continue to invest resources and time in these efforts because it's critical to us that Facebook is a safe and productive place for our community. We were more prepared for these kinds of attacks in the 2018 U.S. midterms than we were in 2016, having expanded our policies against voter suppression and we will continue to improve ahead of the upcoming U.S. elections. By continuing to partner with civil rights organizations, federal and state lawmakers and agencies, voting rights experts, academics and researchers, we believe more than ever that Facebook can empower the voiceless, strengthen our democracies, and ultimately help create a more connected and civically engaged world.”

## **B. Elections Policy & Planning**

### Policy Improvements

Since 2016, Facebook has made policy improvements to help prevent voter suppression and intimidation, and continues to build on these efforts with an eye toward implementing additional policy developments in the near term. These developments are discussed in more detail below.

**Expanded Voter Suppression Policy:** Through the civil rights audit, Facebook strengthened its voter suppression policy in 2018, so that the policy now prohibits:

- Misrepresentations about how to vote (including statements that you can vote using an app);
- Misrepresentations about voting logistics, methods, or requirements;
- Misrepresentations about whether a vote will be counted; and
- Threats of violence relating to voting, voter registration, or the outcome of an election.

Facebook's work in this area continues. The civil rights community has made clear to Facebook that the company needs to focus on voter suppression as a distinct civil rights challenge and recognize how certain communities are often disproportionately targeted and harmed by suppression campaigns. Facebook has committed to doing so and will continue to take steps to prepare for future elections and future attempts at suppression. These actions include the following:

**Voter Interference Protections:** Facebook is currently working on expanding its voting-related policies to prohibit other forms of voter interference. While the specifics are still in development, the goal of the expanded policy is to capture other forms of voting-related interference not addressed by Facebook's existing policy. Facebook has met with external voting rights experts in the United States and abroad in connection with this policy expansion and will continue to do so as the policy development process moves forward. In the next few months, Facebook's policy team expects to make a recommendation regarding the specific contours of the new policy.

**New Policy Against "Don't Vote" Ads:** Facebook understands that in 2016 malicious actors used fake accounts to target specific communities with ads encouraging them not to vote. In 2018, Facebook adopted a number of changes in response to these tactics, including requiring verification of the identity and location of anyone purchasing a political ad and requiring political ads to be transparent in identifying who sponsored the ad. The purpose of these measures is to prevent foreign actors from running political ads and provide greater visibility into who is running a given ad.

Through the course of this civil rights audit, Facebook has acknowledged that those seeking to engage in voter suppression will attempt to circumvent these safeguards and identify alternative ways to communicate suppressive messages. For that reason, Facebook believes it is important to launch a policy that prohibits ads targeting audiences in the U.S. that expressly encourage people not to vote. This new "don't vote" policy prohibition is expected to go into effect before the 2019 gubernatorial elections.

**Addressing Racial Appeals:** One tactic that has long been a concern of the civil and voting rights community is the use of racial appeals. Racial appeals are explicit or implicit efforts to appeal to one's racial identity and/or use race to motivate someone to vote for or against a given candidate. While such appeals have historically been focused on race, they can also apply to religion, ethnicity, or other protected characteristics. Racial appeals can include explicit calls not to vote for a candidate because of the candidate's race; warnings that a candidate is not "one of us;" threats that a racial, religious, or other protected group will "take over" if a certain candidate is elected; or more subtle messaging. These kinds of appeals sow division, advance stereotypes, and promote voter suppression and the disempowerment of minority communities.

Racial appeals have been a significant area of focus for the Audit Team. While the issue has not been resolved, through the audit's work, Facebook has taken several steps that will enable the platform to identify and remove explicit racial appeals more effectively. First, as a starting point, Facebook has gained a better understanding of the impact of racial appeals and acknowledges that racial appeals have appeared on the platform.

Second, in 2018, Facebook adopted an inflammatory content policy, which prohibits attacks in ads based on protected characteristics, such as race, religion, or ethnicity, or immigration status. While this decision was not made specifically to address racial appeals, it does capture some manifestations of this content on the platform. For example, content that would be prohibited under this policy includes fear-mongering claims that all members of a racial group will “take over” or that a religious or immigrant group as a whole represents a terrorist or criminal threat. In addition, Facebook is working on refining this policy to remove more inflammatory and fear-mongering content in ads, while still allowing legitimate speech and debate. Facebook is in the process of developing improved policy guidelines and conducting internal testing to increase consistency of enforcement, with the goal of making the updated policy public before the November 2019 gubernatorial elections.

Third, after the 2016 elections, Facebook updated its hate speech policy to specifically prohibit statements of political exclusion. Prohibited statements of political exclusion include statements that certain people should not be able to run for (or hold) political office or have their vote counted because of their race, religion, gender, sexual orientation, or other protected characteristics. This prohibition also covers explicit racial appeals encouraging voters not to vote for a candidate because of the candidate's race or other protected characteristic.

These steps, while not all-encompassing, are important improvements. Given the substantial harms caused by racial appeals, the Audit Team recognizes that it is critical for Facebook to continue to devote resources to ensuring the effective, consistent, and proactive enforcement of these policies.

#### Election-Related Operations, Resources, and Planning

The civil rights community expressed concern in the lead-up to the 2018 elections that Facebook lacked sufficient infrastructure to handle the voting-related concerns and challenges that would arise, and that Facebook's operational focus shifted to U.S. elections too late in the election season. These concerns have been underscored by the Audit Team and have resulted in the following commitments:

**Combatting Information Operations by Sophisticated Bad Actors:** Facebook defines “Information Operations” as coordinated efforts to manipulate or corrupt public debate for a strategic goal. Both the Auditors and the civil rights community have stressed the significant threat posed by coordinated, deceptive information operations, such as those targeting communities of color in 2016. Facebook understands that the platform continues to be a target for deception through fake accounts and information operations campaigns to manipulate U.S. voters and elections. Facebook has said that it has established a globally-focused team to combat such campaigns by seeking to: (1) identify and dismantle sophisticated networks of bad actors by deploying a team of expert investigators who use tactics imported from the fields of cybersecurity research, digital forensics, law enforcement, national security, and investigative reporting; and (2) combat the most common threats by building technology to detect and automatically remove fake accounts, or most commonly, block them from being created in the first place; and (3) continue to improve the platform's defenses by building lessons learned from their investigations into how their products function to make it harder for bad actors to manipulate people.

While a detailed assessment of Facebook's tactics to detect deception was outside the scope of this audit, the Audit Team understands that Facebook has devoted resources to these issues. Facebook has also [reported improvements](#) in both its ability to block fake accounts and to detect and remove information operations networks.

**Proactive Detection of Voter Suppression:** Because of the time-sensitive nature of elections, civil rights groups have stressed that prompt enforcement of Facebook's voter suppression policy is critical to minimizing the suppressive impact of content that misrepresents voting requirements, logistics, or methods. Facebook reports that it has invested in proactive detection technology since 2016 and is deploying that technology to identify potentially suppressive content, without having to wait for reports from individuals or community partners.

While Facebook understands that it cannot expect suppression tactics to remain unchanged from one year to the next, the company recognizes the importance of learning from past examples. Accordingly, Facebook has used suppressive content identified in 2016 and 2018 to help refine its systems. Facebook also considers tactics and examples seen in elections in other parts of the world to be better prepared should those same approaches appear in the U.S. next. Facebook reports that it updates and retrains its proactive detection algorithms on an ongoing basis to ensure they are up to date.

**Full-Time Dedicated U.S. Elections Team:** Facebook has a full-time dedicated team focused exclusively on the U.S. 2020 elections, including product, engineering, threat intelligence, data science, policy, legal, operations, and other support. This team includes a dedicated U.S.-focused project manager, who is charged with leading the U.S. election work across the company and will ensure all these functions are integrated. Together this team will execute on Facebook's work to prevent interference in the 2020 U.S. presidential election cycle.

Facebook is also continuing to invest in rapid-response capabilities. In connection with the 2018 midterm elections, Facebook piloted a model of bringing key personnel together in a physical operations center (i.e., a war room) as an added layer of security to monitor for and quickly address spikes in voter suppression, hate speech, and other threats that increase in prevalence during an election. Facebook continued to use that model to support key elections around the world in 2019 and has indicated they will continue real-time monitoring and response for the U.S. 2020 elections. This includes an operations center that will be fully up and running for the 2020 election season by end of the year, and has already been deployed to help monitor the presidential debates this past week.

**Voting Rights Expertise:** Both the Audit Team and the civil rights community have emphasized that voting rights knowledge and experience is critical to ensuring that Facebook's polices, operations, and enforcement identify and properly address suppressive and intimidating content. Through these efforts, Facebook has committed to hiring a voting rights expert consultant in the next several months. This expert will be tasked with: (1) developing and delivering U.S. elections/voting rights training (described in more detail below); (2) advising Facebook's product and policy teams on potential gaps in its voter suppression policies and enforcement; (3) advising the company as it handles time-sensitive issues flagged for internal escalation, or grapples with borderline content that falls in the gray area around the edges of Facebook's policies; and (4) interfacing with external voting rights groups and surfacing key concerns internally. The expert consultant will have deep knowledge of voting rights law and both historical and recent trends in voter intimidation and suppression.

One of the key responsibilities of this voting rights expert will be to develop and conduct training for: (1) those responsible for voting-related content and advertising policies, (2) the leadership team responsible for U.S. election operations and (3) election operations center staff and any election-related escalation teams.

The Auditors will work with the voting rights expert to advise Facebook on the best way to implement this training at scale, so that it maximizes value for the content reviewers who enforce Facebook's policies against voter suppression and intimidation.

On a substantive level, the training will provide historical context on voter suppression and intimidation in the United States, including the history of targeting African Americans and minority communities, racial appeals, and the voting protections afforded by the Voting Rights Act. Training will also discuss recent trends in voter suppression and intimidation. In addition, the training will focus on enforcement of Facebook's voter suppression policies, explaining what voting-related content is prohibited on Facebook, and providing specific examples. Those in operations centers or supervising content reviewers will also be trained on Facebook's escalation protocols, ensuring there is a clear understanding of what to watch for and which issues should be escalated for further review or responsive action.

Anticipating that the voting rights consultant will be onboarded by the fall of 2019, Facebook expects the initial voting training to begin no later than early 2020, with an update—including a refresher on key concepts and updates to policies, protocols, or recent trends—conducted in late summer 2020.

**Searchable Political Ads Library:** To promote transparency and public accountability, Facebook has continued to expand its U.S. political/issue ads library, allowing all users to search ads relating to politics or issues of national importance dating back to May 2018. Users can search by keyword or page and then filter those ads by whether they are active or inactive. For each ad in the database, Facebook includes the content of the ad (including photos, videos, text, or graphics) and information about: (1) the ad's sponsor, (2) the range of money spent on the individual ad and both the total and recent amount spent by the page or advertiser, (3) the dates the ad ran, (4) the U.S. state(s) where the ad was shown, (5) the range of impressions received by the ad, and (6) a breakdown by age and gender of who was shown the ad.

Starting in May 2019, Facebook began publicly issuing daily ad reports that provide information about who is advertising on the platform, how many ads are in the library, how much advertisers have spent overall, how much has been spent on advertising in each state, and a searchable list of ads run and money spent by each advertiser.

In addition to these transparency tools, which are available to all users, Facebook has made data available through its Ad Library API. Once a user accesses the API, they can see Facebook's Ad Library data and perform additional, customized analysis in the library.

These measures represent a step in the right direction toward promoting transparency; however civil rights groups feel strongly that Facebook should go even further in disclosing additional data about these ads, such as targeting criteria used by advertisers, which are critical to the analysis of election manipulation and discrimination.

The Auditors encourage Facebook to also address technical limitations of its API, in order to make this research more broadly accessible. Facebook has reported that its product team is working directly with researchers to improve these limitations.

**Expanded Voting Reminders:** In an effort to support an informed and engaged electorate, Facebook plans to run voter registration reminders for U.S. primary and general elections, and voting day reminders for primary and general elections in 2020.

In addition, Facebook will continue to engage with secretaries of state, elections directors, and national organizing bodies such as the National Association of State Secretaries of State and the National Association of State Election Directors, enabling them to report content that may violate Facebook's Community Standards. For example, Facebook continues to maintain an external escalation channel for election directors to report content to Facebook that may violate state election laws.

**Continued External Partnerships & Engagement:** Both through the civil rights audit and outside of the audit context, Facebook has engaged with external groups and organizations on election-related issues. Facebook recognizes the critical value of this engagement and commits to continuing to this work on issues surrounding the election and the census.

In the second half of this year, Facebook will convene meetings with voting rights organizations and advocates to provide a forum for flagging new voter or census suppression concerns or tactics, and to discuss how Facebook's policy and operations improvements are working.

Facebook will also continue its engagement with candidates and officials. In 2019, Facebook has already led dozens of training sessions, briefings, and webinars for officials and candidates at the federal, state, and local level to advise them on how to use Facebook's tools most effectively and how best to protect their offices and campaigns against security breaches. These include trainings at conferences for women entering politics, for local advocacy groups and nonprofits, and for Black and Puerto Rican legislators, as well as Democratic and Republican staff associations and party committees. Facebook will host additional trainings later this year at upcoming conferences, including those representing Black and Hispanic state legislators.

### **C. Prioritizing the 2020 Census**

Apart from the presidential election, 2020 will be a critical year for an entirely different reason: it is a census year. The census is a constitutional requirement that serves as the foundation for allocating federal benefits and electoral representation. The significance of the census, coupled with the historical undercounting of certain groups, including Latinos, African Americans, Asian Americans, Native Americans, and other communities of color, immigrants, and people who are incarcerated, reinforces why the census is a critical focus area.

Through the audit, Facebook has committed to treating the 2020 Census with the same priority and focus as a high priority election. To that end, Facebook has agreed to take the following actions:



**Census Interference Policy:** Extending Facebook's voter suppression protections to the census, Facebook will launch a census interference policy in the fall of 2019. The policy will prohibit misrepresentations of census requirements, methods, or logistics, and census-related misinformation (threats of violence for participating in the census are already covered under Facebook's violence and incitement policy).

**Census Training:** In addition to the voting-related training described above, an expert consultant will also provide training to those responsible for census-related content or ad policies, census-related escalation teams, and anyone responsible for supervising content reviewers who will be enforcing Facebook's policy against census interference. The training will occur in early 2020, and will include explanation of the 2020 Census timeline and the significance of the census, current and/or historical examples of census suppression efforts, as well as discussion of what census-related content is prohibited on Facebook and any census enforcement and escalation protocols.

**Proactive Detection of Census Interference:** Facebook has committed to using proactive detection technology to identify content that violates Facebook's forthcoming census interference policy without waiting for users to report the content. Facebook is conducting research to inform the training of its algorithms for application to the 2020 Census. Facebook will also continue to work with the Audit Team to identify and engage census experts to assist the company in anticipating suppressive tactics and content and help support the company's proactive detection efforts.

**Census Interference Reporting Partnerships:** Paralleling the voting rights reporting partnerships described in the December 2018 Audit Update, Facebook will look to partner with census protection groups to flag for review by Facebook potentially suppressive census-related content they encounter.

**Dedicated Census Resources and Operations:** Facebook is creating a dedicated team focused on the 2020 Census, supervised by a single manager in the U.S. operations center, that will be responsible for developing and deploying Facebook's proactive detection tools. As Facebook continues its operations and resource deployment, Facebook commits to sharing more operational details and 2020 Census planning commitments.

**Promoting Census Participation:** Finally, Facebook will continue to engage with external stakeholders and partners to promote census participation. Facebook has already provided training to more than 20 different organizations on how to use its tools and products to increase awareness of the census and encourage participation. Facebook plans to conduct additional trainings and also establish internal protocols to promptly address census-related concerns escalated by external stakeholders.

## V. Civil Rights Accountability Structure

---

The civil rights audit was born out of longstanding calls from the civil rights community for Facebook to focus its attention on valid concerns the community had repeatedly raised, and to take concrete steps to address those concerns in meaningful, lasting ways. Civil rights groups understand that sporadic consideration of civil rights issues or temporary, ad hoc “fixes” to deeper, more systemic problems are not going to be meaningful in the long run. Even though the groups advocated for the audit, they also do not believe that three audit reports generated by outside consultants will be sufficient to resolve complex, long-term problems. In addition, in a company as large and decentralized as Facebook, civil rights concerns need to be addressed and resolved in a coordinated fashion.

That is why one of the chief issues the civil rights community has raised in the audit process has been Facebook’s civil rights accountability structure—that is, a commitment by Facebook that it has put the appropriate structure in place to ensure that civil rights issues are proactively recognized and considered, and that changes made to address civil rights concerns are implemented in a comprehensive and methodical way, both after problems are brought to Facebook’s attention and before new products and policies launch.

Accordingly, one of the Audit Team’s primary areas of focus since the December report has been Facebook’s civil rights accountability structure. As the largest social media company in the world, what Facebook has committed to here is a consequential and important start, but only if it continues to build upon what it is announcing today. Facebook products and policies impact millions of people, and for many, daily activities have moved online because of Facebook. It is essential that this company do everything in its power to respect and uphold civil rights laws and principles given the huge societal impact of this platform.

Facebook has committed to four actions to expand its commitment to civil rights accountability, while acknowledging that these are just the first steps toward building a long-term accountability structure.

**First**, Facebook has created and institutionalized a **Civil Rights Task Force**. The Task Force meets monthly for the purpose of surfacing, discussing, and addressing civil rights issues. Its membership includes senior leaders who are issue experts in areas such as content policy, fairness in artificial intelligence, and elections, along with C-suite executives and vice presidents from key departments, including:

- Product
- U.S. Policy/External Affairs
- Operations
- Advertising
- Marketing
- Diversity & Inclusion
- Human Resources
- Communications
- Partnerships
- Legal

The purpose of the Task Force is to:

1. Institutionalize the practice of regularly discussing important civil rights concerns or vulnerabilities on a systemic, cross-functional level;
2. Operate as an executive body where employees can escalate civil rights questions or issues to leadership and key decision-makers;
3. Ensure civil rights concerns raised by outside groups are escalated promptly to decision-makers so that they can be considered and acted on quickly, without having to go through layers of management;
4. Share Task Force learnings across and within departments and verticals at the company;
5. Ensure that a centralized team is empowered to influence contemplated product or policy developments that raise civil rights concerns. The Task Force will be in a position to work across company functions to review relevant products and policies with a civil rights lens and help surface vulnerabilities early on.

**Second, COO Sheryl Sandberg will chair the Civil Rights Tasks Force** and has committed to continuing to lead the company's civil rights work after the audit has concluded.

**Third**, Facebook will **onboard strong civil rights expertise** to support the Task Force's work. Beginning this summer, consultants with substantial civil rights expertise will work with both internal Facebook teams and the Task Force itself to ensure this expertise informs the company's work in this area. For example, a voting rights expert will work with Facebook's U.S. elections team starting this fall to provide voting rights expertise and training on both historical and recent trends in voter suppression and intimidation. Facebook and the Task Force will continue to bring in civil rights experts on a range of issues as needed to support the company's ongoing civil rights engagement.

**Fourth**, Facebook has committed to expert **civil rights training** for key groups of employees and managers, as well as all members of the Civil Rights Task Force. Employees working in the early stages of relevant product and policy development will participate in this ongoing training to better equip them to infuse civil rights considerations into their daily product/policy development tasks and decision-making. The goal of the training is to increase awareness of civil rights considerations, issues, and risks so that employees and Task Force members alike are better positioned to proactively identify civil rights concerns and seek solutions. To that end, training will be customized for Facebook to cover the most relevant issues for employees in a given department.

**The Beginning, Not the End:** Both the civil rights groups and the Audit Team agree that meaningful, long-term civil rights accountability structures are critical to ensuring that civil rights issues are identified, evaluated, and appropriately addressed before they have an impact on the community, and that civil rights vulnerabilities identified externally are promptly remedied. While these are noteworthy steps forward, the Audit Team has made clear to Facebook (and Facebook understands) that for the accountability structure to be successful in the long term, these are just the first steps of a longer path forward.