# Facebook's Civil Rights Audit – Final Report

July 8, 2020

## Table of Contents

**About** 3

**Acknowledgements** 4

**Introduction by Laura W. Murphy** 5

**Chapter One: Civil Rights Accountability Structure** 13

**Chapter Two: Elections & Census 2020** 20

**Chapter Three: Content Moderation & Enforcement** 42

**Chapter Four: Diversity & Inclusion** 59

**Chapter Five: Advertising Practices** 72

**Chapter Six: Algorithmic Bias** 76

**Chapter Seven: Privacy** 83

## About the Civil Rights Audit

This investigation into Facebook's policies and practices began in 2018 at the behest and encouragement of the civil rights community and some members of Congress, proceeded with Facebook's cooperation, and is intended to help the company identify, prioritize, and implement sustained and comprehensive improvements to the way it impacts civil rights.

The Audit was led by Laura W. Murphy, a civil rights and civil liberties leader, along with a team from civil rights law firm Relman Colfax, led by firm partner Megan Cacace.

During the first six months of the audit, Laura W. Murphy interviewed and gathered the concerns of over 100 civil rights organizations. Over the course of the Audit's two year engagement, that number exceeded 100 organizations, hundreds of advocates and several members of Congress. The focus areas for the audit, which were informed by those interviews, were described in the first preliminary audit report, released in December 2018. That was followed by a second update in July 2019, which identified areas of increasing concern for the Auditors. This third report will be the Auditors' final analysis.

The Civil Rights Audit is not an audit of Facebook's performance as compared to its tech industry peers. In some areas it may outperform peers with respect to civil rights, and in other areas, it may not. The Auditors are not privy to how other companies operate and therefore do not draw comparisons in this report. The scope of the work on the Audit was focused only on the US and the core Facebook app (rather than Instagram, WhatsApp, or other Facebook, Inc. products).

## Acknowledgements

The Auditors would like to thank the Civil Rights Audit support team from Facebook: Lara Cumberland, Ashley Finch, Trustin Varnado, Soumya Venkat, Shana B. Edwards, and Ruchika Budhraja for facilitating our work with Facebook and for their intelligence, determination and commitment to this task. Also, we would like to thank the External Affairs team, especially Lindsay Elin and Monique Dorsainvil for their work getting the Audit off the ground and also for their intelligence, determination and commitment to the Audit as well. A special thanks to Sheryl Sandberg for her ongoing leadership and support throughout this important process over the last two years.

The Auditors would like to acknowledge Dalia Hashad and Tanya Clay House, consultants to Laura Murphy & Associates.

The Auditors would also like to thank the team from Relman Colfax, which included Stephen Hayes, Eric Sublett, Alexa Milton, Tanya Sehgal, and Zachary Best.

## Introduction by Laura W. Murphy

This report marks the end of a two-year audit process that started in May of 2018 and was led by me and supported by Megan Cacace, a partner at the civil rights law firm Relman Colfax (along with a team from Relman Colfax). The report is cumulative, building on two previous updates that were published in December 2018 and June 2019.

The Audit began at the behest of civil rights organizations and members of Congress, who recognized the need to make sure important civil rights laws and principles are respected, embraced, and robustly incorporated into the work at Facebook.

Civil rights groups have been central to the process, engaging tirelessly and consistently in the Audit effort. We interviewed and solicited input from over 100 civil rights and social justice organizations, hundreds of advocates and several members of Congress. These groups championed the Audit as a collaborative and less adversarial mechanism for effecting systemic change at Facebook. They pointed out that civil rights challenges emerge in almost every aspect of the company, from its products to its Community Standards and enforcement practices.

At the outset, the groups identified the topics on which they wanted Facebook’s greater focus, including voter suppression and voter information, building a civil rights accountability infrastructure, content moderation and enforcement (including hate speech and harassment), advertising targeting and practices, diversity and inclusion, fairness in algorithms and the civil rights implications of privacy practices. All of those topics are addressed in this final report — with varying degrees of depth because of time limitations — in addition to new topics we’ve added to the scope, including COVID-19 and the 2020 census.

The Civil Rights Audit was not limited to racial justice issues. Civil rights are the rights of individuals to be free from unfair treatment or discrimination in the areas of education, employment, housing, credit, voting, public accommodations, and more — based on certain legally-protected characteristics identified in a variety of state and federal laws. Those protected classes include race, sex, sexual orientation, gender identity, disability, national origin, religion, and age, among other characteristics. Our work applies to all of those groups. Our work also applies to every user of Facebook who will benefit from a platform that reduces discrimination, builds inclusion and tamps down on hate speech activity.

When I first started on this project, there was no commitment to publish reports and top management was not actively engaged. With pressure from advocates, that changed. Chief Operating Officer Sheryl Sandberg deserves kudos for taking over as the point person for this work and developing important relationships with civil rights leaders. She also enlisted many other senior executives in this work, including CEO Mark Zuckerberg. Throughout the Audit process, Facebook had dozens of interactions with a broad array of civil rights leaders, resulting in more face-to-face contact with Facebook executives than ever before. This Audit enabled groundbreaking convenings with civil right leaders at Facebook headquarters in Menlo Park, CA, in Atlanta, GA and in Washington, DC.

Many Facebook staff supported the work of Megan Cacace and myself (the Auditors). The Auditors were assigned a three-person full-time program management team, a partially dedicated team of 15+ employees across product, policy, and other functions — and the ongoing support of a team of Executives who, in addition to their full-time positions, sit on the Civil Rights Task Force. It is also worth noting that, since the Audit started, the External Affairs team that manages relationships with the civil rights community has grown in both size and resources.

This collective effort yielded a number of positive outcomes for civil rights that we detail in the report.

### The Seesaw of Progress and Setbacks

The purpose of this Audit has always been to ensure that Facebook makes real and lasting progress on civil rights, and we do believe what's listed below illustrates progress. Facebook is in a different place than it was two years ago — some teams of employees are asking questions about civil rights issues and implications before launching policies and products. But as I've said throughout this process, this progress represents a start, not a destination.

While the audit process has been meaningful, and has led to some significant improvements in the platform, we have also watched the company make painful decisions over the last nine months with real world consequences that are serious setbacks for civil rights.

The Auditors believe it is important to acknowledge that the Civil Rights Audit was a substantive and voluntary process and that the company used the process to listen, plan and deliver on various consequential changes that will help advance the civil rights of its users, including but not limited to:

- **Reaching a historic civil rights settlement** in March 2019, under which Facebook committed to implement **a new advertising system** so advertisers running US housing, employment, and credit ads will no longer be allowed to target by age, gender, or zip code — and Facebook agreed to a much smaller set of targeting categories overall. Since then, the company has delivered on its commitment and gone above and beyond the settlement with additional transparency and targeting measures that are outlined in the report.

- **Expanding their voter suppression policies.** When we started the Audit process in 2018, Facebook had a voter suppression policy in place, but it was more limited. At our urging, the policy is now much more expansive and includes threats that voting will result in adverse law enforcement consequences or statements that encourage coordinated interference in elections. In addition, the company adopted a new policy prohibiting threats of violence relating to voting, voter registration or the outcome of elections. Facebook has engaged two voting rights expert consultants to work with and train the policy, product, and operations teams responsible for enforcing against voter suppression. Nonetheless, recent decisions about Trump posts related to mail-in-ballots in Michigan and Nevada on May 20 and California on May 26 threaten that progress and permit others to use the platform to spread damaging misinformation about voting. Several other voting changes are identified in the elections chapter of the report.

- **Creating a robust census interference policy.** Facebook developed robust policies to combat census interference. It has worked closely with the civil rights community 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. Facebook has also engaged a census expert who consults with and trains policy, product, and operations teams responsible for enforcing against census suppression.

- **Taking steps to build greater civil rights awareness and accountability** across the company on a long-term basis. Facebook has acknowledged that no one on its senior leadership team has expertise in civil rights. Thus, the Auditors are heartened that Facebook has committed to hiring an executive at the VP level to lead its work on civil rights. This person will have expertise in civil rights law and policy and will be empowered to develop processes for identifying and addressing civil rights risks before products and policies are launched. The Civil Rights VP will have dedicated program management support and will work to build out a long-term civil rights infrastructure and team. The company also committed to developing and launching civil rights training for several groups of employees, including the Civil Rights Task Force, which is made up of senior leadership across key verticals in the company. These commitments must be approached with urgency.

- **Improved Appeals and Penalties process.** Facebook adopted several procedural and transparency changes to how people are penalized for what they post on Facebook. For example, the company has introduced an "account status" feature that allows users to view prior Community Standards violations, including which Community Standard was violated, as well as an explanation of restrictions imposed on their account and details on when the restrictions will expire.

- **More frequent consultations with civil rights leaders.** Facebook leadership and staff has more consistently engaged with leaders in the civil rights community and sought their feedback, especially in the voting and census space.

- **Changing various content moderation practices,** including an expanded policy that bans explicit praise, support and representation of white nationalism and white separatism, and a new policy that prohibits content encouraging or calling for the harassment of others, which was a top concern of activists who are often targeted by coordinated harassment campaigns. Facebook also launched a series of pilots to combat hate speech enforcement errors, a well-documented source of frustration for activists and other users who condemn hate speech and violence to be incorrectly kicked off the platform.

- **Taking meaningful steps to create a more diverse and inclusive senior leadership team and culture.** It has, for example, elevated the role of the Chief Diversity Officer to report directly to the Chief Operating Officer and to play an active role in key executive decision meetings — and to increase the number of leadership positions held by people of color by 30%, including 30% more Black people, over the next five years.

- **Investing in diverse businesses and vendors.** Facebook has made commitments to partner with minority vendors and has made more funding available for minority businesses and social justice groups, including a recent announcement that it will spend at least $100 million annually with Black-owned suppliers. This is part of the company's effort to double annual spending with US companies certified as minority, women, veteran, LGBTQ, or disabled-owned suppliers to $1 billion by the end of 2021. Facebook has also committed to support a $100 million investment in Black-owned small businesses, content creators and non-profits who use the platform.

- **Investing in a dedicated team to focus on studying responsible Artificial Intelligence methodologies** and building stronger internal systems to address algorithmic bias.
- **Implementing significant changes to privacy policies and systems** as a result of the Federal Trade Commission settlement that includes a privacy review of every new or modified product, service or practice before it is implemented.

With each success the Auditors became more hopeful that Facebook would develop a more coherent and positive plan of action that demonstrated, in word and deed, the company's commitment to civil rights. Unfortunately, in our view Facebook's approach to civil rights remains too reactive and piecemeal. Many in the civil rights community have become disheartened, frustrated and angry after years of engagement where they implored the company to do more to advance equality and fight discrimination, while also safeguarding free expression. As the final report is being issued, the frustration directed at Facebook from some quarters is at the highest level seen since the company was founded, and certainly since the Civil Rights Audit started in 2018.

The Auditors vigorously advocated for more and would have liked to see the company go further to address civil rights concerns in a host of areas that are described in detail in the report. These include but are not limited to the following:

- **A stronger interpretation of its voter suppression policies** — an interpretation that makes those policies effective against voter suppression and prohibits content like the Trump voting posts — and **more robust and more consistent enforcement of those policies** leading up to the US 2020 election.
- **More visible and consistent prioritization of civil rights** in company decision-making overall.
- **More resources invested to study and address organized hate** against Muslims, Jews and other targeted groups on the platform.
- **A commitment to go beyond banning explicit references to white separatism and white nationalism** to also prohibit express praise, support and representation of white separatism and white nationalism even where the terms themselves are not used.
- **More concrete action** and specific commitments to take steps **to address concerns about algorithmic bias or discrimination.**

This report outlines a number of positive and consequential steps that the company has taken, but at this point in history, the Auditors are concerned that those gains could be obscured by the vexing and heartbreaking decisions Facebook has made that represent significant setbacks for civil rights.

Starting in July of 2019, while the Auditors were embarking on the final phase of the audit, civil rights groups repeatedly emphasized to Facebook that their biggest concerns were that domestic political forces would use the platform as a vehicle to engage in voter and census suppression. They said that they did not want 2020 to be a repeat of 2016, the last presidential election, where minority communities – African Americans especially — were targeted for racial division, disinformation and voter suppression by Russian actors.

The civil rights groups also knew that the Civil Rights Audit was not going to go on forever, and therefore, they sought a commitment from Sheryl Sandberg and Mark Zuckerberg that a robust civil rights infrastructure be put in place at Facebook.

Soon after these civil rights priorities were relayed by the Auditors, in September of 2019 Facebook’s Vice President of Global Affairs and Communications, Nick Clegg, said that Facebook had been and would continue to exempt politicians from its third-party fact checking program. He also announced that the company had a standing policy to treat speech from politicians as newsworthy that should be seen and heard and not interfered with by Facebook unless outweighed by the risk of harm. The civil rights community was deeply dismayed and fearful of the impact of these decisions on our democratic processes, especially their effect on marginalized communities. In their view, Facebook gave the powerful more freedom on the platform to make false, voter-suppressive and divisive statements than the average user.

Facebook CEO, Mark Zuckerberg, in his October 2019 speech at Georgetown University began to amplify his prioritization of a definition of free expression as a governing principle of the platform. In my view as a civil liberties and civil rights expert, Mark elevated a selective view of free expression as Facebook’s most cherished value. Although the speech gave a nod to “voting as voice” and spoke about the ways that Facebook empowers the average user, Mark used part of the speech to double down on the company’s treatment of politicians’ speech.

The Auditors have expressed significant concern about the company’s steadfast commitment since Mark’s October 2019 Georgetown speech to protect a particular definition of free expression, even where that has meant allowing harmful and divisive rhetoric that amplifies hate speech and threatens civil rights. Elevating free expression is a good thing, but it should apply to everyone. When it means that powerful politicians do not have to abide by the same rules that everyone else does, a hierarchy of speech is created that privileges certain voices over less powerful voices. The prioritization of free expression over all other values, such as equality and non-discrimination, is deeply troubling to the Auditors.

Mark Zuckerberg’s speech and Nick Clegg’s announcements deeply impacted our civil rights work and added new challenges to reining in voter suppression.

Ironically, Facebook has no qualms about reining in speech by the proponents of the anti-vaccination movement, or limiting misinformation about COVID -19, but when it comes to voting, Facebook has been far too reluctant to adopt strong rules to limit misinformation and voter suppression. With less than five months before a presidential election, it confounds the Auditors as to why Facebook has failed to grasp the urgency of interpreting existing policies to make them effective against suppression and ensuring that their enforcement tools are as effective as possible. Facebook's failure to remove the Trump voting-related posts and close enforcement gaps seems to reflect a statement of values that protecting free expression is more important than other stated company values.

Facebook's decisions in May of 2020 to let stand on three posts by President Trump, have caused considerable alarm for the Auditors and the civil rights community. One post allowed the propagation of hate/violent speech and two facilitated voter suppression. In all three cases Facebook asserted that the posts did not violate its Community Standards. The Auditors vigorously made known our disagreement, as we believed that these posts clearly violated Facebook's policies. These decisions exposed a major hole in Facebook's understanding and application of civil rights. While these decisions were made ultimately at the highest level, we believe civil rights expertise was not sought and applied to the degree it should have been and the resulting decisions were devastating. Our fear was (and continues to be) that these decisions establish terrible precedent for others to emulate.

The Auditors were not alone. The company's decisions elicited uproar from civil rights leaders, elected officials and former and current staff of the company, forcing urgent dialogues within Facebook. Some civil rights groups are so frustrated that Facebook permitted these Trump posts (among other important issues such as removing hate speech), that they have organized in an effort to enlist advertisers to boycott Facebook. Worse, some civil rights groups have, at this writing, threatened to walk away from future meetings with Facebook.

While Facebook has built a robust mechanism to actively root out foreign actors running coordinated campaigns to interfere with America's democratic processes, Facebook has made policy and enforcement choices that leave our election exposed to interference by the President and others who seek to use misinformation to sow confusion and suppress voting.

Specifically, we have grave concerns that the combination of the company's decision to exempt politicians from fact-checking and the precedents set by its recent decisions on President Trump's posts, leaves the door open for the platform to be used by other politicians to interfere with voting. If politicians are free to mislead people about official voting methods (by labeling ballots illegal or making other misleading statements that go unchecked, for example) and are allowed to use not-so-subtle dog whistles with impunity to incite violence against groups advocating for racial justice, this does not bode well for the hostile voting environment that can be facilitated by Facebook in the United States. We are concerned that politicians, and any other user for that matter, will capitalize on the policy gaps made apparent by the president's posts and target particular communities to suppress the votes of groups based on their race or other characteristics. With only months left before a major election, this is deeply troublesome as misinformation, sowing racial division and calls for violence near elections can do great damage to our democracy.

Nonetheless, there has also been positive movement in reaction to the uproar. On June 5, 2020, Mark Zuckerberg committed to building products to advance racial justice, and promised that Facebook would reconsider a number of existing Community Standards, including how the company treats content dealing with voter suppression and potential incitement of violence. He also promised to create a voting hub to encourage greater participation in the November 2020 elections, and provide access to more authoritative voting information.

On June 26, 2020 Mark announced new policies dealing with voting on topics ranging from prohibitions against inflammatory ads, the labeling of voting posts, guidance on voter interference policy enforcement, processes for addressing local attempts to engage in voter suppression and labeling and transparency on newsworthiness decisions. The Auditors examine these policies at greater length later in this report (in the Elections and Census 2020 Chapter), but simply put: these announcements are improvements, depending on how they are enforced — with the exception of the voting labels, the reaction to which was more mixed.

Nevertheless, Facebook has not, as of this writing, reversed the decisions about the Trump posts and the Auditors are deeply troubled by that because of the precedent they establish for other speakers on the platform and the ways those decisions seem to gut policies the Auditors and the civil rights community worked hard to get Facebook to adopt.

**Where we go from here**

Facebook has a long road ahead on its civil rights journey, and both Megan Cacace and I have agreed to continue to consult with the company, but with the audit behind us, we are discussing what the scope of that engagement will look like. Sheryl Sandberg will continue to sponsor the work at the company. Mark Zuckerberg said that he will continue to revisit its voter suppression policies, as well as its policies relating to calls for violence by state actors.

These policies have direct and consequential implications for the US presidential election in November 2020, and we will be watching closely. The responsibility for implementing strong equality, non-discrimination and inclusion practices rests squarely with the CEO and COO. They have to own it and make sure that managers throughout the company take responsibility for following through.

As we close out the Audit process, we strongly encourage the company to do three things:

- **Seriously consider, debate and make changes on the various recommendations** that Megan Cacace and I have shared throughout the final report, as well as in previous reports. In particular, it's absolutely essential that the company do more to build out its internal civil rights infrastructure. More expertise is needed in-house, as are more robust processes that allow for the integration of civil rights perspectives.
- **Be consistent and clear about the company's commitment to civil rights laws and principles.** When Congress recently and pointedly asked Facebook if it is subject to the civil rights mandates of the federal Fair Housing Act, it vaguely asserted, "We have obligations under civil rights laws, like any other company." In numerous legal filings, Facebook attempts to place itself beyond the reach of civil rights laws, claiming immunity under Section 230 of the Communications Decency Act. On the other hand, leadership has publicly stated that "one of our top priorities is protecting people from discrimination on Facebook." And, as a result of settling four civil rights lawsuits, the company has embraced civil rights principles in redesigning its advertising system to prevent advertisers from discriminating. Thus, what the Auditors have experienced is a very inconsistent approach to civil rights. Facebook must establish clarity about the company's obligations to the spirit and the letter of civil rights laws.

- **Address the tension of civil rights and free speech head on.** Mark's speech at Georgetown seems to represent a turning point for the company, after which it has placed greater emphasis on free expression. But Megan and I would argue that the value of non-discrimination is equally important, and that the two need not be mutually exclusive. As a longtime champion of civil rights and free expression I understand the crucial importance of both. For a 21st century American corporation, and for Facebook, a social media company that has so much influence over our daily lives, the lack of clarity about the relationship between those two values is devastating. It will require hard balancing, but that kind of balancing of rights and interests has been part of the American dialogue since its founding and there is no reason that Facebook cannot harmonize those values, if it really wants to do so.

In publishing an update on our work in June 2019, I compared Facebook's progress to climbing a section of Mount Everest: the company had made progress, but had certainly not reached the summit. As I close out the Civil Rights Audit with this report, many in the civil rights community acknowledge that progress has been made, but many feel it has been inadequate. In our view Facebook has made notable progress in some areas, but it has not yet devoted enough resources or moved with sufficient speed to tackle the multitude of civil rights challenges that are before it. This provokes legitimate questions about Facebook's full-throated commitment to reaching the summit, *i.e.*, fighting discrimination, online hate, promoting inclusion, promoting justice and upholding civil rights. The journey ahead is a long one that will require such a commitment and a reexamination of Facebook's stated priorities and values.

As outlined in the last audit report, the civil rights community has long recognized the need for a permanent civil rights accountability structure at Facebook. Facebook has acknowledged that it must create internal systems and processes to ensure that civil rights concerns based on race, religion, national origin, ethnicity, disability, sex, gender identity, sexual orientation, age, and other categories are proactively identified and addressed in a comprehensive and coordinated way before products and policies are launched, rather than met with reactive, piecemeal, or ad hoc measures after civil rights impacts have already been felt. The Auditors strongly believe that respecting, embracing and upholding civil rights is both a moral and business imperative for such a large and powerful global social media company.

For the duration of their engagement with Facebook, the Auditors have not just reviewed Facebook's policies and practices relating to civil rights, but have also vigorously elevated real-time civil rights concerns that they identified and/or that were raised by the civil rights community. The Auditors played a crucial role in encouraging Facebook to address those concerns. With the Audit coming to an end, calls for an effective civil rights infrastructure have only become louder.

Last year, the company took important, initial steps toward building the foundation for a civil rights accountability structure. It created a Civil Rights Task Force led by Sheryl Sandberg, committed to providing civil rights training for employees, and agreed to add more civil rights expertise to its team. The Auditors and the civil rights community acknowledged Facebook's progress, but made it clear that more needed to be done. Improving upon accountability efforts initiated in 2019 has been a critical focus for the Auditors since the last audit report, and Facebook agrees that having an infrastructure in place to support civil rights work long-term is critical now that the formal audit is over.

This section provides updates on the commitments Facebook made in 2019, and describes the progress Facebook has since made in continuing to build out its civil rights accountability infrastructure. It also identifies where the Auditors think Facebook has not gone far enough and should do more.

While Facebook should be credited for the improvements it has made since the previous report the Auditors urge Facebook to continue to build out its civil rights infrastructure so that it effectively surfaces and addresses civil rights issues at a level commensurate with the scale and scope of Facebook's reach. Given the breadth and depth of Facebook's reach and its impact on people's lives, Facebook's platform, policies, and products can have significant civil rights implications and real-world consequences. It is critical that Facebook establish a structure that is equal to the task.

## A. Update on Prior Commitments

Last year, Facebook made four commitments to lay the foundation for a civil rights accountability structure: (1) create a Civil Rights Task Force designed to continue after the Audit ends; (2) cement Sheryl Sandberg's leadership of the Task Force; (3) onboard civil rights expertise; and (4) commit to civil rights training. Facebook is currently progressing on all four commitments.

Led by **Sheryl Sandberg**, the **Civil Rights Task Force** continues to serve as a forum for leaders of key departments within the company to discuss civil rights issues, identify potential solutions, share lessons learned, and engage in cross-functional coordination and decision-making on civil rights issues. According to Facebook, the Task Force has discussed issues, including new policies and product features, stronger processes that the company could implement, and recent concerns raised by external stakeholders, including civil rights advocates. The membership of the Task Force has evolved over the last year; Facebook reports that it now includes both:

- Decision-makers and executives overseeing departments such as Product, Advertising, Diversity & Inclusion, Legal, and Policy and Communications.
- A cross-functional team of product and policy managers that have been responsible for the implementation of several commitments listed in the previous report as well as a subset of new commitments made in this report. This group represents various functions including key teams working on elections, hate speech, algorithmic bias, and advertising.

Facebook reports that since March 2019, members of the Task Force have consistently met on a monthly basis. The Task Force’s efforts over the past year include:

- Supporting the development of Facebook’s new census interference policy by engaging with key questions regarding the scope and efficacy of the policy.
- Engaging with the civil rights community in various forums including Facebook’s first ever public civil rights convening in Atlanta in September 2019 and in working sessions with the Auditors and sub-groups within the Task Force over the last 12 months.
- Driving cross-functional coordination across teams so that Facebook could be more responsive to stakeholder requests and concerns. For example, Facebook reports that in 2020 a subset of the Task Force has met weekly with subject matter experts from policy, product and operations to specifically address and make progress on a set of voter suppression proposals suggested by the Auditors.
- Advocating for and supporting all of the work behind the Civil Rights Audit — and helping to remove internal barriers to progress.
- Developing the strategy and implementation plan for several of the key commitments outlined in this report, including the implementation of Facebook’s census policy and the new commitments outlined in the Elections and Census Chapter below.

Facebook also has begun to increase its **in-house civil rights expertise**. It has hired voting and census expert consultants, who are developing trainings for key employees and will be supporting efforts to prevent and address voting/census suppression and misinformation on the platform. In addition, Facebook has started to embed civil rights knowledge on core teams. As discussed in more detail below, the Audit Team maintains that bringing civil rights knowledge in-house is a critical component of the accountability structure and encourages Facebook to continue to onboard civil rights expertise.

In an effort to better equip employees to identify and address civil rights issues, Facebook committed that key employees would undergo customized **civil rights training**. Civil rights law firm Relman Colfax and external voting rights and census experts are working with Facebook’s internal Learning and Development team to develop and launch these trainings, which will be developed in 2020. The civil rights trainings include (1) core training on key civil rights concepts and applications that will be available to all employees; (2) in-depth census and voting-related trainings targeted to key employees working in that space; and (3) customized in-person civil rights training for groups of employees in pivotal roles, including members of the Civil Rights Task Force. (All of these trainings are in addition to the fair housing civil rights training Facebook will be receiving from the National Fair Housing Alliance as part of the settlement of the advertising discrimination lawsuits, discussed in the Advertising Chapter below.)

## B. New Commitments and Developments

Since the last Audit Report, Facebook is now committing to expand its existing accountability structure in several key ways.

**First**, Facebook has created **a senior (Vice President) civil rights leadership role** —a civil rights expert who will be hired to develop and oversee the company’s civil rights accountability efforts, and help instill civil rights best practices within the company. The civil rights leader is authorized and expected to:

- identify proactive civil rights priorities for the company and guide the implementation of those priorities;
- develop systems, processes or other measures to improve the company’s ability to spot and address potential civil rights implications in products and policies before they launch; and
- give voice to civil rights risks and concerns in interactions with leadership, executives, and the Task Force.

Unlike the Task Force, cross functional teams, and embedded employees with civil rights backgrounds who have other primary responsibilities, the civil rights leader’s job will center around the leader’s ability to help the company proactively identify and address civil rights issues.

From the beginning, the civil rights leader will have dedicated cross-functional coordination and project management support, in addition to the support of Sheryl Sandberg, the Civil Rights Task Force and Facebook’s External Affairs policy team, which works closely and has relationships with civil rights stakeholders and groups. Facebook will continue to engage Laura Murphy and outside civil rights counsel Relman Colfax on a consulting basis to provide additional civil rights guidance and resources. In addition to these initial resources, the civil rights leader will be authorized to assess needs within the company and build out a team over time.

**Second**, Facebook has committed to **developing systems and processes to help proactively flag civil rights considerations** for its product and policy teams.

### 1. Policy Review

Civil rights input is critical to Facebook's policy development process — the process by which Facebook writes new rules or updates existing ones to govern the type of content that is and is not permitted on the platform. To help ensure civil rights considerations are recognized and addressed in the policy development process, the civil rights leader will have visibility into the content policy development pipeline. In cases where a policy could have civil rights implications, the civil rights leader: (i) will be part of the working group developing that policy; and (ii) will have the opportunity to voice civil rights concerns directly to policy leadership before the policy is launched and when policy enforcement decisions that rely on cross-functional input are escalated internally.

### 2. Product Review

An important element of an effective civil rights infrastructure is a system for identifying civil rights risks or considerations at the front end and throughout the product development process. Facebook is doing two things in this area:

**(i) Civil Rights Screening:** Through the Audit, Facebook has committed to embedding civil rights screening criteria within certain existing product review processes so that teams can better identify and evaluate potential civil rights concerns. As a starting point, the Auditors have worked to develop civil rights issue-spotting questions that teams working on products relating to advertising, election integrity, algorithmic fairness, and content distribution (*e.g.,* News Feed) will embed into existing product review processes. Currently, all but one of the product review processes these questions will be inserted into are voluntary, rather than mandated reviews required of all products. That being said, Facebook has authorized the civil rights leader to look for ways to further develop or improve civil rights screening efforts, and provide input into review processes both to help make sure civil rights risks are correctly identified and to assist teams in addressing concerns raised.

**(ii) Responsible Innovation:** Independent of the Audit, Facebook has been building out a full-time, permanent team within the Product organization that is focused on Responsible Innovation — that is, helping to ensure that Facebook's products minimize harm or negative impacts and maximize good or positive impacts. The Responsible Innovation team's stated priorities include: (a) developing trainings and tools that product teams can use to identify and mitigate potential harms (including civil rights impacts) early on in product development; (b) helping employees understand where to go and what to do if concerns are identified; (c) tracking potential harms identified across products and supporting escalation paths to help ensure risks are effectively addressed; and (d) engaging with outside experts and voices to provide input and subject matter expertise to help product teams integrate diverse perspectives into their product development process.

Facebook indicates that the Responsible Innovation team is growing but currently consists of engineers, researchers, designers, policy experts, anthropologists, ethicists, and diversity, equity, and inclusion experts. The Auditors met with key members of the team, and discussed the importance of avoiding civil rights harms in product development. The Responsible Innovation team is focusing on a handful of key issues or concepts as it builds out its infrastructure; fairness (which includes civil rights) is one of them, along with freedom of expression, inclusive access, economic opportunity, individual autonomy, privacy, and civic participation.

While not limited to civil rights or designed as a civil rights compliance structure specifically, in the Auditors' view, Responsible Innovation is worth mentioning here because the trainings, tools, and processes that team is looking to build may help surface civil rights considerations across a wider set of products than the subset of product review processes into which Facebook has agreed to insert civil rights screening questions (as discussed above). The Auditors also recommend that Facebook add personnel with civil rights expertise to this team.

## C. Recommendations from the Auditors

The Auditorsrecognize Facebook's commitments as important steps forward in building a long-term civil rights accountability structure. These improvements are meaningful, but, in the Auditors' view, they are not sufficient and should not be the end of Facebook's progress.

### 1. Continue to Onboard Expertise

In keeping with Facebook's 2019 commitment to onboard civil rights expertise, the Auditors recommend that Facebook continue to bring civil rights expertise in-house — especially on teams whose work is likely to have civil rights implications (such as elections, hate speech, advertising, algorithmic bias, *etc*.). In the Auditors' view, the more Facebook is able to embed civil rights experts onto existing teams, the better those teams will be at identifying and addressing civil rights risks, and the more civil rights considerations will be built into the company's culture and DNA.

### 2. Build Out the Civil Rights Leader's Team

The Auditors also believe that the civil rights leader will need the resources of a team to meet the demands of the role, and allow for effective civil rights screenings of products and policies.

To the first point, the Auditors believe a team is necessary to ensure the civil rights leader has the capacity to drive a proactive civil rights accountability strategy, as opposed to simply reacting to concerns raised externally or through review processes. There is a difference between working full-time (as members of the civil rights leader's team) on identifying and resolving civil rights concerns, and having civil rights be one component of a job otherwise focused on other areas (as is the case for members of the Civil Rights Task Force). While the Auditors recognize that Facebook has agreed to allow the civil rights leader to build a team over time, they are concerned that without more resources up front, the leader will be overwhelmed before there is any opportunity to do so. From the Auditors' vantage point, the civil rights leader's responsibilities — identifying civil rights priorities, designing macro-level systems for effective civil rights product reviews, participating in policy development, being involved in real-time escalations on precedential policy enforcement decisions, and providing guidance on civil rights issues raised by stakeholders — is far more than any one person can do successfully, even with support.

To the second point, equipping the civil rights leader with the resources of a team would likely make civil rights review processes more successful. It would allow those reviews to be conducted and/or supervised by people with civil rights backgrounds who sit outside the product and policy teams and whose job performance depends on successfully flagging risks — as opposed to having reviews done by those with different job goals (such as launching products) which may not always align with civil rights risk mitigation. The Auditors believe that for review processes to be most effective, those conducting civil rights screens must be able (through internal escalation if necessary) to pause or stop products or policies from going live until civil rights concerns can be addressed. The civil rights leader (and those on his/her team) will be best equipped to do so because it is aligned with their job duties and incentives. Because the civil rights leader will not be able to personally supervise reviews effectively at the scale that Facebook operates (especially if further review processes are built out) a team seems mandatory.

### 3. Expand Civil Rights Product Review Processes.

The Auditors acknowledge that embedding civil rights considerations into existing (largely voluntary) product review processes relating to advertising, election integrity, algorithmic fairness, and News Feed is progress, and a meaningful step forward. But, as Facebook continues to develop its civil rights infrastructure, the Auditors recommend: (i) adopting comprehensive civil rights screening processes or programs that assess civil rights risks and implications across all products; and (ii) making such screens mandatory and able to require product changes.

While Responsible Innovation is a positive development, it was not designed to replace a civil rights accountability infrastructure. The Responsible Innovation approach involves considering a host of dimensions or interests (which may be competing) and providing tools to help employees decide which interests should be prioritized and which should give way in making a given product decision. The framework does not dictate which dimensions need to be prioritized in which circumstances, and as a result, is not designed to ensure that civil rights concerns or impacts will be sufficient to require a product change or delay a product launch.

### 4. Require Civil Rights Perspective in Escalation of Key Content Decisions.

Difficult content moderation questions — in particular those that involve gray areas of content policy or new applications not explicitly contemplated by existing policy language — are sometimes escalated to leadership. These “escalations” are typically reactive and time-sensitive. But, as seen with recent decisions regarding President Trump’s posts, escalations can have substantial and precedent-setting implications for how policies are applied in the future — including policies with significant civil rights impacts. To help prevent civil rights risks from being overlooked during this expedited “escalation,” the Auditors recommend that the civil rights leader be an essential (not optional) voice in the internal escalation process for decisions with civil rights implications (as determined by the civil rights leader). To the Auditors, this means that the civil rights leader must be “in the room” (meaning in direct dialogue with decision-makers) when decisions are being made and have direct conversations with leadership.

### 5. Prioritize Civil Rights

In sum, the Auditors’ goal has long been to build a civil rights infrastructure at Facebook that ensures that the work of the Audit — the focused attention on civil rights concerns, and the company’s commitment to listen, accept sometimes critical feedback, and make improvements — continues long after the formal Civil Rights Audit comes to a close. The Auditors recognizes that Facebook is on the path toward long-term civil rights accountability, but it is not there yet. We urge the company to build and infrastructure that is commensurate to the significant civil rights challenges the company encounters.

The Auditors believe it is imperative that Facebook commit to building upon the foundation it has laid. It is critical that Facebook not only invest in its civil rights leader (and his or her team), in bringing on expertise, and in developing civil rights review processes, but it must also invest in civil rights as a priority. At bottom, all of these people, processes, and structures depend for their effectiveness on civil rights being vigorously embraced and championed by Facebook leadership and being a core value of the company.

## Chapter Two: Elections & Census 2020

With both a presidential election and a decennial census, 2020 is an incredibly important year for Facebook to focus on preventing suppression and intimidation, improving policy enforcement, and shoring up its defenses against coordinated threats and interference. As such, the Audit Team has prioritized Facebook's election and census policies, practices, and monitoring and enforcement infrastructure since the last report.

The current COVID-19 pandemic has had a huge impact on Americans' ability to engage in all forms of civic participation. Understandably, it has changed how elections and the census are carried out and has influenced the flow of information about how to participate in both. On social media, in particular, the pandemic has led candidates and elected officials to find new ways to reach their communities online, but it has also prompted misinformation and new tactics designed to suppress participation, making Facebook's preparation and responsiveness more important than ever.

This chapter provides an update on Facebook's prior elections and census commitments (made in the June 2019 Audit Report), describes Facebook's response to the COVID-19 pandemic as it relates to elections and census, and details new commitments and developments. Facebook has made consequential improvements directed at promoting census and voter participation, addressing suppression, preventing foreign interference, and increasing transparency, the details of which are described below.

This report is also transparent about the places where the Auditors believe that the company has taken harmful steps backward on suppression issues, primarily in its decision to exempt politicians' speech from fact checking, and its failure to remove viral posts, such as those by President Trump, that the Auditors (and the civil rights community) believe are in direct violation of the company's voter suppression policies.

### A. Update on Prior Elections and Census Commitments

In the June 2019 Audit Report, Facebook made commitments to develop or improve voting and census-related policies and build out its elections/census operations, resources, and planning. Updates on these commitments are provided below.

#### 1. Policy Improvements

**(i) Prohibiting Census Suppression and Interference.** After listening to repeated feedback and concern raised by stakeholders about census interference on social media, in 2019, Facebook committed to treating the 2020 census like an election, which included developing and launching a census interference policy designed to protect the census from suppression and interference as the company has done for voting. Facebook made good on its policy commitment. Through a months-long process involving the Audit Team, the US Census Bureau, civil rights groups, and census experts, Facebook developed a robust census interference policy, which was formally launched in December 2019.

The census interference policy extends beyond mere misrepresentations of how and when to fill out the census to the types of threats of harm or negative consequences that census experts identify as particularly dangerous and intimidating forms of suppression — especially when targeted (as they often are) toward specific communities. This new policy is supported by both proactive detection technology and human review, and violating content is removed regardless of who posts it. Notably, the policy applies equally to content posted by politicians and any other speakers on the platform.

Specifically, Facebook’s new census interference policy prohibits:

- Misrepresentation of the dates, locations, times and methods for census participation;
- Misrepresentation of who can participate in the census and what information and/or materials must be provided in order to participate;
- Content stating that census participation may or will result in law enforcement consequences;
- Misrepresentation of government involvement in the census, including that an individual’s census information will be shared with another government agency; and
- Calls for coordinated interference that would affect an individual’s ability to participate in the census (enforcement of which often requires additional information and context).

Many in the civil rights community and the US Census Bureau lauded the policy. The Leadership Conference on Civil and Human Rights described it as industry leading: “the most comprehensive policy to date to combat census interference efforts on its platform,” and the Census Bureau thanked Facebook for the policy and its efforts to “ensure a complete and accurate 2020 Census.” Others celebrated the policy while taking a cautionary tone. Rashad Robinson of Color of Change said, “Facebook is taking an important step forward by attempting to promote an accurate Census count, but success will depend on consistent enforcement and implementation [...] This updated policy is only as good as its enforcement and transparency, which, to be clear, is an area that Facebook has failed in the past.”

Enforcement of the policy had an early setback, but according to Facebook, has since improved. In March 2020, an ad was posted by the Trump Campaign that appeared to violate the new census interference policy, but it took Facebook over 24 hours to complete its internal escalation review and reach its final conclusion that the ad did, in fact, violate the new policy and should be removed. The delay caused considerable concern within the civil rights community (and among the Auditors) — concern that Facebook’s enforcement would negate the robustness of the policy. After the incident, however, Facebook conducted an internal assessment to identify what went wrong in its enforcement/escalation process and make corrections. While Facebook developed its census interference policy with expert input, it is difficult to anticipate in advance all the different ways census interference or suppression content could manifest. Because the violating content in the ad took a form that had not been squarely anticipated, Facebook’s removal of the ad was delayed. After this initial enforcement experience, Facebook focused attention on ensuring that its enforcement scheme is sufficiently nimble to promptly address interference and suppression manifested in unanticipated ways.

Since then, Facebook has identified and removed a variety of content under the policy, including false assertions by multiple public figures that only citizens may participate in the census. Facebook has also demonstrated an improved ability to adapt to unanticipated circumstances: as described in more detail below, it has proactively identified and removed violating content using the COVID-19 pandemic to suppress or interfere with census participation — content which Facebook could not have anticipated at the time the census interference policy was developed.

**(ii) Policy Updates to Prevent Voter Intimidation.** Since the 2016 election, Facebook has expanded its voter suppression policy to prohibit:

- Misrepresentation of the dates, locations, and times, and methods for voting or voter registration;
- Misrepresentation of who can vote, qualifications for voting, whether a vote will be counted, and what information and/or materials must be provided in order to vote; and
- Threats of violence relating to voting, voter registration, or the outcome of an election.

The June 2019 Audit Report acknowledged, however, that further improvements could be made to prevent intimidation and suppression, which all too often is disproportionately targeted to specific communities. Facebook committed to exploring further policy updates, specifically updates directed at voter interference and inflammatory ads.

Since the last report, Facebook has further expanded its policies against voter intimidation to now also prohibit:

- Content stating that voting participation may or will result in law enforcement consequences (*e.g*., arrest, deportation, imprisonment);
- Calls for coordinated interference that would affect an individual's ability to participate in an election (enforcement of which often requires additional information and context); and
- Statements of intent or advocacy, calls to action, or aspirational or conditional statements to bring weapons to polling places (or encouraging others to do the same).

These policy updates prohibit additional types of intimidation and threats that can chill voter participation and stifle users from exercising their right to vote.

**(iii) Expansion of Inflammatory Ads Policy.** In the June 2019 report, Facebook committed to further refining and expanding the Inflammatory Ads policy it adopted in 2018. That policy prohibits certain types of attacks or fear-mongering claims made against people based on their race, religion, or other protected characteristics that would not otherwise be prohibited under Facebook's hate speech policies. When the policy was first adopted, it prohibited claims such as allegations that a racial group will "take over" or that a religious or immigrant group as a whole represents a criminal threat.

Facebook expanded its Inflammatory Ads policy (which goes beyond its Community Standards for hate speech) on June 26, 2020 to also prohibit ads stating that people represent a "threat to physical safety, health or survival" based on their race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, serious disease or disability, or immigration status. Content that would be prohibited under this policy include claims that a racial group wants to "destroy us from within" or that an immigrant group "is infested with disease and therefore a threat to health and survival of our community." The Auditors recognize this expansion as progress and a step in the right direction. The Auditors believe, however, that this expansion does not go far enough in that it is limited to physical threats; it still permits advertisers to run ads that paint minority groups as a threat to things like our culture or values (*e.g.*, claiming a religious group poses a threat to the "American way of life." The Auditors are concerned that allowing minority groups to be labeled as a threat to important values or ideals in targeted advertising can be equally dangerous and can lead to real-world harms, and the Auditors urge Facebook to continue to explore ways to expand this policy.

As part of the same policy update, Facebook will also prohibit ads with statements of inferiority, expressions of contempt, disgust or dismissal and cursing when directed at immigrants, asylum seekers, migrants, or refugees. (Such attacks are already prohibited based on race, gender, ethnicity, religious affiliation, caste, gender identity, sexual orientation, and serious disease or disability.) The Auditors believe this is an important and necessary policy expansion, and are pleased that Facebook made the change.

**(iv) Don't Vote Ads Policy & Don't Participate in Census Ads Policy.** Through the Audit, Facebook committed in 2019 to launching a "Don't Vote Ads Policy"— a policy designed to prohibit ads targeting the US expressing suppressive messages encouraging people not to vote, including the types of demobilizing ads that foreign actors targeted to minority and other communities in 2016. Facebook launched that policy in September 2019.

In keeping with its commitment to civic groups and the Census Bureau to treat the 2020 census like an election, Facebook also created a parallel policy prohibiting ads designed to suppress participation in the census through messages encouraging people not to fill it out. Facebook launched the parallel "Don't Participate in Census Ads Policy" at the same time in September 2019.

Together, these new ads policies prohibit "ads targeting the US that portray voting or census participation as useless or meaningless and/or advise users not to vote or participate in a census."

### 2. Elections & Census Operations, Resources, and Planning

After the 2018 midterm elections, civil rights groups were encouraged by the operational resources Facebook placed in its war room, but expressed great concern that the war room capability was created just 30 days prior to Election Day in 2018. With that concern in mind, the Auditors urged Facebook to take a more proactive posture for the 2020 elections, and in the June 2019 Audit Report, Facebook communicated its plan to stand up a dedicated team focused on US Elections and Census, supervised by a single manager in the US.

**(i) 24/7 Detection and Enforcement.** In the lead up to the 2020 election and census, Facebook put dedicated teams and technology in place 24/7 to detect and enforce its rules against content that violates the voting and census interference policies. The teams bring together subject matter experts from across the company — including employees in threat intelligence, data science, software engineering, research, community operations and legal.

Facebook reports that having these fully dedicated teams in place allows them to: (1) conduct real-time monitoring to find and quickly remediate potential harm, including content that violates the company's policies on voter suppression or hate speech; (2) investigate problems quickly and take action when warranted; and (3) track trends in adversarial behavior and spikes in volume that are observed on the platform.

Facebook's 24/7 detection and enforcement is further supplemented by its Election Operations Center ("EOC") (formerly known as the war room), which allows for increased coordination and rapid response when content volumes and escalations are higher than normal (*e.g.,* in the days leading up to a presidential primary election). The EOC was assembled during this election cycle for each democratic presidential debate and has been in operation for all primary elections. The company has used the debates and primary season to refine playbooks and protocols for improved operations.

Prior to COVID-19 EOC personnel would come together in the same physical rooms, but since the outbreak Facebook has shifted to a virtual model, where the same teams coordinate in real-time by video conference. Facebook asserts that the groundwork laid prior to COVID-19 (*e.g.,* playbooks and protocols) have been important in ensuring that remote work conditions have not had a negative impact on the EOC's effectiveness.

**(ii) New Partnerships.** To enhance its elections and census operations and resources, Facebook has also created new partnerships. On the census side, Facebook has been working closely with the Census Bureau to help ensure a fair and accurate census, sharing information during weekly calls to discuss emerging threats and to coordinate efforts to disrupt attempted census interference. Facebook has also partnered with the Census Bureau, civil rights groups, and non-profit organizations with expertise reaching under-represented communities to allow for increased monitoring and reporting of Facebook content that appears to violate the company's census-related policies. Facebook provided these partners with tools and training to enable them to monitor the platform in real time for census interference and suppression, and flag content that may violate Facebook's policies for review by Facebook's operations team.

Facebook has a similar program of partnership on the voting side—partnering with voting rights and election protection organizations and outfitting them with training and Facebook tools that allow partners to conduct platform-wide searches, track content spreading online, and flag potentially violating content. Content flagged by these partners as violating can then be reviewed by trained reviewers at Facebook. Facebook has expanded this program for 2020, extending the opportunity to participate to more than 30 voting rights and election protection groups.

As stated in the 2019 audit report, Facebook continues to engage with secretaries of state, elections directors, and national organizing bodies such as the National Association of Secretaries of State and the National Association of State Election Directors. Facebook works with these offices and organizations to help track violating content and misinformation related to elections and the census, so teams can review and take appropriate action. The company also works directly with election authorities to connect people with accurate information about when and how to vote. Connecting people to accurate voting information is especially critical in light of COVID-19’s impact on the 2020 election.

Facebook has provided each of these reporting partners (census protection groups, voting rights and election protection groups, and state election officials) access to CrowdTangle, (a free social media monitoring tool owned by Facebook) to help them quickly identify misinformation and voter and census interference and suppression. CrowdTangle surfaces content from elected officials, government agencies, colleges and universities, as well as local media and other public accounts. They have also created several public live displays that anyone can use, for example, a display that shows what US presidential candidates are posting on Facebook and Instagram in one dashboard. In addition to the 2020 presidential candidates, CrowdTangle allows anyone to track what Senators, Representatives in the House and Governors are posting on both their official and campaign Pages.

**(iii) Expert Consultants and Training.** In response to concerns that knowledge of civil rights, voting rights, the census process, and forms of suppression and intimidation is critical to policy development and enforcement strategies, Facebook agreed in 2019 to hire a voting rights consultant and a census consultant to provide guidance and training to voting/census policy and ads teams, content review supervisors, and those on internal escalation teams in advance of the census and 2020 elections. The training would cover topics such as the history of voter/census suppression, examples of suppression, and Facebook’s voting and census policies and escalation protocols. In addition to training, the consultants would provide guidance on policy gaps, surface voting/census related concerns raised by external groups, and help advise the company in real time as tricky voting/census-related questions are escalated internally.

Facebook has hired and onboarded Franita Tolson and Justin Levitt as expert voting rights consultants, and for the company’s expert census consultant, it is working with Beth Lynk from the Leadership Conference on Civil and Human Rights. Franita Tolson is a Professor at the USC Gould School of Law, and focuses on voting rights, election law, and the Fourteenth and Fifteenth Amendments. Justin Levitt served as Deputy Assistant Attorney General for Civil Rights of the US Department of Justice, with voting rights as one of his primary areas of focus. He is now a Professor at Loyola School of Law. Beth Lynk is the Census Counts Campaign Director at the Leadership Conference on Civil and Human Rights.

These three consultants have met with relevant internal teams and begun preparing trainings. Beth Lynk will provide supplemental training and ongoing guidance as Facebook continues to enforce its census interference policy. Franita Tolson is in the process of preparing a voting training, which will be delivered in July 2020. Aside from preparing trainings, the consultants will provide voting-related guidance to Facebook on an ongoing basis, including support for the Election Operations Center.

While the Auditors were disappointed that it took longer than anticipated to hire and onboard expert consultants, the Auditors also acknowledge that the delay was due, in large part, to the fact that it took longer than expected to compile a sufficiently large pool of qualified applicants to conduct a thorough and inclusive hiring process. While Facebook has thoughtfully engaged with external voting rights and census advocates, the Auditors believe that by not onboarding and embedding the consulting experts before the end of 2019, as was originally planned, Facebook lost meaningful opportunities to integrate their guidance and advice into their policy and enforcement decision-making process.

## B. Response to COVID-19 Pandemic and Impact on Elections/Census

The COVID-19 pandemic has had a monumental impact on our country and the world, and will likely continue to do so for some time. In addition to impacting lives, jobs, and our daily existences, the pandemic will have ripple effects on elections, voting, and the census — the full implications of which are not yet certain. Accordingly, while not a planned Audit Report topic, because of COVID-19's potential impact on voting and the census, an update on Facebook's COVID response in these areas is warranted.

**Impact on Voter/Census Suppression & Enforcement**

As the pandemic spread, Facebook recognized the possibility of new forms of voter and census suppression relating to the virus. Facebook focused resources on detecting such content and proactively provided content reviewers with clear enforcement guidance on how its policies applied to COVID to ensure that violating content would be removed. For example, Facebook took steps to ensure that content reviewers were trained to remove, as violations of Facebook's voter and census interference policies, false statements that the election or census had been cancelled because of COVID-19.

Facebook has detected and removed various forms of suppressive content related to COVID, such as false statements about the timing of elections or methods for voting or participating in the census. Facebook says that from March to May 2020, it removed more than 100,000 pieces of Facebook and Instagram content in the US (a majority of which were COVID-related) for violating its voter interference policies – virtually all of which were removed proactively before being reported.

While Facebook closed many of its physical content review locations as a result of the pandemic and sent contract content reviewers home for their own safety (while continuing to pay them), Facebook has made clear this shift should not negatively impact its ability to enforce elections and census-related policies. Facebook indicates that unlike some other policies, the content reviewers that help enforce Facebook's elections and census policies include full-time Facebook employees (in addition to contract reviewers) who are able to work remotely. Facebook's elections and census policies are also enforced in part via proactive detection, which is not impacted by the closing of content review sites.

Because COVID-19 has resulted in changes to election times and voting methods, Facebook has focused on both removing voting misinformation and proactively disseminating correct voting information. This includes providing notices (via banners in News Feed) to users in areas where voting by mail is available to everyone, or where there have been last-minute changes to the election. For example, when the Ohio Primary was postponed at the last minute due to COVID-19 concerns, Facebook worked with the Ohio Secretary of State's office and ran a notification at the top of the News Feeds of Ohio users on the original election date confirming that the election had been moved and providing a link to official Ohio elections information. Where election deadlines have been changed, Facebook has been incorporating those changes into its voting products and reminders so that users receive registration or election day information on the correct, updated dates.

In order to prevent suppression and provide access to accurate information, Facebook has committed to continuing to focus attention and resources on COVID-19's implications for elections and the census. The company represents that it will continue to work with state election officials as they make plans for the fall, recognizing that COVID is likely to impact fall general elections as well.

## C. New Elections/Census Developments and Commitments

Since the last report, Facebook has made a number of improvements and new commitments related to promoting census and voter participation, addressing suppression, preventing foreign interference, and increasing transparency. These developments are detailed below.

### 1. Promoting Census Participation

Facebook has taken a number of steps to proactively promote census participation. Because this is the first year that all households can complete the census online, Facebook has a new opportunity to spread awareness of the census and encourage participation. Working in partnership with the US Census Bureau and other non-profits, Facebook launched notifications in the two weeks leading up to Census Day that appeared at the top of Facebook and Instagram feeds, reminding people to complete the census and describing its importance. The notifications also included a link to the Census Bureau's website to facilitate completion of the census. Between Facebook and Instagram, more than 11 million people clicked on the notification link to access the Census Bureau's website where the census could be completed.

Facebook has provided training to organizations leading get-out-the-count outreach to under-represented and hard to count communities, as well as state and local governments, on how to best to leverage Facebook tools to encourage census participation. Facebook also committed $8.7M (in the form of monetary and ad credit donations) with the goal of supporting census coalitions conducting outreach to undercounted communities — African American, Latinx, Youth, Arab American, Native American, LGBTQ+, and Asian American communities and people with disabilities — to ensure an accurate count and broad national reach. Facebook recognizes that this support has likely been even more important to organizations as they shifted to digital outreach strategies and engagement in light of COVID-19. Facebook states that donations supported work intending to highlight the importance of completing the census, and provide information about how to get counted, and therefore were directed toward actions such as phone banking, peer-to peer messaging, and digital and media engagement.

### 2. Promoting Civic Participation

Facebook has a host of products and programs focused on promoting civic participation. It has stated that it coordinates closely with election authorities to provide up-to-date information to its users.

(i) **Election Day Reminders.** On Election Day and the days leading up to it, Facebook is reminding people about the election with a notification at the top of their News Feed. These notices also encourage users to make a post about voting and connect them to election information. Facebook is issuing these reminders for all elections that are municipal-wide or higher and cover a population of more than 5,000 (even in years where there are no national elections) and globally for all nationwide elections considered free or partly-free by Freedom House (a trusted third party entity that evaluates elections worldwide).

(ii) **Facebook Voter Registration Drives.** Facebook launches voter registration reminders via top-of-Feed notifications that provide voter registration information and deadlines, and connect people directly to their state government's voter registration websites (where online registration is available; if online registration is not available, it links to a trusted third-party website, such as TurboVote). These notices also allow people to encourage their friends to register to vote via custom News Feed tools.

(iii) **Voting Information Center and Voter Participation Campaigns.** On June 17, Facebook announced its planned Voting Information Center, which the company hopes will give millions of people accurate information about voting and voting registration. The company has set the goal of helping 4 million voters register this year using Facebook, Instagram and Messenger, and also use the Voting Information Center to help people get to the polls. Facebook's goal is double the estimated 2 million people they helped register in both 2018 and 2016.

Facebook surveyed potential voters and 62% said they believe people will need more information on how to vote this year than they needed in previous elections (presumably due to the impact of COVID-19 on voting). The Voting Information Center is modeled after the COVID-19 information center that the company deployed to connect users to trusted information from health authorities about the pandemic.

Facebook intends to include in the Voting Information Center, information about registering to vote, or requesting an absentee or mail-in ballot, depending on the rules in their state, as well as information on early voting. Facebook reports that people will also be able to see local election alerts from their officials about changes to the voting process, and will include information about polling locations and ID requirements.

Facebook is working with state election officials and other experts to ensure the Voting Information Center accurately reflects the latest information in each state. (Notably, the civil rights community is wary of Facebook relying solely on state election officials for accurate, up-to-date information; both recent and historical examples of state election officials not providing accurate information on Election Day underscore why the civil rights community has urged Facebook to also partner with non-partisan election protection organizations.) The information highlighted in the Voting Information Center will change to meet the needs of voters through different phases of the election like registration periods, deadlines to request a vote-by-mail ballot, the start of early voting, and Election Day.

Starting this summer, Facebook will put the Voting Information Center at the top of people's Facebook and Instagram feeds. Facebook expects more than 160 million people in the US will see the Voting Information Center from July through November.

The civil rights community's response to Facebook's announcement of the Voting Information Center was measured. While they generally viewed it as a positive development, they were clear that it does not make up for the company's seeming failure to enforce its voter suppression policies (as described in Section E.1 below) and recent decisions that allow viral posts labeling officially issued ballots illegal to remain up. In order for users to see the information in the Voting Information Center, they have to take an affirmative step of navigating to the Center or clicking on a link. They have to be affirmatively looking for voting information, whereas viral suppressive content is delivered right to users' or shown in their News Feeds. For many users who view false statements from politicians or viral voting misinformation on Facebook, the damage is already done; without knowing that the information they've seen is false, they may not have reason to visit the Voting Information Center or seek out correct information.

**(iv) Vote-By-Mail Reminders.** In response to COVID, Facebook added an additional product that gives people information about key vote-by-mail deadlines. Typically, this product is sent prior to a state's vote-by-mail ballot request deadline in states where every eligible voter in the state is able to request an absentee ballot. The reminder links to more information about vote by mail and encourages people to share the information with their friends.

**(v) Sponsoring MediaWise's MVP First-time Voter Program.** Facebook is sponsoring MediaWise's program to train first-time voters on media literacy skills in order to make them more informed voters when they go to the polls this November. MediaWise's program includes a voter's guide, in-person (pre-COVID) and virtual training and awareness campaigns. This program has reached 8 million first time voters since January 2020, and will continue to reach additional first-time voters in the lead up to the general election.

**(vi) Partnering with Poynter to Launch MediaWise for Seniors.** Older Americans are increasingly engaged on social media, and as a result, they're exposed to more potential misinformation and false news stories. In June 2020, the Poynter Institute expanded its partnership with Facebook to launch the MediaWise for Seniors program. The purpose of this program is to teach older people key digital media literacy and fact-checking skills — including how to find reliable information and spot inaccurate information about the presidential election as well as COVID-19 — to ensure they make decisions based on fact and not fiction. Facebook reports that through this partnership, MediaWise will host a series of Facebook Lives teaching media literacy, working with Poynter's PolitiFact, create two engaging online classes for seniors on Poynter's e-learning platform, News University, and launch a social media campaign teaching MediaWise tips across platforms.

### 3. Labeling Voting Posts

Civil rights groups have expressed considerable concern about potential voting misinformation on the platform, especially in light of Facebook's exemption of politicians from fact-checking. In light of these concerns, civil rights groups have urged Facebook to take steps to cabin the harm from voting misinformation.

Facebook announced on June 26, 2020 that posts about voting would receive a neutrally worded label that does not opine on the accuracy of the post's content, but directs users to the Voting Information Center for accurate voting information. In other words, the label is not placed on just content that is demobilizing (*e.g.*, posts encouraging people not to vote) or content that is likely to be misinformation or at the edge of what Facebook's policies permit, but instead inserts the label on voting-related content. Facebook reports that this label will be placed on posts that discuss voting, including posts connecting voting with COVID-19, as well as posts that are about the election but do not use those terms (*e.g.*, posts containing words such as ballots, polling places, poll watchers, election day, voter fraud, stolen election, deadline to register, *etc*.).

The reaction to Facebook's new labeling program within the civil rights community (and among the Auditors) was mixed. On the one hand, they recognize the need to ensure access to correct voting information and value the dissemination of correct information, particularly at a time when confusion about voting and the US presidential election may be rampant. On the other hand, there is concern that labeling all voting-related posts (both those that are accurate and those that are spreading misinformation) with neutral language will ultimately be confusing to users and make it more difficult for them to discern accurate from misleading information. While Facebook has asserted that it will remove detected content that violates its voter interference policy, regardless of whether it has a label, civil rights groups remain wary that Facebook could view the labeling as reducing its responsibility to aggressively enforce its Voter Interference Policy — that the company may not have a sense of urgency in removing false information regarding voting methods or logistics because those posts will already have a label directing users to the Voting Information Center. The Auditors have stressed that the new voting labels do not diminish the urgency for Facebook to revisit its interpretation of what constitutes "misrepresentations of methods for voting" under its Voter Interference Policy. For example, voting labels will not alleviate the harm caused by narrow readings of that policy that allow content such as posts falsely alleging that official ballots are illegal. These types of false statements sow suppression and confusion among voters and should be taken down, not merely given a label. Further, because of the likely saturation of labels — the frequency with which users may see them — there is concern that users may quickly ignore them and, as a result, the labels will ultimately not be effective at cabining the harm caused by false voter information. Facebook states it is researching and exploring the best way to implement the labeling program to maximize traffic to the Voting Information Center without oversaturating users. Facebook has represented to the Auditors it will observe how people interact with labels and updateits analysis to increase the labeling program's effectiveness.

### 4. Voter Suppression Improvements

**(i) Voter Interference Policy Enforcement Guidance.** In December 2019, Facebook expanded its Voter Interference Policy to prohibit content that indicates that voting will result in law enforcement consequences. On June 26, 2020, Facebook issued further guidance clarifying what that provision prohibits. Specifically, Facebook made clear that assertions indicating that ICE or other federal immigration enforcement agencies will be at polling places are prohibited under the policy (even if those posts do not explicitly threaten deportation or arrest). The Auditors believe this clarification is an important one, as it signals that Facebook recognizes that messages warning of surveillance of the polls by law enforcement or immigration officials sends the same (suppressive) message as posts that explicitly use words like "arrest" or "deportation."

**(ii) Process for Addressing Local Suppression.** Civil rights groups encouraged Facebook to do more to address one common form of voter suppression: targeted, false claims about conditions at polling places that are designed to discourage or dissuade people from voting, or trick people into missing their opportunity to vote. This form of localized suppression includes things like false claims that a polling place is closed due to defective machines or that a voting location has been moved. (The purpose being to influence would-be voters not to go to their polling place to vote.) Facebook recognizes that if these statements are false they could unfairly interfere with the right to vote and would be in violation of its policy. The challenge historically, has been determining the veracity of localized content in a timely manner. Facebook has since begun exploring ways to distinguish between false claims about voting conditions that are suppressive, and accurate statements about problems at polling places that people (including voting rights advocates or election protection monitors) should be aware of.

In June 2020, Facebook announced that it has committed to a process for evaluating the accuracy of statements about polling conditions and removing those statements that are confirmed to be false and violating its policies. Specifically, Facebook announced that it will continue working with state election authorities, including in the 72 hours prior to Election Day, when voter interference is most likely and also most harmful, to confirm or refute the accuracy of the statements, and remove them when they are confirmed false. Facebook's expert voting rights consultants will also be available to share relevant regional and historical factors in real-time to help ensure the company approaches enforcement decisions with full awareness and context. The civil rights community (and the Auditors) view this as a positive step forward that could help reduce the amount of suppressive content on the platform. The Auditors and the civil rights community have some concern, however, that state election officials may not always provide accurate and timely information; indeed some state election officials have, at times, participated in suppression efforts, ignored them, or provided inaccurate information about voting conditions. Accordingly, the Auditors have recommended that Facebook supplement their process with other sources of reliable information on polling conditions including non-partisan election protection monitors and/or reliable news sources.

### 5. Improving Proactive Detection

Time is of the essence during elections, and civil rights groups have pushed for Facebook's prompt enforcement of its voting policies to minimize the impact of voter-suppressive content. Facebook has directed their enforcement teams to look for new tactics while also accounting for tactics seen on its platforms in past elections. In 2020, with the support of its voting rights consultants and the Auditors, Facebook's enforcement teams are also being familiarized with common off-platform tactics from past elections (including tactics used in paper flyers, emails, and robocalls) to target communities of color and race and language minorities.

For example in Philadelphia in 2008, flyers posted near Drexel University incorrectly warned that police officers would be at polling places looking for individuals with outstanding arrest warrants or parking tickets. If a similar hoax were to be disseminated on Facebook or Instagram in 2020, this would be a direct violation of Facebook's Voter Interference Policy, prohibiting "content stating that voting participation may result in law enforcement consequences (*e.g.*, arrest, deportation, imprisonment)." The Auditors believe expanding its proactive detection to account for a broader set of historical examples should allow Facebook to more rapidly identify more content that violates its existing policies, and better protect communities targeted by voter suppression on its platforms.

### 6. User Reporting & Reporter Appeals

**(i) User Reporting.** Facebook gives users the option to report content they think goes against the Community Standards. In 2018, Facebook added a new option for users to specifically report "incorrect voting information" they found on Facebook during the US midterm elections. As Facebook's Voter Interference Policy has evolved to include additional prohibitions beyond incorrect voting information, the Auditors advocated for Facebook to update this reporting option to better reflect the content prohibited under Facebook's Voter Interference Policy. Facebook has accepted the Auditors' recommendation and as of June 2020, the reporting option for users now reads "voter interference," which better tracks the scope of Facebook's policy.

While the Auditors are pleased that Facebook has updated its reporting options to better capture the scope of content prohibited under Facebook's Voter Interference Policy, the Auditors are concerned that this form of reporting is extremely limited. Currently, content reported by users as voter interference is only evaluated and monitored for aggregate trends. If user feedback indicates to Facebook that the same content or meme is being posted by multiple users and is receiving a high number of user reports, only then will Facebook have the content reviewed by policy and operational teams. This means that most posts reported as "voter interference" are *not* sent to human content reviewers to make a determination if posts should stay up or be taken down.

Facebook justifies this decision by citing its findings during the 2018 midterms, that an extremely low number of reports reviewed by its human reviewers were found to be violating its voter interference policy. In 2018, Facebook observed the vast majority of content reported as "incorrect voting information" were not posts that violated Facebook's voting policies, but instead were posts by people expressing different political opinions. Facebook reports that during the 2018 midterms, over 90% of the content Facebook removed as violating its voter suppression policy (as it existed at the time) was detected proactively by its technology before a user reported it. Ahead of the 2020 election, Facebook states that it was concerned that sending all reported voter interference content for human review could unintentionally slow its review process and reduce its ability to remove suppressive content by diverting reviewers from assessing and removing content more likely to be violating (*i.e.*, content proactively detected by Facebook or content flagged to Facebook by voting rights groups) to reviewing user reports that Facebook says have in the past often been non-violating and unactionable.

As stated previously in the New Partnerships section of this chapter (Section A.2.ii), Facebook indicated it has expanded its program partnering with voting rights organizations, which provides a dedicated channel for partner organizations to flag potentially violating voter interference content. Facebook recognizes these organizations possess unique expertise and in some instances may surface complex or nuanced content missed by its proactive detection. Facebook states that reports generated from this program in combination with regular inputs from its voting rights consultants will allow it to identify new trends and developments in how suppressive content appears on its platform, and continuously improve its overall detection and enforcement strategy.

Even if Facebook's proactive detection has improved in some ways, the Auditors remain concerned that Facebook's technology may not effectively anticipate and identify all forms of voter suppression that would be in violation of its policies, especially forms that are new, unique, or do not follow the same patterns as 2018. And, statistics on what percentage of content Facebook removed was initially flagged by proactive detection technology, of course, do not indicate whether or how much content that actually violated Facebook's policy was not detected and therefore allowed to stay up. Further, because Facebook's Voter Interference Policy has expanded since 2018, the Auditors are concerned that some forms of content prohibited under the current policy may be more nuanced and context-specific, making it more difficult to accurately detect with proactive detection technology. Because online voter suppression and misinformation pose such a grave risk to elections, the Auditors believe that it is insufficient and highly problematic to not send user reports of "voter interference" content to human reviewers. The Auditors believe that not routing user reports to content reviewers likely creates a critical gap in reporting for Facebook, a gap that is unreasonable for Facebook to expect can be filled by reports from partner organizations (with other obligations and already limited resources), even if external partners are experts in voter suppression.

**(ii) Reporter Appeals.** Facebook's decision not to send user-reported voter interference content to human reviewers has downstream effects on the ability of users to appeal reported content that is not taken down. In order for content to be eligible for appeal it must first be reviewed by Facebook and given a formal determination as to whether or not it violates Community Standards. As stated above, posts reported as potentially violating Facebook's Voter Interference Policy are treated as "user feedback" and are not formally assessed for violation (unless the post is also detected as potentially violating by Facebook). Given the significant harm that can be caused by voter interference content — including suppression and interference with users' ability to exercise their right to vote — the Auditors believe it is critical that there be a way to report and subsequently appeal potential missed violations to further ensure that violating suppressive content gets taken down. Further, content decisions that are unappealable cannot be appealed to the Oversight Board (for more details on the *Oversight Board*, see the *Content Moderation & Enforcement chapter*) by users once it is operational. This makes it impossible for election or census-related content to be reviewed by the Oversight Board, thereby excluding a critically important category of content — one that can impact the very operation of our democratic processes. The Auditors believe such exclusion is deeply problematic and must be changed.

### 7. Increased Capacity to Combat Coordinated Inauthentic Behavior

The last report included an update on Facebook's efforts to combat "information operations" or coordinated inauthentic behavior, which are coordinated, deceptive efforts to manipulate or disrupt public debate, including surrounding elections. The danger of such coordinated deceptive activity was illustrated in powerful detail in 2016, when foreign actors engaged in coordinated, deceptive campaigns to influence the US election, including targeting communities of color with demobilizing content. Since 2016, Facebook has built out a team of over 200 people globally — including experts in cybersecurity, disinformation research, digital forensics, law enforcement, national security and investigative journalism — that is focused on combating these operations.

Since the last Audit Report, Facebook states that it has continued to devote energy and resources to combating these threats and has built out its strategies for detecting coordinated inauthentic behavior. Facebook has adopted a three-pronged approach focused on detecting and removing: (1) violating content; (2) known bad actors; and (3) coordinated deceptive behavior. In other words:

- Facebook states that it removes suppressive content that violates Facebook policy regardless of who posts it;
- Facebook attempts to identify and remove representatives of groups that have been banned from the platform (like the IRA) regardless of what they post; and
- Facebook attempts to detect and dismantle coordinated efforts to deceive through fake accounts, impersonation, or use of bots or other computer-controlled accounts.

The purpose of these three strategies is to have the flexibility to catch these information operations, understanding that tactics are likely to evolve and change as bad actors attempt to evade detection.

Using this strategy, Facebook recently took down a network of accounts based in Ghana and Nigeria that was operating on behalf of individuals in Russia. The network used fake accounts and coordinated activity to operate Pages ostensibly run by nonprofits and to post in Groups. In 2016, Russian actors used fake accounts to build audiences with non-political content targeting issues relevant to specific communities, and then pivoted to more explicitly political or demobilizing messages. Here, this network of accounts was identified and removed when they appeared to be attempting to build their audiences — posting on topics such as black history, celebrity gossip, black fashion, and LGBTQ issues — before the accounts could pivot to possibly demobilizing messages. Facebook reported that it removed 49 Facebook accounts, 69 Pages, and 85 Instagram accounts as part of this enforcement action.

Facebook's systems are also used to detect coordinated deceptive behavior — no matter where it is coming from, including the United States. Facebook recently reported taking down two domestic networks engaged in coordinated inauthentic behavior, resulting in the removal of 35 Facebook accounts, 24 pages, and 7 Groups. These networks posted on topics relating to US politics including the presidential election and specific candidates, as well as COVID-19 and hate speech and/or conspiracy theories targeting Asian Americans.

While it is concerning that information operations continue to be a threat that is often targeted at particular communities, the Auditors are encouraged by Facebook's response and reported investment of time and resources into increasing their detection capabilities.

### 8. New Voting/Census Landing Page

Facebook has a number of different policies that touch on voting, elections, or census-related content, but not all of the policies live within the same section of the Community Standards. As a result, it can be difficult for civil rights groups (or users generally) who are trying to understand what voting/census-related content is permitted (and what is not) to easily review the relevant policies and understand where the lines are.

Both the Auditors and civil rights groups urged Facebook to provide more clarity and make voting/census related policies more accessible. Facebook agreed. Facebook is developing a landing page where all the different policy, product, and operational changes the company has implemented to prevent voter and census suppression are detailed in one place — additionally this page would include clarifications of how the policies fit together and answers to frequently asked questions.

## D. Political Ads Transparency Updates for Ads About Social Issues, Elections or Politics

Since the last report, Facebook has adopted several new transparency measures and made additional improvements to its public Ad Library. These updates are described in more detail below.

### 1. Policy Update

In 2019, Facebook updated its Policy on Social Issues, Elections or Politics to require authorizations in order to run ads related to the census. These ads are included in the Ad Library for transparency.

### 2. Labeling for Shared Ads

Facebook requires ads about social issues, elections or politics to indicate the name of the person or entity responsible for the ad through a disclaimer displayed when the ad is shown. However, when a user subsequently shared or forwarded an ad, neither the "Sponsored" designation nor the disclaimer designation used to follow the ad once it was shared, leaving viewers of the shared ad without notice that the content was originally an ad, or indication of the entity responsible for the ad. Civil rights advocates and others had expressed concern that this loophole could undermine the purpose of the labeling by allowing circumvention of transparency features, and leaving users vulnerable to manipulation. Facebook has now closed this loophole. Ads that are shared will retain their transparency labeling.

### 3. More Accessible Ad Information and Options to See Fewer Political, Electoral, and Social Issue Ads

Facebook has also developed new transparency features for these ads that compile and display relevant information and options all in one place. Since 2018, users have been able to click on political, electoral, or social issue ads to access information about the ad's reach, who was shown the ad, and the entity responsible for the ad, but now users can additionally see information about why they received the ad — that is, which of the ad targeting categories selected by the advertiser the user fell into.

In addition, the same pop-up that appears by clicking on an ad now gives users more control over the ads they see. Users now have the opportunity to opt into seeing fewer ads about social issues, elections or politics. Alternatively, users can block future ads from just the specific advertiser responsible for a given ad or adjust how they are reached through customer lists (*e.g.*, disallow advertisers from showing ads to them based on this type of audience list or make themselves eligible to see ads if an advertiser used a list to exclude them).

### 4. Ad Library Updates

Since 2018, Facebook has maintained a library of ads about social issues, elections or politics that ran on the platform. These ads are either classified as being about social issues, elections or politics or the advertisers self-declare that the ads require a "Paid for by" disclaimer. The last audit report announced updates and enhancements Facebook had made to the Ad Library to increase transparency and provide more information about who is behind each ad, the advertiser's prior spending, and basic information about the ad's audience. However, the civil rights community has continued to urge Facebook to both improve the Ad Library's search functionality and provide additional transparency (specifically information regarding targeting of political ads) so that the Ad Library could be a more effective tool for identifying patterns of suppression and misinformation.

Since the last report, Facebook has made a number of additional updates to the Ad Library in an effort to increase transparency and provide useful data to researchers, advocates, and the public generally. These improvements include:

- Potential Reach & Micro-Targeting: The Ad Library now permits users to search and filter ads based on the estimated audience size, which allows researchers to identify and study ads intended for smaller, more narrowly defined audiences. This new search function should make it easier to uncover efforts to "micro-target" smaller, specifically identified communities with suppressive or false information.

- Sorting Ad Search Results: When users run searches on the Ad Library, they can now sort their results so that the ads with the most impressions (*i.e.*, the number of times an ad was seen on a screen) appear first, allowing researchers to focus their attention on the ads that have had the most potential impact. Or, if recency is most important, ad search results can instead be sorted to appear in order of when the ad ran.

- Searching by State/Region: While the Ad Library has contained information about the state(s) in which a given ad ran since 2018, that information used to only be available by clicking on an individual ad. Facebook has updated its search functionality so that users interested in the ads that have run in a specific state or group of states can limit their searches to just those areas.

- Grouping Duplicate Ads: Advertisers often run the same ad multiple times targeted at different audiences or issued on different dates. When reviewing search results, users searching the Ad Library previously had to wade through all the duplicate ads — making it tedious and difficult to sort through search results. Facebook has now added a feature that groups duplicate ads by the same advertiser together so that it is easier to distinguish duplicates from distinct ads and duplicate versions of the same ad can be reviewed all at once.

- Ads Library Report Updates: In addition to these updates to the Ad Library, Facebook also updated its Ads Library Report to provide additional categories of information, including aggregate trend information showing the amount presidential candidates have spent on Facebook ads over time (either in aggregate or breaking down spend by date) and as well as searchable spend information for other (non-presidential) candidates.

By improving searching options and identifying ads that are targeted to smaller audiences, these updates to some extent advance the civil rights community's goals of making the Ad Library a useful tool for uncovering and analyzing voter suppression and misinformation targeted at specific communities. Facebook has asserted that privacy risks could be potentially created by sharing additional information about the audiences targeted for political ads, such as ZIP codes, or more granular location information of those receiving the ads. Due to this limiting factor, the current Ad Library updates do not fully respond to the civil rights community's assertions that such information is critical to identifying and addressing patterns of online voter suppression. Civil rights groups have provided Facebook with specific suggestions about ways to provide transparency without sacrificing privacy, ands continue to recommend that Facebook explore and commit to privacy-protective ways to provide more visibility into the targeting criteria used by advertisers so that the Ad Library can be a more effective tool for shedding light on election manipulation, suppression, and discrimination.

## E. Additional Auditor Concerns

### 1. Recent Troubling Enforcement Decisions

The Auditors are deeply concerned that Facebook's recent decisions on posts by President Trump indicate a tremendous setback for all of the policies that attempt to ban voter suppression on Facebook. From the Auditors' perspective, allowing the Trump posts to remain establishes a terrible precedent that may lead other politicians and non-politicians to spread false information about legal voting methods, which would effectively allow the platform to be weaponized to suppress voting. Mark Zuckerberg asserted in his 2019 speech at Georgetown University that "voting is voice" and is a crucial form of free expression. If that is the case, then the Auditors cannot understand why Facebook has allowed misrepresentations of methods of voting that undermine Facebook's protection and promotion of this crucial form of free expression.

In May 2020, President Trump made a series of posts in which he labeled official, state-issued ballots or ballot applications "illegal" and gave false information about how to obtain a ballot. Specifically, his posts included the following statements:

- "State of Nevada "thinks" that they can send out illegal vote by mail ballots, creating a great Voter Fraud scenario for the State and the U.S. They can't! If they do, "I think" I can hold up funds to the State. Sorry, but you must not cheat in elections"
- "Michigan sends absentee ballots to 7.7 million people ahead of Primaries and the General Election. This was done illegally and without authorization by a rogue Secretary of State. . ." (Note: reference to absentee "ballots" was subsequently changed to "ballot applications")
- "There is NO WAY (ZERO!) that Mail-In Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, ballots will be forged & even illegally printed out & fraudulently signed. The Governor

> of California is sending Ballots to millions of people, anyone living in the state, no matter who they are or how they got there, will get one. . ."

On its face, Facebook's voter interference policy prohibits false misrepresentations regarding the "methods for voting or voter registration" and "what information and/or materials must be provided in order to vote." The ballots and ballot applications issued in Nevada and Michigan were officially issued and are current, lawful forms of voter registration and participation in those states. In California, ballots are *not* being issued to "anyone living in the state, no matter who they are." In fact, in order to obtain a mail-in ballot in California one has to register to vote.

Facebook decided that none of the posts violated its policies. Facebook read the Michigan and Nevada posts to be accusations by President Trump that state officials had acted illegally, and that content challenging the legality of officials is allowed under Facebook's policy. Facebook deemed the California post to be non-violating of its provision for "misrepresentation of methods for voter registration." Facebook cited that people often use shorthand to describe registered voters (*e.g.,* "Anyone who hasn't cast their ballot yet, needs to vote today."). It wasn't clear to Facebook that the post — which said "anyone living in the state, no matter who they are" would get a ballot when, in fact, only those who registered would get one — was purposefully and explicitly stating "you don't have to register to get a ballot," and therefore was determined to be non-violating.

The Auditors vehemently expressed their views that these posts were prohibited under Facebook's policy (a position also expressed by Facebook's expert voting consultant), but the Auditors were not afforded an opportunity to speak directly to decision-makers until the decisions were already made.

To the civil rights community, there was no question that these posts fell squarely within the prohibitions of Facebook's voter interference policy. Facebook's constrained reading of its policies was both astounding and deeply troubling for the precedents it seemed to set. The civil rights community identified the posts as false for labeling official ballots and voting methods illegal. They explained that for an authoritative figure like a sitting President to label a ballot issued by a state "illegal" amounted to suppression on a massive scale, as it would reasonably cause recipients of such official ballots to hesitate to use them. Persons seeing the President's posts would be encouraged to question whether they would be doing something illegal or fraudulent by using the state's ballots to exercise their right to vote.

Civil rights leaders viewed the decision as opening the door to all manners of suppressive assertions that existing voting methods or ballots — the very means through which one votes — are impermissible or unlawful, sowing suppression and confusion among voters. They were alarmed that Facebook had failed to draw any line or distinction between expressing *opinions* about what voting rules or methods states *should* (or should not) adopt, and making false *factual* assertions that officially issued ballots are fraudulent, illegal, or not issued through official channels. Civil rights leaders expressed concern that the decision sent Facebook hurtling down a slippery slope, whereby the facts of how to vote in a given state or what ballots will be accepted in given jurisdiction can be freely misrepresented and obscured by being labeled unlawful or fraudulent.

Similarly, the civil rights community viewed the California post as a straightforward misrepresentation of how one gets a ballot — a misrepresentation that if relied upon would trick a user into missing his or her opportunity go obtain a ballot (by failing to register for one). That is the very kind of misrepresentation that Facebook’s policy was supposed to prohibit. As elections approach and updates are made to voting and voter registration methods due to COVID-19, both the civil rights groups and the Auditors worry that Facebook’s narrow policy interpretation will open the floodgates to suppression and false statements tricking people into missing deadlines or other prerequisites to register or vote.

In response to the press coverage around these decisions, Mark Zuckerberg has reasserted publicly that platforms should not be “arbiters of truth.” To civil rights groups, those comments suggested the renunciation of Facebook’s Voter Interference Policy; Facebook seemed to be celebrating its refusal to be the “arbiter of truth” on factual assertions regarding what methods of voting are permitted in a state or how one obtains a ballot—despite having a policy that prohibits factual misrepresentations of those very facts.

Two weeks after these decisions, and following continuing criticism from members of Congress, employees, and other groups, Mark Zuckerberg announced that the company would agree to review the company’s policies around voter suppression “to make sure [Facebook is] taking into account the realities of voting in the midst of a pandemic.” Zuckerberg warned, however, that while the company is committing to review its voter suppression policies, that review is not guaranteed to result in changes. Facebook also announced that it would be creating a voting hub (modeled after the COVID-19 hub it created) that would provide authoritative and accurate voting information, as well as tools for registering to vote and encouraging others to do the same.

The Auditors strongly encourage Facebook to expeditiously revise or reinterpret its policies to ensure that they prohibit content that labels official voting methods or ballots as illegal, fraudulent, or issued through unofficial channels, and that Facebook prohibit content that misrepresents the steps or requirements for obtaining or submitting a ballot.

### 2. Announcements Regarding Politicians’ Speech

In Fall 2019, Facebook made a series of announcements relating to speech by politicians. These included a September 2019 speech (and accompanying Newsroom Post) in which Nick Clegg, Vice-President for Global Affairs and Communications, stated that Facebook does not subject politicians’ speech to fact-checking, based on the company’s position that it should not “prevent a politician’s speech from reaching its audience and being subject to public debate and scrutiny.” Facebook asserts that the fact-checking program was never intended to police politicians’ speech. This public moment in September 2019 brought increased attention and scrutiny to Facebook’s standing guidance to its fact-checking partners that politicians’ direct statements were exempt from fact-checking. In that same speech, Clegg described Facebook’s newsworthiness policy, by which content that otherwise violates Facebook’s Community Standards is allowed to remain on the platform. Clegg clarified that in balancing the public’s interest in the speech against potential harm to determine whether to apply the newsworthiness exception, politicians’ speech is presumed to meet the public interest prong of Facebook’s newsworthy analysis. That is, politicians' speech will be allowed (and not get removed despite violating Facebook's content policies) unless the content could lead to real world violence or the harm otherwise outweighs the public's interest in hearing the speech.

These announcements were uniformly criticized in the civil rights community as being dangerously incongruent with the realities of voter suppression. In short, the civil rights community expressed deep concern because politicians have historically been some of the greatest perpetrators of voter suppression in this country. By continuing to exempt them from fact-checking at a time when politicians appear to be increasingly relying on using misinformation, and giving them a presumption of newsworthiness in favor of allowing their speech to remain up, the civil rights community felt like Facebook was inviting opportunities for increased voter suppression.

The Auditors shared the civil rights community's concerns and repeatedly (and vigorously) expressed those concerns directly to Facebook. Facebook has not made any clarifications on the scope of its definition for politicians nor has it made adjustments to its exemption of politicians from fact-checking. However, with respect to its newsworthiness policy, Facebook insists the most common application for its newsworthiness treatment is content that is violating but educational and important for the public's awareness (*e.g.*, images of children suffering from a chemical weapons attack in Syria). Facebook has since informed the Auditors that over the last year it has only applied "newsworthiness" to speech posted by politicians 15 times globally, with only 1 instance occurring in the United States.

Facebook has since clarified that voter interference and census interference as defined under the Coordinating Harm section of the Community Standards are exempt from the newsworthiness policy -- meaning they would not stay up as newsworthy even if expressed by a politician -- and newsworthiness does not apply to ads. After continued engagement by the Auditors and civil rights groups, Facebook recently extended the exemption from newsworthiness to Facebook's policies prohibiting threats of violence for voting or registering to vote and statements of intent or advocating for people to bring weapons to polling places. There is one voting-related policy where the exemption does not apply. Content could potentially stay up as "newsworthy" even if it violates Facebook policy prohibiting calls for people to be excluded from political participation based on their race, religion or other protected characteristics (*e.g.*, "don't vote for X Candidate because she's Black" or "keep Muslims out of Congress"). While the Auditors agree with Facebook's decision not to allow content violating these other voting policies to stay up as newsworthy, the Auditors urged Facebook to take the same position when politicians violate its policies by making calls for exclusion from political participation on the basis of protected characteristics, and are deeply concerned that Facebook has not done so. The Auditors believe that this exemption is highly problematic and demonstrates a failure to adequately protect democratic processes from racial appeals by politicians during elections.

The Auditors continue to have substantial concern about these policies and their potential to be exploited to target specific communities with false information, inaccurate content designed to perpetuate and promote discrimination and stereotypes, and/or for other targeted manipulation, intimidation, or suppression. While Facebook has made progress in other areas related to elections and the Census, to the Auditors, these political speech  exemptions constitute significant steps backward that undermine the company's progress and call into question the company's priorities.

Finally, in June 2020, Facebook announced that it would start being more transparent about when it deems content “newsworthy” and makes the decision to leave up otherwise violating content. Facebook reports that it will now be inserting a label on such content informing users that the content violates Community Standards but Facebook has left it up because it believes the content is newsworthy and that the public interest value of the content outweighs its risk of harm.

Setting aside the Auditors’ concerns about the newsworthiness policy itself (especially its potential application to voting-related content), the Auditors believe this move toward greater transparency is important because it enables Facebook to be held accountable for its application of the policy. By labeling content left up as newsworthy, users will be able to better understand how often Facebook is applying the policy and in what circumstances.

## Chapter Three: Content Moderation & Enforcement

Content moderation — what content Facebook allows and removes from the platform — continues to be an area of concern for the civil rights community. While Facebook's Community Standards prohibit hate speech, harassment, and attempts to incite violence through the platform, civil rights advocates contend that not only do Facebook's policies not go far enough in capturing hateful and harmful content, they also assert that Facebook unevenly enforces or fails to enforce its own policies against prohibited content. Thus harmful content is left on the platform for too long. These criticisms have come from a broad swath of the civil rights community, and are especially acute with respect to content targeting African Americans, Jews, and Muslims—communities which have increasingly been targeted for on- and off-platform hate and violence.

Given this concern, content moderation was a major focus of the 2019 Audit Report, which described developments in Facebook's approach to content moderation, specifically with respect to hate speech. The Auditors focused on Facebook's prohibition of explicit praise, support, or representation of white nationalism and white separatism. The Auditors also worked on a new events policy prohibiting calls to action to bring weapons to places of worship or to other locations with the intent to intimidate or harass. The prior report made recommendations for further improvements and commitments from Facebook to make specific changes.

This section provides an update on progress in the areas outlined in the prior report and identifies additional steps Facebook has taken to address content moderation concerns. It also offers the Auditors' observations and recommendations about where Facebook needs to focus further attention and make improvements, and where Facebook has made devastating errors.

### A. Update on Prior Commitments

As context, Facebook identifies hate speech on its platform in two ways: (1) user reporting and (2) proactive detection using technology. Both are important. As of March 2019, in the last audit report, Facebook reported that 65% of hate speech that it removed was detected proactively, without having to wait for a user to report it. With advances in technology, including in artificial intelligence, Facebook reports as of March 2020 that 89% of removals were identified by its technology before users had to report it. Facebook reports that it removes some posts automatically, but only when the content is either identical or near-identical to text or images previously removed by its content review team as violating Community Standards, or where content very closely matches common attacks that violated policies. Facebook states that automated removal has only recently become possible because its automated systems have been trained on hundreds of thousands of different examples of violating content and common attacks. Facebook reports that, in all other cases when its systems proactively detect potential hate speech, the content is still sent to its review teams to make a final determination. Facebook relies on human reviewers to assess context (*e.g.*, is the user using hate speech for purposes of condemning it) and also to assess usage nuances in ways that artificial intelligence cannot.

Facebook made a number of commitments in the 2019 Audit Report about steps it would take in the content moderation space. An update on those commitments and Facebook's follow-through is provided below.

### 1. Hate Speech Pilots

The June 2019 Audit Report described two ongoing pilot studies that Facebook was conducting in an effort to help reduce errors in enforcement of its hate speech policies: (1) hate speech reviewer specialization; and (2) information-first guided review.

With the hate speech reviewer specialization pilot, Facebook was examining whether allowing content reviewers to focus on only a few types of violations (rather than reviewing each piece of content against all of Facebook's Community Standards) would yield more accurate results, without negatively impacting reviewer well-being and resilience. Facebook completed its initial six-month long pilot, with pilot participants demonstrating increased accuracy in their decisions, and fewer false positives (erroneous decisions finding a violation when the content does not actually go against Community Standards). While more needs to be done to study the long-term impacts of reviewer specialization on the emotional well-being of moderators, Facebook reported that participants in the pilot indicated they preferred specialization to the regular approach, and that attrition among pilot participants was generally lower than average for content reviewers at the same review site.

Given those results, Facebook will explore a semi-specialization approach in the future where reviewers will specialize on a subset of related policy areas (*e.g.*, hate speech, bullying, and harassment) in order to significantly reduce pressure on reviewers to know and enforce on all policy areas. Facebook is choosing to pursue semi-specialization instead of specialization in any given Community Standard area to limit the amount of time that any reviewer spends on a single violation type to reduce risks of reviewer fatigue and over-exposure to the same kind of graphic or troubling content. At the same time, the company continues to build out its tools and resources supporting reviewer well-being and resiliency. Facebook reports that it is also working on establishing a set of well-being and resiliency metrics to better evaluate which efforts are most effective so that the company's future efforts can be adjusted to be made more effective, if necessary.

The other pilot, Facebook's information-first guided review pilot, was designed to evaluate whether modifying the tool content reviewers use to evaluate content would improve accuracy. The standard review tool requires reviewers to decide whether the content is violating first, and then note the basis for the violation. Under the pilot, the order was reversed: reviewers are asked a series of questions that help them more objectively arrive at a conclusion as to whether the content is violating.

Facebook completed a successful pilot of the information-first guided approach to content review, with positive results. Facebook states that content reviewers have found the approach more intuitive and easier to apply. Because switching to information-first guided review required creating new review tools, training, and workflows, Facebook felt the need to fully validate the approach before operationalizing it on a broader scale. Having now sufficiently tested the approach, Facebook plans to switch to information-first review for all content flagged as hate speech in North America, and then continue to expand to more countries and regions, and more categories of content. While COVID-19 has impacted the availability of content reviewers and capacity to train reviewers on the new approach, Facebook indicates it is working through those issues and looking to continue its progress toward more widespread adoption of the information-first approach.

### 2. Content Moderator Settlement

It is important to note that civil rights organizations have expressed concern about the psychological well-being of content reviewers, many of whom are contractors, who may be exposed to disturbing and offensive content. Facebook recently agreed to create a $52 million fund, accessible to a class of thousands of US workers who have asserted that they suffered psychological harm from reviewing graphic and objectionable content. The fund was created as part of the settlement of a class action lawsuit brought by US-based moderators in California, Arizona, Texas and Florida who worked for third party firms that provide services to Facebook.

In the settlement, Facebook also agreed to roll out changes to its content moderation tools designed to reduce the impact of viewing harmful images and videos. Specifically, Facebook will offer moderators customizable preferences such as muting audio by default and changing videos to black and white when evaluating content against Community Standards relating to graphic violence, murder, sexual abuse and exploitation, child sexual exploitation, and physical abuse.

Moderators who view graphic and objectionable content on a regular basis will also get access to weekly, one-on-one coaching sessions with a licensed mental health professional. Workers who request an expedited session will get access to a licensed mental health professional within the next working day, and vendor partners will also make monthly group coaching sessions available to moderators.

Other changes Facebook will require of those operating content review sites include:

- Screening applicants for resiliency as part of the recruiting and hiring process;
- Posting information about psychological support resources at each moderator’s workstation; and
- Informing moderators of Facebook’s whistleblower hotline, which may be used to report violations of workplace standards by their employers.

### 3. Changes to Community Standards

In the last report, the Auditors recommended a handful of specific changes to the Community Standards in an effort to improve Facebook’s enforcement consistency and ensure that the Community Standards prohibited key forms of hateful content.

The Auditors recommended that Facebook remove humor as an exception to its prohibition on hate speech because humor was not well-defined and was largely left to the eye of the beholder — increasing the risk that the exception was applied both inconsistently and far too frequently. Facebook followed through on that commitment. It has eliminated humor as an exception to its prohibition on hate speech, instead allowing only a narrower exception for content meeting the detailed definition of satire. Facebook defines satire as content that “includes the use of irony, exaggeration, mockery and/or absurdity with the intent to expose or critique people, behaviors, or opinions, particularly in the context of political, religious, or social issues. Its purpose is to draw attention to and voice criticism about wider societal issues or trends.”

The Auditors also recommended that Facebook broaden how it defined hate targeted at people based on their national origin to ensure that hate targeted at people from a region was prohibited (*e.g.*, people from Central America, the Middle East, or Southeast Asia) in addition to hate targeting people from specific countries. Facebook made that change and now uses a more expansive definition of national origin when applying its hate speech policies.

### 4. Updated Reviewer Guidance

In the last report, Facebook committed to providing more guidance to reviewers to improve accuracy when it comes to content condemning the use of slurs or hate speech. Recognizing that too many mistakes were being made removing content that was actually condemning hate speech, Facebook updated its reviewer guidelines to clarify the criteria for condemnation to make it clearer and more explicit that content denouncing or criticizing hate speech is permitted. Facebook reports that these changes have resulted in increased accuracy and fewer false positives where permissible content is mistakenly removed as violating.

## B. New Developments & Additional Recommendations

### 1. Hate Speech Enforcement Developments

In addition to completing the pilots discussed in the last Audit Report, Facebook made a number of other improvements designed to increase the accuracy of its hate speech enforcement. For example, Facebook made the following changes to its hate speech enforcement guidance and tools:

(i) **Improved Reviewer Tools.** Separate and apart from the commitments Facebook made as part of the content reviewer settlement, the company has further upgraded the tool reviewers use to evaluate content that has been reported or flagged as potentially violating. The review tool now highlights terms that may be slurs or references to proxies (stand-ins) for protected characteristics to more clearly bring them to reviewers’ attention. In addition, when a reviewer clicks on the highlighted term, the reviewer is provided additional context on the term, such as the definition, alternative meanings/caveats, term variations, and the targeted protected characteristic. The purpose of these changes is to help make potentially violating content more visible, and provide reviewers with more information and context to enable them to make more accurate determinations. Facebook plans to build tooling to assess whether and to what extent these changes improve reviewer accuracy.

(ii) **Self-Referential Use of Slurs.** While Facebook’s policies have always permitted the self-referential use of certain slurs to acknowledge when communities have reclaimed the use of the slur, Facebook reports that it recently refined its guidelines on self-referential use of slurs. Specifically, Facebook indicates that it provided content reviewers with policy clarifications on the slur uses that have historically been most confusing or difficult for content reviewers to accurately evaluate. Separately, Facebook reports that it clarified what criteria must be present for the use of a slur to be treated as a permissible “self-referential” use. These refinements were made to increase accuracy, especially with respect to users’ self-referential posts.

### 2. Oversight Board

The June 2019 Report described Facebook's commitment to establish an Oversight Board independent of Facebook that would have the capacity to review individual content decisions and make determinations as to whether the content should stay up or be removed—determinations which would be binding on Facebook (unless implementing the determination could violate the law). While the concept and creation of the Oversight Board was independent of the Audit, Facebook nevertheless requested that the Auditors provide input on the structure, governance, and composition of the board. Facebook states that its Oversight Board charter was the product of a robust global consultation process of workshops and roundtables in 88 different countries, a public proposal process, and consultations with over 2,200 stakeholders, including civil rights experts. The charter, which was published in September 2019, describes the Board's function, operation, and design. Facebook also commissioned and published a detailed human rights review of the Oversight Board in order to inform the Board's final charter, bylaws, and operations, and create a means for ensuring consistency with human rights-based approaches.

Once the charter was published, Facebook selected 4 co-chairs of the Board. Those co-chairs and Facebook then together selected the next 16 Board members. All 20 members were announced in May 2020. Looking ahead, in partnership with Facebook, the Board will select an additional 20 members. Once the initial Board reaches its 40 members, the Board's Membership Committee will have the exclusive responsibility of selecting members to fill any vacancies and to grow the Board beyond 40 members, if they so choose. All members, once selected by Facebook and the Board, are formally appointed by the Trustees who govern the Oversight Board Trust (the independent entity established to maintain procedural and administrative oversight over theBoard). Facebook compiled feedback and recommendations on Board member composition and selection process from external partners, consultants, and Facebook employees; and through a Recommendations Portal that the company initiated in September 2019 to allow individual members of the public to make recommendations. In the future, the Recommendations Portal will be the sole mechanism by which the Board will receive recommendations about potential new members.

The Auditors were repeatedly consulted during the process of building the initial slate of Board members and strongly advocated for the Board's membership to be diverse, representative, and inclusive of those with expertise in civil rights. While the Auditors did not have input into all Board member selections or veto power over specific nominees, the inclusion of diverse views and experiences, human rights advocates, and civil rights experts are positive developments that help lend the Board credibility in the Auditors' view.

### 3. Appeals & Penalties

#### (i) Appeals.

In 2018, Facebook launched a process allowing users to appeal content decisions. The process allows for appeals by both the person that posted content found to violate Community Standards and by users who report someone else's content as violating. Still, Facebook users have felt that the company's appeals system was opaque and ineffective at correcting errors made by content moderators. The Auditors have met with several users who explained that they felt that they landed in "Facebook jail" (temporarily suspended from posting content) in a manner that they thought was discriminatory and wrongly decided because of errors made by Facebook content moderators. After continued criticism, including by the civil rights community, Facebook committed in the 2019 Audit Report to improving the transparency and consistency of its appeals decision-making.

As a result, Facebook has made a number of changes to its appeals system and the notices provided to users explaining their appeal options. These changes include providing: (a) better notice to users when a content decision has been made; (b) clearer and more transparent explanations as to why the content was removed (or not removed); and (c) the opportunity for users to make more informed choices about whether they want to appeal the content decision. Specifically, Facebook has changed many of the interface and message screens that users see throughout the appeals process to provide more explanations, context, and information.

Facebook also reports that it studies the accuracy of content decisions and seeks to identify the underlying causes of reviewer errors—whether they be policy gaps, deficiencies in guidance or training, or something else. Facebook is exploring whether adjustments to the structure of its appeals process could improve accuracy while still being operational on the massive scale at which Facebook operates.

### (ii) Appeals Recommendations

- **Voter/Census Interference Policy appeals:** the details of this recommendation were presented in the User Reporting & Reporter Appeals section of the Elections & Census chapter.

- **Appeals data:** Facebook's Community Standards Enforcement Report details by policy area how much content was appealed, and how much content was restored after appeals. While this transparency is useful, Facebook should do more with its appeals data. The company should more systematically examine its appeals data by violation type and use these insights to internally assess where the appeals process is working well, where it may need additional resources, and where there may be gaps, ambiguity, or unanticipated consequences in policies or enforcement protocols. For example, if the data revealed that decisions on certain categories of hate speech were being overturned on appeal at a disproportionate rate, Facebook could use that information to help identify areas where reviewers need additional guidance or training.

- **Description of Community Standards:** As part of the enforcement and appeals user experience described above, Facebook has done more to inform users that content was taken down and to describe the Community Standard that was violated. While the increased transparency is important, the Auditors have found that Facebook's descriptions of the Community Standards are inconsistent. For example:

  - In some contexts, Facebook describes the hate speech policy by saying, "We have these standards to protect certain groups of people from being described as less than human."

  - In other circumstancs (such as describing violatons by groups), Facebook describes the hate speech policy as "content that directly attacks people based on their race, ethnicity, national origin, religious affiliation, sexual orientation, sex, gender or gender identity, or serious disabilities or diseases."

- o And in the contexts of violations by a page, Facebook describes hate speech as "verbal abuse directed at individuals."

In some cases Facebook reports that these differences are driven by Facebook's attempt to give users a description of the specific subsection of the policy that they violated to help improve user understanding and better explain the appeals process. In other instances, however, the differences are driven by inconsistent use of language across different products (*e.g.*, Groups, Pages). This is problematic because using inconsistent language to describe the relevant policy may create confusion for users trying to understand what Facebook's policies prohibit and whether/how their content may have violated those policies. Such confusion leaves Facebook susceptible to criticism around the consistency of its review.

The Auditors recommend that Facebook ensure its Community Standards are described accurately and consistently across different appeals contexts (*e.g.*, appeals regarding an individual post, a violation by a group, a violation by a page, *etc.*)

- **Frequently Reported Accounts:** The high-volume nature of Facebook's content moderation review process means that when an account attracts an unusually large number of reports, some of those reports are likely to result in at least some content being found to violate the Community Standards — even if the content is unobjectionable. Anti-racism activists and other users have reported being subjected to coordinated reporting attacks designed to exploit this potential for content reviewing errors. Those users have reported difficulty managing the large number of appeals, resulting in improper use restrictions and other penalties.

  Facebook's current appeal system does not address the particular vulnerabilities of users subjected to coordinated reporting campaigns (*e.g.*, reporting everything a user posts in the hope that some will be found violating and subject the user to penalties).

  The Auditors recommend that Facebook adopt mechanisms to ensure that accounts that receive a large number of reports, and that are frequently successful upon appeal, are not subjected to penalties as a result of inaccurate content moderation decisions and coordinated reporting campaigns.

**(iii) Penalties.**

Facebook's penalty system—the system for imposing consequences on users for repeatedly violating Facebook's Community Standards—has also been criticized for lacking transparency or notice before penalties are imposed, and leaving users in "Facebook jail" for extended periods seemingly out of nowhere. The company has faced criticism that penalties often seem disproportionate and to come without warning.

Since the last Audit Report, Facebook has made significant changes to its penalty system. To provide users greater context and ability to understand when a violation might lead to a penalty, Facebook has created an "account status" page on which users can view prior violations (including which Community Standard was violated) and an explanation of any restrictions imposed on their account as a result of those violations (including when those restrictions expire). Facebook similarly improved the messaging it sends to users to notify them that a penalty is being imposed — adding in details about the prior violations that led to the imposition of the penalty and including further explanation of the specific restrictions being imposed. Facebook has also begun informing users that further penalties will be applied in the future if they continue to violate its standards. Facebook is in the process of rolling out these new features, which the Auditors believe will be a helpful resource for users and will substantially increase transparency.

After the horrific attack in Christchurch, New Zealand in 2019, Facebook took steps to understand what more the company could do to limit its services from being used to cause harm or spread hate. Two months after the terrorist attack, the company imposed restrictions on the use of Facebook Live such that people who commit any of its most severe policy violations such as terrorism, suicide, or sexual exploitation, will not be permitted to use the Live feature for set periods of time. While these restrictions will not alleviate the fears about future live streaming of horrific events, they are an important step.

Taken together, the Auditors believe that these changes to Facebook's appeals and penalties processes are important improvements that will improve transparency, and reduce confusion and some of the resulting frustration. In the Auditors' view, however, there are additional improvements that Facebook should make.

### (iv) Penalties Recommendation.

**Transparency:** As noted above, Facebook has partially implemented increased transparency in the form of additional user messaging identifying the reasons behind a penalty at the time it is imposed, including the specific underlying content violations. However, in some settings users still receive earlier versions of the penalty messaging, which do not provide the user with context regarding the underlying content violations that led to the penalty.

The Auditors recommend that Facebook fully implement this additional user messaging across all products, interfaces, and types of violations.

## 4. Harassment

The civil rights community has expressed great concern that Facebook is too often used as a tool to orchestrate targeted harassment campaigns against users and activists. The Auditors have asked Facebook to do more to protect its users and prevent large numbers of users from flooding individual activists with harassing messages and comments. In the June 2019 report, the Auditors flagged a number of ways to better address and protect against coordinated harassment, including:

- Expressly prohibiting attempts to organize coordinated harassment campaigns;
- Creating features allowing for the bulk reporting of content as violating or harassing; and
- Improving detection and enforcement of coordinated harassment efforts.

This section describes the steps Facebook has taken to more effectively prohibit and combat harassment on the platform, and identifies areas for further improvement. Due to time constraints caused by the Auditors being pulled into address intervening events or provide input on time-sensitive challenges (as well as the COVID-19 crisis), the Auditors and Facebook were unable to conduct a detailed, comprehensive assessment of Facebook's harassment infrastructure as was done on hate speech in the 2019 report or as was done on Appeals and Penalties in this report. As a result, the Auditors cannot speak directly to the effectiveness of the changes Facebook has implemented over the last year, which are described here. However, the Auditors felt it was still important to describe these changes for purposes of transparency, and to flag the specific areas where the Auditors believe there is more work to be done.

On the policy side, Facebook has now adopted the Auditors' recommendation to ban content that explicitly calls for harassment on the platform, and will begin enforcement in July 2020. This policy update responds to concerns raised by the civil rights community regarding Facebook being too reactive and piecemeal in responding to organized harassment. In addition, Facebook has begun working with human rights activists outside the US to better understand their experiences and the impact of Facebook's policies from a human rights perspective, which could ultimately lead to recommendations for additional policy, product, and operational improvements to protect activists.

On the enforcement side, Facebook reports it has built new tools to detect harassing behavior proactively, including detection of language that is harassing, hateful, or sexual in nature. Content surfaced by the technology is sent to specialized operations teams that take a two-pronged approach, looking both at the content itself and the cluster of accounts targeting the user. Facebook reports using these tools to detect harassment against certain categories of users at a heightened risk of being attacked (*e.g.*, journalists), but is exploring how to scale application and enforcement more broadly to better mitigate the harm of organized harassment for all users, including activists.

When it comes to bulk reporting of harassment, however, Facebook has made less tangible progress. Last year the Auditors recommended 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 (which can be burdensome, emotionally draining, and time-consuming). In October 2018, Facebook launched a feature that allowed people to hide or delete multiple comments at once from the options menu of their post, but did not allow multiple comments to be reported as violating. The feature is no longer available due to negative feedback on the user experience. Facebook reports that it is exploring a reboot of this feature and/or other product interventions that could better address mass harassment — which may or may not be coordinated. A feature was recently launched on Instagram that allows users to select up to 25 comments and then delete comments or block the accounts posting them in bulk; the Auditors believe that Facebook should explore doing something similar because it is important that Facebook users are able to report comments in bulk so that harassers (including those not expressly coordinating harassment campaigns with others) face penalties for their behavior.

### 5. White Nationalism

In the last Audit Report, the Auditors restrained their praise for Facebook's then-new ban on white nationalism and white separatism because, in the Auditors' view, the policy is too narrow in that it only prohibits content expressly using the phrase(s) "white nationalism" or "white separatism," and does not prohibit content that explicitly espouses the very same ideology without using those exact phrases. At that time, the Auditors recommended that Facebook look to expand the policy to prohibit content which expressly praises, supports, or represents white nationalist or separatist ideology even if it does not explicitly use those terms. Facebook has not made that policy change.

Instead, Facebook reports that it is continuing to look for ways to improve its handling of white nationalist and white separatist content in other ways. According to the company, it has 350 people who work exclusively on combating dangerous individuals and organizations, including white nationalist and separatist groups and other organized hate groups. This multi-disciplinary team brings together subject matter experts from policy, operations, product, engineering, safety investigations, threat intelligence, law enforcement investigations, and legal.

Facebook further notes that the collective work of this cross-functional team has resulted in a ban on more than 250 white supremacist organizations from its platform, and that the company uses a combination of AI and human expertise to remove content praising or supporting these organizations. Through this process, Facebook states that it has learned behavioral patterns in organized hate and terrorist content that make them distinctive from one another, which may aid in their detection. For example, Facebook has observed that violations for organized hate are more likely to involve memes while terrorist propaganda is often dispersed from a central media arm of the organization and includes formalized branding. Facebook states that understanding these nuances may help the company continue to improve its detection of organized hate content. In its May 2020 Community Standards Enforcement Report, Facebook reported that in the first three months of 2020, it removed about 4.7 million pieces of content connected to organized hate — an increase of over 3 million pieces of content from the end of 2019. While this is an impressive figure, the Auditors are unable to assess its significance without greater context (*e.g.*, the amount of hate content that is on the platform but goes undetected, or whether hate is increasing on the platform overall, such that removing more does not necessarily signal better detection).

Facebook has also said that it is able to take more aggressive action against dangerous individuals and organizations by working with its Threat Intelligence and Safety Investigations team, who are responsible for combating coordinated inauthentic behavior. The team states that it uses signals to identify if a banned organization has a presence on the platform and then proactively investigates associated accounts, Pages and Groups — removing them all at once and taking steps to protect against recidivist behavior.

That being said, the civil rights community continues to express significant concern with Facebook's detection and removal of extremist and white nationalist content and its identification and removal of hate organizations. Civil rights advocates continue to take issue with Facebook's definition of a "dangerous organization," contending that the definition is too narrow and excludes hate figures and hate organizations designated by civil rights groups that track such content on social media. Furthermore, civil rights groups have challenged the accuracy and effectiveness of Facebook's enforcement of these policies; for example, a 2020 report published by the Tech Transparency Project (TTP) concluded that more than 100 groups identified by the Southern Poverty Law Center and/or Anti-Defamation League as white supremacist organizations had a presence on Facebook.

Because Facebook uses its own criteria for designating hate organizations, they are not in agreement with the hate designation of organizations that are identified by the TTP report. In some ways Facebook's designations are more expansive (*e.g.*, Facebook indicates it has designated 15 US-based white supremacist groups as hate organizations that are not so-designated by the Southern Poverty Law Center or Anti-Defamation League) and in some ways civil rights groups feel that Facebook's designations are under inclusive.

Of course, even if a group is not formally designated, it still must follow Facebook's content policies which can result in the removal of individual posts or the disabling of Pages if they violate Community Standards. In other words, an organization need not meet Facebook's definition of a hate organization for the organization's Page to be disabled; the Page can be disabled for containing hate symbols, hate content, or otherwise violating Community Standards. However, for the very same reasons that Facebook designates and removes whole organizations, civil rights groups contend that piecemeal removal of individuals posts or even Pages, while helpful, is insufficient for groups they think should be removed at the organizational level.

In addition, while Facebook announced in 2019 that searches for white supremacist terms would lead users to the page for Life After Hate (a group that works to rehabilitate extremists), the report also found that this redirection only happened a fraction of the time — even when searches contained the words "Klu Klux Klan." Facebook indicates that the redirection is controlled by the trigger words selected by Facebook in collaboration with Life After Hate and that "Klu Klux Klan" is on the list and that should have triggered redirection. The Auditors are heartened that Facebook has already begun an independent evaluation of its redirection program and the Auditors encourage Facebook to assess and expand capacity (including redirecting to additional organizations if needed) to better ensure users who search for extremist terms are more consistently redirected to rehabilitation resources.

The TTP report also noted how Facebook's "Related Pages" feature, which suggests other pages a person might be interested in, could push users who engage with white supremacist content toward further white supremacist content. While Facebook indicates that it already considers a page or group's history of Community Standards violations in determining whether that page or group is eligible to be recommended to users, the Auditors urge Facebook to further examine the impact of the feature and look into additional ways to ensure that Facebook is not pushing users toward extremist echo chambers.

At bottom, while the Auditors are encouraged by some of the steps Facebook is taking to detect and remove organized hate, including white nationalist and white separatist groups, the Auditors believe the company should be doing more. The company has not implemented the Auditors' specific recommendation that it work to prohibit expressly – even if not explicit – references to white nationalist or white separatist ideology. The Auditors continue to think this recommendation must be prioritized, even as the company expands its efforts to detect and remove white nationalist or separatist organizations or networks. In addition, the Auditors urge Facebook to take steps to ensure its efforts to remove hate organizations and redirect users away from (rather than toward) extremist organizations efforts are working as effectively as possible, and that Facebook's tools are not pushing people toward more hate or extremist content.

### 6. COVID-19 Updates

Facebook has taken a number of affirmative and proactive steps to identify and remove harmful content that is surfacing in response to the current COVID-19 pandemic. How COVID-19 is handled by Facebook is of deep concern to the civil rights community because of the disease's disproportionate impact on racial and ethnic groups, seniors, people who are incarcerated or in institutionalized settings, and the LGBTQ community among other groups. COVID-19 has also fueled an increase in hate crimes, scapegoating and bigotry toward Asians and people of Asian descent, Muslims and immigrants, to name a few. Lastly, civil rights groups are concerned that minority groups have been targeted to receive COVID-19 misinformation.

Since the World Health Organization (WHO) declared COVID-19 a public health emergency in January, Facebook has taken aggressive steps to remove misinformation that contributes to the risk of imminent physical harm. (While the company does not typically remove misinformation, its Community Standards do allow for removal of misinformation that contribute to the risk of imminent violence or physical harm.) Relying on guidance from external experts, such as the WHO and local health authorities to identify false claims, Facebook has removed false claims about: the existence or severity of COVID-19, how to prevent COVID-19, how COVID-19 is transmitted (such as false claims that some racial groups are immune to the virus), cures for COVID-19, and access to or the availability of essential services. The list of specific claims removed has evolved, with new claims being added as new guidance is provided by experts.

Facebook has also started showing messages in News Feed to people who have interacted with (*e.g.*, liked, reacted, commented on, shared) harmful misinformation about COVID-19 that was later removed as false. The company uses these messages to connect people to the WHO's COVID-19 mythbuster website that has authoritative information.

Facebook also updated its content reviewer guidance to make clear that claims that people of certain races or religions have the virus, created the virus, or are spreading the virus violate Facebook's hate speech policies. Facebook has similarly provided guidance that content attempting to identify individuals as having COVID-19 violates Facebook's harassment and bullying Community Standards.

Facebook's proactive moderation of content related to COVID-19 is, in the Auditors' view, commendable, but not without concerns. Ads that have patently false COVID-19 information have been generated and not captured by Facebook's algorithm. The strength of its strong policies is not only measured in words, but also how well those policies are enforced. Nonetheless, the Auditors strongly recommend that Facebook take lessons from its COVID-19 response (such as expanding the staff devoted to this effort, a commitment to public education and vigorously strengthening and enforcing its policies) and apply them to other areas, like voter suppression, to improve its content moderation and enforcement.

## 7. Additional Auditor Concerns and Recommendations

### (i) Recent Troubling Content Decisions.

The civil rights community found Facebook's recent enforcement decision finding content posted by President Trump to be outside the scope of its Violence and Incitement Policy dangerous and deeply troubling because it reflected a seeming impassivity toward racial violence in this country.

Facebook's Violence and Incitement Community Standard is intended to "remove language that incites or facilitates serious violence." The policy prohibits "threats that could lead to death" including "calls for high-severity violence," "statements of intent to commit violence," and "aspirational or conditional statements to commit high-severity violence." The policy also prohibits "statements of intent or advocacy or calls to action or aspirational or conditional statements to bring weapons to locations."

In the midst of nationwide protests regarding police violence against the Black community, President Trump posted statements on Facebook and Twitter that:

> "These THUGS are dishonoring the memory of George Floyd, and I won't let that happen. Just spoke to Governor Tim Walz and told him that the Military is with him all the way. Any difficulty and we will assume control but, when the looting starts, the shooting starts."

The phrase, "when the looting starts the shooting starts" is not new. A Florida police chief famously used the phrase in the 1960s when faced with civil rights unrest to explain that lethal force had been authorized against alleged looters.

In contrast to Twitter, which labeled the post as violating its policy against glorifying violence, Facebook deemed the post non-violating of its policies and left it up. Facebook's stated rationale was the post served as a warning about impending state action and its Violence and Incitement policy does not prohibit such content relating to "state action." Facebook asserted that the exception for state action had long predated the Trump posts. Mark Zuckerberg later elaborated in a meeting with employees that although the company understood the "when the looting starts, the shooting starts" phrase referred to excessive policing but that the company did not think it had a "history of being read as a dog whistle for vigilante supporters to take justice into their own hands."

The civil rights community and the Auditors were deeply troubled by Facebook's decision, believing that it ignores how such statements, especially when made by those in power and targeted toward an identifiable, minority community, condone vigilantism and legitimize violence against that community. Civil rights advocates likewise viewed the decision as ignoring the fact that the "state action" being discussed — shooting people for stealing or looting — would amount to unlawful, extrajudicial capital punishment. In encounters with criminal conduct, police are not authorized to randomly shoot people; they are trained to intercept and arrest, so that individuals can be prosecuted by a court of law to determine their guilt or innocence. Random shooting is not a legitimate state use of force. Facebook articulated that under its policy, threats of state use of force (even lethal force) against people alleged to have committed crimes are permitted. The idea that those in positions of authority could wield that power and use language widely interpreted by the public to be threatening violence against specific groups (thereby legitimizing targeted attacks against them) seemed plainly contrary to the letter and spirit of the Violence and Incitement Policy. Externally, that reading could not be squared with Mark Zuckerberg's prior assurances that it would take down statements that could lead to "real world violence" even if made by politicians.

The Auditors shared the civil rights community's concerns, and strongly urged Facebook to remove the post, but did not have the opportunity to speak directly to any decision-makers until after Facebook had already decided to leave it up.

As with the company's decisions regarding President Trump's recent voting-related posts, the external criticism of this decision was far from limited to the civil rights community. Some Facebook employees posted public messages disagreeing with the decision and staged a virtual walkout. Several former employees of the company published a joint letter criticizing the decision — warning that, "We know the speech of the powerful matters most of all. It establishes norms, creates a permission structure, and implicitly authorizes violence, all of which is made worse by algorithmic amplification." Members of the House Committee on Homeland Security sent a letter demanding an explanation for the decision, explaining "There is a difference between being a platform that facilitates public discourse and one that peddles incendiary, race-baiting innuendo guised as political speech for profit."

After the company publicly left up the looting and shooting post, more than five political and merchandise ads have run on Facebook sending the same dangerous message that "looters" and "ANTIFA terrorists" can or should be shot by armed citizens. These have ranged from ads by Congressional candidate Paul Broun referring to this AR-15 rifle as a "liberty machine" and urging its use against "looting hordes from Atlanta", to T-shirts depicting guns saying "loot this" or targets to be used as shooting practice for when "looters" come. To be clear, Facebook agreed these ads violated their policies (ads for T-shirts or targets are clearly not "warnings about state action"). Facebook ultimately removed the ads after they were brought to Facebook's attention, although only after the ads collectively received more than two hundred thousand impressions. The civil rights community expressed concern that the ads illustrated how Facebook's public decision to permit the President's looting and shooting post could have ripple effects that magnify the impact of the decision and further spread its violent messages on the platform. The fact that these violating ads calling for violence were not initially caught and taken down by Facebook's content reviewers is also concerning to the Auditors.

Facebook has since announced a willingness to revisit its Violence and Incitement Policy and the scope of its exception for threats of state action. As of this writing, it is unclear whether that revisiting will result in any policy or enforcement changes, and if so, what those changes will be. However, to many in the civil rights community the damage has already been done — the trust that the company will interpret and enforce its policies in ways that reflect a prioritization of civil rights has been broken.

### (ii) Polarization.

The civil rights groups and members of Congress also have questions about Facebook's potential role in pushing people toward extreme and divisive content. A number of them have flagged an article in the *Wall Street Journal* that asserts that Facebook leadership "shut down efforts to make the site less divisive" and "largely shelved" internal research on whether social media increases polarization. Additionally, the Chairman of the House Intelligence Committee said on June 18, 2020, "I'm concerned about whether social media platforms like YouTube, Facebook, Instagram and others, wittingly or otherwise, optimize for extreme content.

These technologies are designed to engage users and keep them coming back, which is pushing us further apart and isolating Americans into information silos." The Chairman further expressed concern about how Facebook's algorithm works and whether it prioritizes engagement and attention in a manner that rewards extreme and divisive content.

Facebook argues that the *Wall Street Journal* article used isolated incidents where leadership chose not to approve a possible intervention to make the argument that Facebook doesn't care about polarization in general. Facebook reports it has commissioned internal & external research, which have informed several measures the company has taken to fight polarization. Examples include:

- **Recalibrating News Feed.** In 2018, Facebook changed News Feed ranking to prioritize posts from friends and family over news content. Additionally, Facebook reports reducing clickbait headlines, reducing links to spam and misleading posts, and improving comment rankings to show people higher quality information.
- **Growth of Its Integrity Team.** Facebook has spent the last four years building a global integrity team that addresses safety and security issues, including polarization. This dedicated team was not in place when some of the internal research referenced was produced.
- **Restricting Recommendations.** If Pages and Groups repeatedly share content that violates Facebook's Community Standards, or is rated false by fact-checkers, Facebook reports that it reduces those Pages' distribution, and removes them from recommendations.

The Auditors do not believe that Facebook is sufficiently attuned to the depth of concern on the issue of polarization and the way that the algorithms used by Facebook inadvertently fuel extreme and polarizing content (even with the measures above). The Auditors believe that Facebook should do everything in its power to prevent its tools and algorithms from driving people toward self-reinforcing echo chambers of extremism, and that the company must recognize that failure to do so can have dangerous (and life-threatening) real-world consequences.

### (iii) Hate Speech Data & Analysis.

The Auditors recommend that Facebook compile data and further study how hate speech manifests on the platform against particular protected groups to enable it to devote additional resources to understanding the form and prevalence of different kinds of hate on the platform, its causes (*e.g.*, policy gaps, global enforcement trends or training issues, *etc*.), and to identify potential remedial steps the company could take.

Currently, when content reviewers remove content for expressing hate against a protected group or groups, Facebook does not capture data as to the protected group(s) against whom the hate speech was directed. Similarly, when users report content as violating hate speech policies, they do not have a way to note which protected class(es) are being attacked in the post. Without this information, Facebook lacks specific metrics for evaluating and understanding: (1) the volume of hate broken down by the group targeted, (2) whether there are categories of attacks on particular groups that are prevalent but not consistently removed, (3) whether there is a gap in policy guidance that has resulted in hate attacks against one religion, race, gender identity, falling through the cracks, based on the particular way those attacks manifested, *etc*.

Because the data would focus on the content of posts and the reasons that content violates Facebook's hate speech policies (rather than anything about the users reporting or posting it), the Auditors are confident that this kind of data collection need not involve collection of any data on users or otherwise implicate privacy concerns.

Facebook and the Auditors have repeatedly heard concerns from civil rights groups that particular forms of hate are prevalent on the platform but the absence of data for analysis and study seems to undercut efforts to document and define the problem, identify its source, and explore potential mitigation.

Take anti-Muslim hate speech, for example. For years the civil rights community has expressed increasing alarm at the level of anti-Muslim hate speech on (and off) the platform. While Christchurch was an inflection point for the Muslim community and its relationship to Facebook, the community's concerns with Facebook existed long before and extend beyond that tragedy. From the organization of events designed to intimidate members of the Muslim community at gathering places, to the prevalence of content demonizing Islam and Muslims, and the use of Facebook Live during the Christchurch massacre, civil rights advocates have expressed alarm that Muslims feel under siege on Facebook — and have criticized Facebook for not doing enough to address it. (Of course, this is not to say that Muslims are alone in experiencing persistent hate on the platform or the sense that they are under attack. Indeed, hate speech and efforts to incite violence targeting African Americans, Jews, Asians and the LGBTQ and LatinX communities, to name a few, have gained national attention in recent months. But, Facebook has not yet publicly studied or acknowledged the particular ways anti-Muslim bigotry manifests on its platform in the same manner it has discussed its root cause analysis of hate speech false positives removals of the posts of African American users and publicly launched pilots to test potential remedies).

Facebook's existing policy prohibits attacks against people based on their religion, including those disguised as attacks against religious concepts (*e.g.*, attacks against "Islam" which use pronouns like "they" or depict people). However, reports from civil rights groups and anecdotal examples suggest that these kinds of attacks persist on the platform and may seem to be more frequent than attacks mentioning Christianity, Judaism, or other religious concepts, making Facebook's distinction between attacks targeted at people versus concepts all the more blurry (and potentially problematic) when it comes to anti-Muslim sentiment.

Having data on the prevalence of anti-Muslim hate speech on the platform, what kinds of content is being flagged as anti-Muslim hate speech, and what percentage and types of content is being removed as anti-Muslim hate speech would be incredibly useful in defining the issue and identifying potential remedies. The Auditors recommend that Facebook (1) capture data on which protected characteristic is referenced by the perpetrator in the attacking post, and then (2) study the issue and evaluate potential solutions or ways to better distinguish between discussion of religious concepts and dehumanizing or hateful attacks masquerading as references to religious concepts or ideologies.

Facebook’s events policy provides another illustration of the need for focused study and analysis on particular manifestations of hate. Facebook policy prohibits both calls to bring weapons to houses of worship (including mosques) and calls to bring weapons to other religious gatherings or events to intimidate or harass people. Civil rights groups have expressed ongoing concern that Facebook’s enforcement of its events policy is too slow, often pointing to an August 2019 incident in which efforts to organize intimidation at the Islamic Society of North America’s annual convening in Houston, Texas took just over 24 hours to remove. Facebook agrees that 24 hours is too long and acknowledges that the Houston incident represents an enforcement misstep. Facebook should study the incident to pinpoint what went wrong and update protocols to ensure faster enforcement in the future. The Auditors believe having an effective expedited review process to remove such content quickly is critical given its potential for real-world harm, and that such post-incident analysis assessments are vital to that end. In the midst of nationwide protests, it is all the more important that Facebook get its events policy enforcement and expedited review process right — to ensure that people cannot use Facebook to organize calls to arms to harm or intimidate specific groups.

For that reason, the Auditors recommend that Facebook gather data on its enforcement of its events policies to identify how long it takes Facebook to remove violating content (and whether those response times vary based on the type of content or group targeted). Those kinds of metrics can be critical to identifying patterns, gaps, or areas for improvement.

Of course, the civil rights community’s concerns with hate on Facebook are not limited to anti-Muslim bigotry. And as we’ve seen with the COVID-19 pandemic and recent incidents of racism that have captured national (and international) attention, new manifestations and targets of hate speech can arise all the time, which, in the Auditors’ view, only reinforces the need to capture data so that new spikes and trends can be identified quickly and systematically.

At bottom, the Auditors recommend that Facebook invest in further study and analysis of hate on the platform and commit to taking steps to address trends, policy gaps, or enforcement issues it identifies. It is important that Facebook understand how different groups are targeted for hate, how well Facebook is alerting content reviewers to the specific ways that violating content manifests against certain groups, to more quickly identify and remove attempts to organize events designed to intimidate and harass targeted groups, and where Facebook could focus its improvement efforts. For many forms of hate, including anti-Muslim bigotry, documenting and publicly acknowledging the issue is an important first step to studying the issue and building solutions. For that reason, the Auditors not only recommend that Facebook capture, analyze, and act on this data as described above, but that it also include in its Community Standards Enforcement Report more detailed information about the type of hate speech being reported and removed from the platform, including information on the groups being targeted.

## Chapter Four: Diversity and Inclusion

As the nation becomes more attuned to systemic exclusion and inequities, companies should recognize diversity and inclusion as paramount and they should expect to be held accountable for their success (or failure) to embody these principles. In recent weeks, the tragedies and ensuing protests against police violence and systemic racism have led to a wave of corporate statements against the racism and injustice facing communities of color. For some, these expressions of solidarity ring hollow from companies whose workforce and leadership fail to reflect the diversity of this country or whose work environments feel far from welcoming or inclusive to underrepresented groups. The civil rights community hopes that these company commitments to doing "better" or "more" start with actual, concrete progress to further instill principles of diversity, equity, and **inclusion** in corporate America and Silicon Valley. Progress includes more diverse workforces at every level and inclusive environments with structures in place to promote equity and remove barriers. It includes a path to C-Suite or senior leadership posts for people of color (in roles that are not limited to diversity officer positions as is often the case in corporate America), and company-wide recognition that diversity and inclusion is a critical function of all senior leadership and managers (rather than the responsibility of those in underrepresented groups). This chapter provides a window into the status of diversity and inclusion at Facebook — its stated goals, policies, and programs — contextualized through the lens of concerns that have been raised in the civil rights community.

The civil rights community has long expressed concern regarding diversity and inclusion at Facebook—from staff and contractors (like those who are content reviewers), to senior management, and outside vendors or service providers that are used by the company to furnish everything from supplies to financial services. These concerns are multi-faceted. Civil rights groups have raised alarms about the relative dearth of people of color, older workers, people with disabilities, women, and other traditionally underrepresented minorities ("URMs") (including African Americans, Hispanic, Native Americans and Pacific Islanders) at Facebook—across multiple positions and levels, but particularly in technical roles and in leadership positions. Civil rights leaders have characterized the current numbers for Hispanic and African American staff as abysmal across every category (*e.g.*, technical roles, non-technical roles, management, *etc*.). Because of this lack of representation, civil rights groups have advocated for Facebook to do more to grow a strong and effective recruiting pipeline bringing underrepresented minorities into the company. Aside from recruiting and hiring, civil rights advocates also have challenged Facebook to ensure that those underrepresented minorities hired are retained, included, and promoted to positions of leadership—so that experiences of isolation or exclusion by URM employees do not lead to attrition reducing already low employment numbers. Concerns about the URM employee experience have been heightened in recent years following public memos and posts from current or former employees alleging experiences with bias, exclusion, and/or microaggressions.

The House Committee on Financial Services summarized many of these concerns in a public memo issued in advance of its 2019 hearing on Facebook in which it stated:

- "Facebook's 2019 diversity report highlights the company's slow progress with diversity metrics. From 2018 to 2019, Facebook reported less than a one percent increase in the total number of female employees. A majority

> of its employees are white (44%) and Asian (43%), with less than 13% of its total workforce representative of African Americans, Hispanics and other ethnic groups combined. Facebook's corporate board of directors and senior leadership are mostly comprised of white men, with the first appointment of an African American female in April 2019.[1] Facebook provides statistics on its supplier diversity, including spending $404.3 million in 2018 with diverse suppliers, an increase of more than $170 million from the previous year.[2] However, the report does not provide details on the total amount of spending with all suppliers nor has the company published specific data on its use of diverse-owned financial services firms, such as investment with diverse asset managers or deposits with minority-owned depository institutions."

In light of these concerns, the Audit Team has spent time drilling down on Facebook's diversity and inclusion strategy, programs, and practices. The Audit Team has met with policy and program leaders at the company, several members of the Diversity & Inclusion team, a small group of employees who lead Facebook Resource Groups (FBRGs), as well as the executives who sponsor those groups. This section reviews the Auditors observations, and acknowledges both the progress and the areas for improvement.

The Auditors have been pleased with recent announcements and changes by the company — they are both critical and signal a strong commitment to recognizing the importance of diversity and inclusion in all aspects of company operations. These include:

- Elevating the role of the Chief Diversity Officer to report directly to the COO and sit in on all management team meetings led by either the CEO or COO.
- A diverse supplier commitment of $1 billion in 2021 and every year thereafter. As part of that commitment, Facebook committed to spending at least $100 million annually with Black-owned suppliers.
- A commitment to have 50% of Facebook's workforce be from underrepresented communities by the end of 2023. (Facebook defines URM to include: women, people who are Black, Hispanic, Native American, or Pacific Islander, people with two or more ethnicities, people with disabilities, and veterans.) And, over the next five years, a commitment to have 30% more people of color, including 30% more Black people, in leadership positions.

Training 1 million members of the Black community, in addition to giving 100,000 scholarships to Black students working toward digital skills certifications. Facebook's goal in making this commitment is to ensure people have the opportunity to develop the skills necessary to succeed as we adjust to the COVID-19 world.

Increasing Facebook's previous $100 million global grant commitment by an additional $75 million available to Black-owned businesses in the US and to non-profits who support Black communities – as well as $25 million to Black creators to help amplify their stories on Facebook.

The resources that Facebook has committed over the last seven years to develop new Diversity & Inclusion projects, initiatives and programs (which are described in detail below) are noteworthy. In at least some of these areas, the company has made progress. Yet, as Facebook leadership has publicly acknowledged, there is more work to do.

[1] The Auditors note that this has since changed. There are now two African American women on Facebook's board of directors.

[2] The Auditors note that Facebook has updated these figures in its recently released annual supplier diversity report which is referenced below.

As a part of the Audit process, the Auditors had conversations with a small group of employees in winter 2019 who lead the company resource groups representing the URM populations. (Because the Auditors only had access to a small group of employees, comprehensive employee surveys or interviews were outside the scope of this Audit.) While employees did share positive sentiments on feeling empowered to build community, these conversations were primarily designed to elicit their general concerns and recommendations for approaches to improve the experience of URM populations at Facebook. Given the concerns expressed publicly by current and former employees, the Auditors wanted to include some themes of feedback here. The Auditors emphasize that the themes outlined here only reflect some of the views expressed by a small group of employees and are not to be construed as the views of all of the members of the Facebook Resource groups, or employees at large. Themes that emerged in the Auditors' conversations included:

- a concern about the lack of representation in senior management and the number of people of color (with the exception of Asians and Asian Americans) in technical roles;
- concerns about the performance evaluation process being consistently applied;
- a lack of recognition for the time URM employees spent on mentoring and recruiting other minorities to work at Facebook — this feedback was particularly pronounced with resource group leaders who are also managers;
- a greater sense of isolation because of their limited numbers compared to the overall workforce, especially in technical roles;
- a lack of awareness of all the internal programs available to report racial bias and/or discrimination;
- a desire to have more of a say in policies and products that affect their communities;
- a desire to see more data about rates of attrition.

To be sure, many of these diversity and inclusion issues are not unique to Facebook. Other tech companies and social media platforms have similarly low representation of URMs, and have similarly faced criticism for failing to bring employment opportunities to minority communities or foster inclusive environments where URMs stay and succeed. A recent report (Internet Association Inaugural Diversity & Inclusion Benchmark Report) highlights the lack of progress throughout the tech industry. Civil rights leaders continue to press the business case for inclusion and argue that diversity is a source of competitive advantage and an enabler of growth in a demographically changing society.

However, the fact that this is an industry-wide issue, does not absolve Facebook of its responsibility to do its part. Indeed, given the substantial role that Facebook plays in the tech industry and the outsized influence it has on the lives of millions of Americans and billions of users worldwide, it is particularly important for Facebook to maintain a diverse and inclusive workforce from top to bottom. The civil rights community and members of Congress are concerned that Facebook is not doing enough in that regard.

There is a strongly held belief by civil rights leaders that a diverse workforce is necessary and complementary to a robust civil rights infrastructure. That widely held belief was elevated by the House Committee on Energy and Commerce in its hearing on “Inclusion in Tech: How Diversity Benefits All Americans.” Without meaningful diversity or the right people in decision making, companies may not be able to account for blind spots and biases.

That said, having people of color in leadership roles is not the same as having people who have been deeply educated and experienced in understanding civil rights law and policy. People of color and civil rights expertise are not interchangeable. Treating them as such risks both reducing people of color to one-dimensional representatives of their race or national origin and unfairly saddling them with the responsibility, burden, and emotional labor of identifying civil rights concerns and advocating internally for them to be addressed. Facebook needs to continue to both drive meaningful progress on diversity and inclusion and build out its civil rights infrastructure, including bringing civil rights expertise in-house.

This chapter proceeds in five parts. First, it explains the strategies animating Facebook’s diversity and inclusion programs and the company’s D & I resources. Second, it describes the trajectory of Facebook’s employment figures and discusses Facebook’s hiring goals. Third, it summarizes relevant programs and initiatives intended to advance the company’s D & I goals. Fourth, it offers the Auditors’ observations on Facebook’s internal D & I efforts and suggested improvements. Fifth, it discusses the status of Facebook’s partner, vendor, and supplier diversity efforts and provides the Auditors’ observations on those efforts.

### 1. Facebook’s Diversity & Inclusion Strategy & Resources

Facebook’s diversity and inclusion program began in earnest in 2014, when it hired its first Global Chief Diversity Officer to define Facebook’s diversity and inclusion strategy and begin to build out a diversity and inclusion department at the company. Facebook states that its ultimate objective in pursuing diversity and inclusion efforts is to make better products and policies by leveraging employees’ different perspectives, skills and experience. With that goal in mind, diversity and inclusion strategies are aimed at:

- increasing the number of employees from underrepresented groups;
- building fair and inclusive systems for employee performance and development, including cultivating an environment that promotes employee retention of talent and leverages different perspectives, and implementing processes that support all people in their growth; and
- integrating D & I principles into company-wide systems.

The Auditors are not taking a position of support or opposition to these diversity strategies but are merely sharing what Facebook says it is doing. Facebook reports that it has created a number of programs and initiatives to generate progress on diversity and inclusion, which are outlined in Section 3 below.

When it comes to D & I at Facebook, the Auditors understand that the D & I team is strongly resourced (although the Auditors are not privy to exact budget numbers). It is also supported by approximately 40 members of the People Analytics team including data scientists, sociologists, social scientists, race and bias experts, and the People Growth team (whose expertise is in talent planning and career development). Furthermore, with its Global Chief Diversity Officer now sitting on Facebook’s executive management team and (as of June 2020) reporting directly to

Sheryl Sandberg, there is at least an increased opportunity to integrate diversity and inclusion considerations into decision-making.

### 2. Facebook's Workforce Figures and Hiring Goals

The figures Facebook published in its 2019 Diversity Report show Black and Hispanic employees make up 3.8% and 5.2% of employees across all positions, respectively, 1.5% and 3.5% of employees in technical roles, 8.2% and 8.8% of employees in business and sales roles, and 3.1% and 3.5% of employees in senior leadership roles. While Asian employees represent 43% of the workforce (and 52% of employees in technical roles), they represent only 24.9% of senior leadership roles.

Although Facebook has a long way to go, there are signs of progress. Facebook points out that there has been substantial change within individual subgroups and in specific roles. The company's latest published employment statistics show that since 2014 they have increased the number of Black women at Facebook by over 40x and the number of Black men by over 15x. This is spanning a period in which the overall company's growth was only 6.5x. This is good news even while the overall percentages remain small. On the non-technical side, Facebook has increased the percentage of Black people from 2% to 8%.

Facebook has also increased the representation of women from 31% of its population in 2014 to 37% in 2019 with the numbers in leadership over the same period moving from 23% to 33% women. In the technical realm, the company's most significant growth has been seen among women, who represented only 15% of people in technical roles in 2014 but increased to 23% by 2019.

In 2020, Facebook will report that 8% of its US workforce self-identified as LGBTQA+ (based on a 54% response rate), noting a 1% rise in representation from 2016, which is the first year that the company began collecting and publishing this data. Facebook's representation of veteran workers in the US has remained relatively steady at 2% between 2018 and 2020. As for people with disabilities, Facebook will report that 3.9% of its US workforce identified as being a person with a disability in 2020, which is the first year this data is being shared. (Facebook does not publicly report statistics on older workers. Figures for this category are absent from this report due to lack of access to data, not deprioritization by the Auditors.)

The Auditors' view into the 2020 numbers suggests that this trajectory of increasing representation generally continues in 2020.

Facebook also recently committed to a goal of diversifying its employee base such that by 2024 at least 50% of Facebook employees will be women, people who are Black, Hispanic, Native American, Pacific Islanders, people with two or more ethnicities, people with disabilities, and veterans (referred to as the "50 in 5" goal). (Currently 43% of Facebook's workforce fall into these categories.) In establishing this goal, Facebook aims to double the number of women it employs globally and the number Black and Hispanic employees working in the US. While the goal is ambitious, Facebook reports that it was set to signal commitment, help focus the company's efforts, and drive results. Facebook asserts that in order to set the company up for success, the company instituted the 50 in 5 goal only after building out its D & I, Human Resources, Learning & Development, Analytics and Recruiting teams and strategies, and taking steps to build out its internal infrastructure by, for example, starting to evaluate senior leaders on their effectiveness at meeting D&I goals. This goal (and the principle of representation it reflects) has been embraced by civil rights leaders.

On June 18 of this year, Facebook further enhanced its 50 in 5 goal by announcing that it would aim to increase the number of people of color in leadership positions over the next years by 30%, including increasing the representation of Black employees in such roles by 30%. The Auditors recognize that diversity in leadership is important and view these goals as important steps forward to be achieved.

The Auditors believe in public goal setting for the recruitment of URMs, and recognize that these aspirations are important signals of the desire for diversity. However, the Auditors are wary that it would send a problematic message if Facebook does not come close enough to meeting its goals. The Auditors would like to know more about how the commitment to these goals has changed behavior or prompted action, and how the company plans to ensure representation of each sub-group in these goals. The Auditors were unable to poll leaders on this topic, but would like to see continued public commitments to and discussion of these goals by the Facebook senior leadership team.

The Auditors recognize that workforce statistics are not a sufficient or meaningful metric for providing transparency into the current state of inclusion at Facebook, and a sense of whether and to what extent Facebook has created an inclusive environment. The absence of helpful measures of equity or inclusion at Facebook is not intended to suggest that those goals are subordinate or insignificant but merely reflect the Auditors’ lack of access to such data or resources.

### 3. Details on Facebook’s Diversity and Inclusion Programs & Systems

The D & I strategy the company has adopted (and refined) since 2014 has three main components which operate simultaneously and build off each other: (i) recruiting (ii) inclusion; and (iii) the integration of D & I principles into company-wide systems.

By design, not all of Facebook’s programs are housed within the diversity and inclusion or human resources departments; a number of them are in education and civic engagement partnerships, based on the company’s belief that that for D & I to become a core component of company operations it must be embedded into all systems rather than stand alone. Some of these programs are relatively longstanding (*e.g.*, five years old) and some have been rolled out within the last year. These programs, which are intended to address short, medium, and long-term goals, are described in more detail below. The Auditors recount these efforts not for the purpose of supporting (or critiquing) any particular initiative, but to provide transparency into what Facebook is doing.

In the Auditors’ view, these programs and initiatives demonstrate that Facebook is investing in D & I and taking concrete steps to help create a diverse and inclusive culture. At the same time, the Auditors maintain that there are additional steps that Facebook can and should take to ensure that the benefits of these programs are fully realized. The Auditors’ recommendations and observations about potential areas for improvement or growth are set out in Section 4.

### (i) Recruiting.

The goal of Facebook's recruiting policies and programs are to recruit and hire candidates from diverse backgrounds — understanding that Facebook cannot build a diverse culture without diverse representation.

Facebook has instituted a number of programs and commitments designed to increase diversity in hiring. For example, Facebook introduced the "Diverse Slate Approach" as a pilot in 2015, which sets the "expectation that candidates from under-represented backgrounds be considered when interviewing for an open position." Akin to the "Rooney Rule" in the National Football League, the idea is to promote diverse hiring by ensuring that a more diverse set of candidates are given careful consideration. As applied to Facebook, the company states that for every competitive hire (*e.g.*, not for an internal lateral transfer to an open position), hiring managers are expected to interview qualified candidates from groups currently underrepresented in the position. The purpose of the strategy is to focus recruiters' attention on diversifying the candidate pool and push hiring managers to ensure they have truly considered a range of qualified talent before making a hiring decision. Facebook asserts that it has seen increases in diversity with the application of the strategy (without causing significant hiring delays). Facebook has now adopted the Diverse Slate Approach globally and also applied it to open positions on its Board of Directors in 2018. Facebook does not, however, tie executive pay to achieving diversity metrics and that is something it may want to consider to accelerate its ability to meet targets.

In addition, as discussed above, Facebook has also set aggressive hiring goals of 50% representation in five years, prioritizing hiring at the leadership levels and in technical functions. (Although it remains to be seen whether Facebook will meet those goals.)

Part of diversifying hiring has also included efforts to look outside of Silicon Valley for qualified candidates. Facebook states that it is recruiting talent from more than 300 schools across the United States for entry level jobs (including from HSIs and HBCUs) and from thousands of companies globally across multiple industries for experienced hires.

In addition to hiring, Facebook has adopted a host of programs and initiatives designed to build out the pipeline of underrepresented minorities into tech jobs. The target audiences for these programs range from post-graduate level students to college students, high school students, and even elementary-school age children and their families or caregivers. These programs include, for example:

- **Engineer in Residence:** Facebook engineers are embedded on university campuses at institutions with high minority enrollment (including HBCUs and HSIs) to design and teach undergraduate computer science courses and extracurricular programs to provide underrepresented groups with access to innovative computer science curricula and programming.

- **Facebook University:** an 8-week summer training program where college freshmen intern at Facebook across roles in engineering, analytics, product design, operations, and sales and advertising, with the goal of building connections between students from underrepresented communities and Facebook.

- **Align Program:** Facebook is sponsoring Northeastern University's Align Program, which helps non-computer science graduates, especially those from traditionally underrepresented groups, change careers to transition to computer science.

- **Co-Teaching AI:** Facebook's Artificial Intelligence (AI) team has partnered with Georgia Tech to co-create and co-teach an AI course designed to help diversify exposure to the AI field.

- **Above & Beyond CS Program:** A 10-week program designed for college juniors and seniors from underrepresented groups to help prepare students in computer science fields for the technical interviews that are an integral part of the hiring process for these jobs.

- **CodeFWD:** Facebook provides a free online program to educators and non-profit organizations designed to allow them to introduce students in grades 4 through 8 to computer programming. After completing the program, the educators and organizations can apply to receive additional resources like programmable robots to provide further coding opportunities to their students.

- **TechPrep:** Facebook provides a free online resource hub (in English, Spanish, and Portuguese) to help students ages 8-25 and their parents or guardians learn what computer science is, what jobs are available to computer programmers, and how to get started learning to code.

**(ii) Inclusive Programming.**

The goal of Facebook's inclusion efforts is to ensure Facebook employees — especially members of under-represented groups — feel seen, heard, and valued. These initiatives range from community-building opportunities and resources to trainings and tools for managing or addressing bias and promoting inclusion.

Facebook's community building opportunities and resources include:

- **Facebook Resource Groups (FBRGs):** These are inclusive groups that anyone who works at Facebook can join, which are focused on underrepresented and marginalized communities, and provide professional development, community support, and opportunities to build connections with other group members and engage on important issues.

- **Community Summits:** Facebook also supports its underrepresented workforce through annual gatherings or community summits that bring together people who work at Facebook across the globe and provide a forum for various communities to gather, share and grow.

Facebook has also developed and deployed a number of trainings intended to advance its inclusion goals. These include its Managing Bias and Managing Inclusion trainings, which provide tools and practical skills designed to help limit the impact of biases (including unconscious ones) and promote inclusion within teams and in day-to-day interactions, and a "Be the Ally" training, which provides guidance to help employees support each other and take steps to counteract examples of exclusion or bias they observe. Additional guidance in this area is included in the onboarding training managers undergo as well as Facebook's Managing a Respectful Workplace training. Facebook also offers a "Design for Inclusion" training which Facebook describes as a multi-day immersive workshop for senior leaders in the company that focuses on exploring the root causes of inequities that influence decision-making, and works towards creating a more inclusive and innovative company culture. While these trainings have been available to all employees for years, Managing Bias, manager onboarding and Managing a Respectful Workplace are now mandatory.

Along with developing its suite of trainings, in 2019 Facebook created a new tool for anonymously reporting microaggressions as well as positive examples of allyship or supportive behaviors that have an impact on day-to-day life at Facebook. The tool, called the "Micro-Phone," provides employees (and contingent workers) an outlet for sharing these experiences, and gives Facebook insight into common themes and trends. Facebook states that it includes insights from the Micro-Phone in reports regularly provided to senior leadership (to flag issues and push for implementation of D & I action plans), and uses Micro-Phone lessons to inform trainings and help build D & I and HR strategies

**(iii) The Integration of Diversity and Inclusion Principles into Company-Wide Systems.** The third component of Facebook's diversity and inclusion strategy is focused on integrating a D & I lens into processes, policies and products. That is, building out internal systems to help promote consistent implementation of D & I policies and practices, and looking for ways to ensure Facebook considers and accounts for diverse experiences and perspectives in developing policies and products.

For example, Facebook has examined its performance review process to look for ways that bias or stereotyped assumptions could seep in, and is making changes to the process to try to counteract those risks. These changes include requiring mid-cycle performance conversations designed to provide more uniform opportunities for direct communication (rather than presumptions) and more consistent feedback. Similarly, Facebook has adopted scorecards to better hold department leaders accountable for implementing the company's diversity and inclusion policies; Facebook states that department leaders will be given clear criteria (*e.g.*, their team's consistent use of the Diverse Slate Approach, consistent and quality career conversations with direct reports, ensuring that their teams complete the Facebook's trainings on bias, inclusion, and allyship, *etc*.), and be assessed against that criteria.

In addition, Facebook is in the early stages of developing a plan to better integrate into its policy and product development process consideration of how different policies and products will impact, speak to, or work for people across a wide spectrum of experiences, identities, and backgrounds. To that end, Facebook has begun piloting this strategy by inserting the Chief Diversity Officer into product and policy discussions. To begin formalizing that integration, Facebook recently announced that it has moved the Chief Diversity Officer within Facebook's organizational structure so that the role now directly reports to COO Sheryl Sandberg. With this change, Facebook states that it intends to involve the Chief Diversity Officer in high-level decision-making affecting products, business, and policy on a more consistent basis. Facebook also recently hired a full-time employee to focus on this D & I integration work. Facebook indicates its next goal is to determine how to build up the concept into a systemic and scalable approach, as opposed to more ad-hoc injections of D & I team members into policy or product decision-making processes.

### 4. Auditors' Observations Regarding Facebook's Internal D & I Efforts

Overall, the constant change in diversity and inclusion at Facebook— driven by the development of new projects and initiatives and the expansion of existing programming — reflects ongoing innovation and interest in D & I. The Auditors further believe that Facebook's new focus on D & I integration and ensuring greater accountability in the application of D & I policies and strategies through things like leadership scorecards are steps in the right direction.

To identify issues and assess program effectiveness, Facebook reports that the company uses quantitative and qualitative assessments, feedback from surveys and regular focus groups with under-represented people, coupled with established third-party research. The Auditors urge Facebook to make at least some of this data and feedback public (in its annual Diversity Report) so that the civil rights community and the general public can better understand the effectiveness of the company's myriad programs and initiatives. However, because the Auditors are not privy to this data or feedback, the Auditors cannot speak to the effectiveness of any particular program or initiative. Further, while the Auditors did not have an opportunity to conduct surveys or interviews of employees, in their discussions with employees they observed a disconnect between the experiences described by a number of the employee resource group representatives and the diversity and inclusion policies, practices, and initiatives described by Facebook. The Auditors have made a number of recommendations based on conversations with ERG representations and company leadership.

(i) **Comprehensive study.** Anecdotal accounts the Auditors heard suggest that efforts to instill inclusive practices or ensure consistent application of diversity-enhancing policies may have not yet taken hold on a systemic level. These observations signal that a more comprehensive (both quantitative and qualitative) study of how consistently Facebook's diversity and inclusion-based policies or strategies are being applied internally would be valuable.

The Auditors believe that continuing to develop data and metrics for assessing the effectiveness of its inclusion, and D & I integration efforts is critical to evaluating and guiding Facebook's D & I strategy. While Facebook publishes its employment figures annually in its diversity report, those figures primarily speak to Facebook's hiring and recruiting efforts — they do not offer a clear illustration of whether/how Facebook's initiatives, policies, trainings, and tools designed to advance inclusion and D & I integration have impacted employee experiences or have translated to progress in cultivating a culture of inclusion. These additional metrics would provide critical insight in those areas. Further, the results could help Facebook identify where it may need to refocus attention and consider ways to revise, expand, improve, and/or redesign their existing programs

**(ii) Continued improvement on infrastructure.**

The Auditors encourage Facebook to continue to invest in building out systems and internal infrastructure to make sure diversity and inclusion strategies are prioritized, applied with consistency, embedded in everyday company practices, and ultimately create an inclusive culture.

For example, the Auditors believe that practices such as the consistent application of the Diverse Slate Approach and exhibiting inclusive behavior are metrics upon which all employees, managers, and executives (not just senior leaders) should be evaluated in performance reviews. (As of 2019, senior leaders started to be given goals against the Diverse Slate Approach and Inclusion metrics, which is progress, but the Auditors believe is not enough.). Given the company's ongoing exponential growth, and its diffuse and siloed organizational structure, and the other pressures that employees face to innovate and get products to market quickly, focusing on accountability, consistency, and D & I integration seems critical for diversity and inclusion practices to be effectively adopted at scale. It is important for managers and employees to be deeply familiar with tools and practices designed to impact the culture at Facebook and create a more inclusive environment.

(Given the COVID-19 pandemic and Facebook's recent announcement that remote work will continue indefinitely for many employees, Facebook should assess whether adjustments need to be made to inclusion and D & I integration strategies to account for the impact of prolonged remote work — especially on efforts to instill community, combat URM isolation, and ensure consistency in feedback, mentoring, and application of D & I strategies across the board.)

**(iii) Stronger Communication.**

Based on the Auditors' observations and conversations, one of the unfortunate side effects of this development and expansion is that programs can sometimes be siloed and diffuse, which can result in a lack of awareness of different initiatives, how they fit together, and what needs to be done to advance them. As an initial step, the Auditors believe that describing all of Facebook's diversity and inclusion programs and initiatives in a single user-friendly resource, and explaining how the programs all fit together, and the strategies behind them would help address information gaps and focus conversations. (This report does not substitute for such a resource because it is merely an outline of Facebook's efforts and is not exhaustive.)

Both in the civil rights community and inside Facebook, conversations about how to improve diversity and inclusion at the company can be more targeted if there is greater transparency and clarity about what Facebook is currently doing (and not doing) and what Facebook's policies are — as compared with employees' lived experiences.

### 5. Partner, Vendor, and Supplier Diversity

The civil rights community has criticized Facebook for not doing enough to ensure that the vendors and service providers it chooses to partner with reflect the diversity of our society. They contend that partnering with more diverse vendors, media companies, and law and financial management firms is also good business, as it promotes innovation and brings new audiences, perspectives, and ideas to the table.

Facebook launched its supplier diversity program in late 2016 with the goal of helping diverse suppliers do business with Facebook and with the people and communities that Facebook connects. Through the program, Facebook has sought to increase its use of vendors owned by racial and ethnic minorities, women, members of the LGBT community, veterans, and people with disabilities. In July 2020, Facebook reported spend of $515 million with certified diverse suppliers in 2019 — a 40% increase over 2018 ($365M) — bringing its cumulative spend to over $1.1 billion since the launch of these efforts.

In June 2020, Facebook set a new goal: to spend at least $1 billion with diverse suppliers starting in 2021 and continuing each year thereafter. As part of that goal, the company committed to spending at least $100 million per year with Black-owned suppliers.

Because vendor decisions are diffuse rather than centralized in a single team, changing the way Facebook makes vendor decisions required building a tool that would promote more diverse choices at scale. Facebook has now developed an internal vendor portal to facilitate selection of diverse-owned companies when Facebook teams are looking for vendors for everything from office supplies to coffee to cables for data centers.

With its rapid expansion (and the large-scale construction projects accompanying such expansion), Facebook is now turning its attention to diversifying its construction contracting for both primary contracts and subcontracts. Working in partnership with its Global Real Estate and Facilities team, Facebook states that it has established aggressive internal goals for increasing opportunities and awarding competitive contracts to diverse suppliers starting with general contractors and directly sourced professional services (*e.g.*, architects, interior design, fixtures, furnishing and equipment). In addition, Facebook indicates it will launch its Tier 2 (subcontractor) reporting program in 2020, which will require eligible Facebook contractors to report their direct subcontracting with diverse suppliers on a quarterly basis. This will include key categories of spend like construction, facilities operations, marketing and events, where the prime supplier relies heavily on subcontracted suppliers to deliver the scope of work for which Facebook engaged them. Facebook states that it will also strengthen its contract language and program to more affirmatively encourage and support prime suppliers in identifying and contracting with qualified diverse subcontractors.

Facebook has also made commitments to increase diversity and inclusion within consumer marketing. The consumer marketing team works with hundreds of creative supply chain vendors a year to create marketing and advertising campaigns for Facebook and its family of apps. The consumer marketing team has committed to increasing diversity and inclusion in the following areas within their supply chain: supplier diversity (owner/operator), on camera talent, key production crew roles including photographer, director, first assistant director editor, director of photography, visual effects artist, audio mixer and colorist. To implement this commitment Facebook has taken steps such as:

- Prioritizing diversity in selecting vendors to work on projects.
- Partnering with the non-profit Free the Work, pledging to always consider/bid at least one female director every time there is a commercial production over $500K.
- Creating an economic pipeline program for production assistants.
- Tracking the production commitments across our external agencies and internal teams on a quarterly basis to ensure accountability.

Facebook has also taken steps to require diversity when engaging other service providers, such as outside legal counsel. When Facebook hires outside law firms, it now requires that those firms staff its Facebook projects with teams that are at least one-third diverse (meaning racial or ethnic minorities, women, people with disabilities, or members of the LGBT community). Facebook’s outside counsel agreements also require that diverse team members be given meaningful roles and responsibilities, such as being the day-to-day contact with Facebook, leading presentations, or having a speaking role at court hearings.

In 2019, Facebook launched an annual survey of its top 40 law firms (by spend) it engages as outside counsel to evaluate the firms’ performance in meeting these diversity requirements. Facebook celebrated the firm with the highest score and is directing all firms, especially low-scoring firms to improve. (Facebook has indicated that penalties, including cancellation of outside counsel contracts, were not imposed but may be imposed in the future should firms persist in failing to meet expectations for diversity.) In addition to these diversity commitments, Facebook is starting to build partnerships with law firms to promote greater diversity in the legal profession through programs designed to provide greater opportunities for law students from diverse backgrounds.

In the Auditors’ opinion, Facebook has demonstrated less progress on the financial management side. Facebook has faced strong criticism from the civil rights community (and members of Congress) regarding the lack of diversity of its asset managers and financial services providers. During testimony before the House Financial Services Committee in 2019, Mark Zuckerberg was grilled about Facebook’s asset management and whether sufficient attention has been paid to the diversity of Facebook’s asset management firms. Of the 10 investment management firms Facebook works with, one is self-identified (but not certified) as female owned, and none are minority-owned.

Facebook states that its engagements with financial institutions center around capital markets activities (share repurchases) and investment management. The company notes that in 2020, it hired a diverse firm to execute share repurchases on their behalf. Facebook also engaged a diverse consulting firm to conduct a search for diverse investment managers capable of meeting the company’s needs. Facebook indicates that the results of this search are being used to develop an RFP, with the intent to hire qualified vendors.

### 6. Auditors’ Observations

Facebook has made important progress in some areas, especially its vendor diversity program. But, it can and should do more. Its efforts to expand construction-related contracting with diverse-owned companies is a step in the right direction. Given that millions of businesses use Facebook products and services, Facebook could also do more to enable diverse-owned companies to be identified and surfaced through Facebook’s products to provide more visibility for those seeking to partnership with diverse-owned companies. With respect to outside counsel engagements, including updating its contracts to require diverse representation and meaningful participation are positive, affirmative steps. The Auditors encourage Facebook to continue to explore ways to give those words meaning by ensuring that firms that fall short of these obligations are held accountable. On the financial management side, Facebook should redouble its efforts to engage with more diverse companies. While Facebook states that many of its financial needs are limited and therefore do not result in significant financial gains for asset management firms, engaging with diverse institutions can have positive impacts that are not reducible or limited to brokerage fees earned.

When so much of our world has moved online, Facebook's advertising tools can have a significant impact. They can help small businesses find new customers and build their customer base, and can enable nonprofits and public service organizations to get important information and resources to the communities that need them the most. They also can determine whether one learns of an advertised, available job, housing, or credit opportunity, or does not. While recognizing that there are positive uses for advertising tools, the civil rights community has long been concerned that Facebook's advertising tools could be used in discriminatory ways.

Over the last few years, several discrimination lawsuits were filed against Facebook alleging that its ad tools allowed advertisers to choose who received their ads and, in doing so, permitted advertisers to discriminate by excluding people from seeing ads for housing, employment, or credit opportunities based gender, race, age, and other personal characteristics. In March 2019, Facebook settled discrimination lawsuits brought by the National Fair Housing Alliance, Communications Workers of America, the American Civil Liberties Union, and private parties.

The June 2019 Audit Report described five major changes Facebook was making to its ad targeting system to prevent Facebook's ad tools from being used for discrimination. This chapter provides updates on Facebook's progress implementing these five commitments, describes new developments, and identifies areas for further analysis and improvement.

*First*, Facebook agreed to build a separate advertising flow for creating US housing, employment, and credit ("HEC") opportunity ads on Facebook, Instagram, and Messenger with limited targeting options. Facebook states that it fulfilled this commitment in December 2019 when this flow became mandatory across all the tools businesses use to buy ads on Facebook. When an advertiser identifies their ad as offering housing, employment or credit, they are not permitted to target based on gender, age, or any interests that appear to describe people of a certain race, religion, ethnicity, sexual orientation, disability status, or other protected class. They are also prohibited from targeting ads based on narrow location options, including ZIP code (which can correlate with protected class given residential segregation patterns). Facebook has made Lookalike targeting unavailable to advertisers using the HEC flow (Lookalike targeting is when an advertiser provides Facebook a customer list and Facebook identifies users who are similar to those on the list who are then targeted for advertising). Instead of Lookalike targeting, Facebook states that advertisers using the HEC flow are only able to create Special Ad Audiences — audiences selected based on similarities in online behavior and activity to those on a customer list but without considering age, gender, ZIP code or FB group membership.

There has been some criticism or skepticism as to whether and how effectively Facebook will ensure that HEC ads are actually sent through the restricted flow (as opposed to sneaking into the old system where protected class targeting options remain available). Facebook indicates that it uses a combination of automated detection and human review to catch advertisers that may attempt to circumvent these restrictions. As part of its settlement, Facebook has committed to continuous refinement of the automated detection system so it is as effective as possible.

*Second,* Facebook committed to providing advertisers with information about Facebook’s non-discrimination policy and requiring them to certify that they will comply with the policy as a condition of using Facebook’s advertising tools. Although Facebook’s Terms and Advertising Policies had contained prohibitions against discrimination even before the settlement, that policy was not widely known or well-enforced. Facebook updated its “Discriminatory Practices” ad policy in June, 2019 to state: “Any United States advertiser or advertiser targeting the United States that is running credit, housing or employment ads, must self identify as a Special Ad Category, as it becomes available, and run such ads with approved targeting options.” Before certifying, advertisers are directed to Facebook’s non-discrimination policy, and are shown examples illustrating what ad targeting behavior is permitted and not permitted under the policy. Advertisers are also provided with external links where they can find more information about complying with non-discrimination laws.

Facebook began asking advertisers to certify compliance with its non-discrimination policy in 2018, but in 2019 it made the certification mandatory and began requiring all advertisers to comply. Facebook reports that since late August 2019, all advertisers must certify compliance with the non-discrimination policy; those who attempt to place an ad but have not yet completed the certification receive a notice preventing their ad from running until the certification is complete. Facebook designed the certification experience in consultation with outside experts to underscore the difference between acceptable ad targeting and ad discrimination.

*Third,* Facebook committed to building a section in its Ad Library for US housing ads that includes all active ads for housing (sale or rental), housing-related financing (*e.g.*, home mortgages), and related real estate transactions (*e.g.*, homeowners’ insurance or appraisal services). The purpose of this section is to help ensure that all housing ads are available to everyone (including non-Facebook users), regardless whether a user was in the advertiser’s intended audience for the ad or actually received the ad. The Library is searchable by the name of the Page running an ad or the city or state to which the ad is targeted. The housing section of Facebook’s Ad Library went live on December 4, 2019. Facebook reports that the Library now contains all active housing opportunity ads targeted at the US that started running or were edited on or after that date.

In addition to following through on the commitments discussed in the last report, Facebook also expanded on those commitments by agreeing to extend all of these changes to Canada by the end of the year.

Facebook committed in the June 2019 Audit Report to go above and beyond its obligations as part of its settlement of the discrimination cases and build Ad Library sections for employment and credit ads too. Like the housing section, Facebook agreed to also make all active ads for job opportunities or credit offers (*e.g.*, credit card or loan ads) available to everyone, including non-Facebook users. Facebook reports that it is actively building the employment and credit sections of the Ad Library now, and plans to launch them by the end of the year.

*Fourth,* Facebook committed to engage the National Fair Housing Alliance to conduct a training for key employees with advertising-related responsibilities on fair housing and fair lending laws. Facebook indicates that the National Fair Housing Alliance is in the process of developing the training (in partnership with Facebook’s Learning and Development team), and expects to deliver the training in early 2021. Given the importance of understanding these issues, the Auditors would like to see more than one training take place, whether through periodic refresher training, or training updates, or some other training format.

*Fifth,* while Facebook did not make any specific commitments in the last report regarding its algorithmic system for delivering ads, it did agree to engage academics, researchers, civil rights and privacy advocates, and other experts to study the use of algorithms by social media platforms. Part of that commitment included studying the potential for bias in such systems. While concepts of discrimination and bias have long been applied to models, advancements in the complexity of algorithms or machine learning models, along with their increasingly widespread use, have led to new and unsettled questions about how best to identify and remedy potential bias in such complicated systems. Facebook reports that since the last report it has participated in several ongoing engagements, including:

- Creating a series of "Design Jams" workshops through Facebook's Trust Transparency and Control (TTC) Labs initiative, in which stakeholders from industry, civil society and academia focused on topics like algorithmic transparency and fairness both in the advertising context and more broadly. Facebook states that more such workshops are planned over the coming months.

- Conducting roundtable discussions and consultations with stakeholders (*e.g.*, The Center for Democracy and Technology, The Future of Privacy Forum) on ways of advancing both algorithmic fairness and privacy—many approaches to measuring fairness in algorithms require collecting or estimating additional sensitive data about people, such as their race, which can raise privacy and other concerns. Facebook reports that it is working to better understand expectations and recommendations in this area.

Facebook also agreed to meet regularly with the Plaintiffs in the lawsuits and permit them to engage in testing of Facebook's ad platform to ensure reforms promised under the settlements are implemented effectively. Both of these commitments are underway.

While Facebook deserves credit for implementing these prior advertising commitments, it is important to note that these improvements have not fully resolved the civil rights community's discrimination concerns. Most of the changes Facebook made in 2019 focused on the targeting of ads and the choices advertisers were making on the front end of the advertising process; civil rights advocates remain concerned about the back end of Facebook's advertising process: ad delivery.

In March 2019, the Department of Housing and Urban Development (HUD) filed charges against Facebook alleging not only that Facebook's ad targeting tools allow for discrimination, but that Facebook also discriminated in delivering ads (choosing which of the users within an ad's target audience should be shown a given ad) in violation of fair housing laws. That charge remains pending.

Furthermore, in December 2019, Northeastern University and the non-profit Upturn released a new study of Facebook's advertising system that was carried out after the 2019 targeting restrictions were put into place. The study suggested that Facebook's Special Ad Audiences algorithms may lead to biased results despite the removal of protected class information.

In addition to the efforts referenced above, Facebook has said that it is continuing to invest in approaches to studying and addressing such issues, and is consulting with experts globally to help refine its approach to algorithmic fairness generally and concerns related to ads delivery in particular. The Auditors believe that it is critical that Facebook's expert consultations include engagement with those who have specific expertise in civil rights, bias, and discrimination concepts (including specifically fair housing, fair lending, and employment discrimination), and their application to algorithms. More details on Facebook's work can be found in the Algorithmic Bias section of this report.

From the Auditors' perspective, participating in stakeholder meetings and engaging with academics and experts is generally positive, but it does not reflect the level of urgency felt in the civil rights community for Facebook to take action to address long-standing discrimination concerns with Facebook's ad system — specifically ad delivery. The civil rights community views the most recent Upturn study as further indication that the concern they have been expressing for years — that Facebook's ad system can lead to biased or discriminatory results — may be well-placed. And while civil rights advocates certainly do not want Facebook to get it wrong when it comes to data about sensitive personal characteristics or measuring algorithmic fairness, they are concerned that it is taking Facebook too long to get it right — and harm is being done in the interim.

Algorithms, machine-learning models, and artificial intelligence (collectively "AI") are models that make connections or identify patterns in data and use that information to make predictions or draw conclusions. AI is often presented as objective, scientific and accurate, but in many cases it is not. Algorithms are created by people who inevitably have biases and assumptions, and those biases can be injected into algorithms through decisions about what data is important or how the algorithm is structured, and by trusting data that reflects past practices, existing or historic inequalities, assumptions, or stereotypes. Algorithms can also drive and exacerbate unnecessary adverse disparities. Oftentimes by repeating past patterns, inequality can be automated, obfuscating and perpetuating inequalities. For example, as one leading tech company learned, algorithms used to screen resumes to identify qualified candidates may only perpetuate existing gender or racial disparities if the data used to train the model on what a qualified candidate looks like is based on who chose to apply in the past and who the employer hired; in the case of Amazon the algorithm "learned" that references to being a woman (*e.g.*, attending an all-female college, or membership in a women's club) was a reason to downgrade the candidate.

Facebook uses AI in myriad ways, such as predicting whether someone will click on an ad or be interested in a Facebook Group, whether content is likely to violate Facebook policy, or whether someone would be interested in an item in Facebook's News Feed. However, as algorithms become more ubiquitous in our society it becomes increasingly imperative to ensure that they are fair, unbiased, and non-discriminatory, and that they do not merely magnify pre-existing stereotypes or disparities. Facebook's algorithms have enormous reach. They can impact whether someone will see a piece of news, be shown a job opportunity, or buy a product; they influence what content will be proactively removed from the platform, whose account will be challenged as potentially inauthentic, and which election-related ads one is shown. The algorithms that Facebook uses to flag content as potential hate speech could inadvertently flag posts that condemn hate speech. Algorithms that make it far more likely that someone of one age group, one race or one sex will see something can create significant disparities — with some people being advantaged by being selected to view something on Facebook while others are disadvantaged.

When it comes to algorithms, assessing fairness and providing accountability are critical. Because algorithms work behind the scenes, poorly designed, biased, or discriminatory algorithms can silently create disparities that go undetected for a long time unless systems are in place to assess them. The Auditors believe that it is essential that Facebook develop ways to evaluate whether the artificial intelligence models it uses are accurate across different groups and whether they needlessly assign disproportionately negative outcomes to certain groups.

## A. Responsible AI Overview

Given the critical implications of algorithms, machine-learning models, and artificial intelligence for increasing or decreasing bias in technology, Facebook has been building and growing its Responsible Artificial Intelligence capabilities over the last two years. As part of its Responsible AI (RAI) efforts, Facebook has established a multi-disciplinary team of ethicists, social and political scientists, policy experts, AI researchers and engineers focused on understanding fairness and inclusion concerns associated with the deployment of AI in Facebook products. The team's goal is to develop guidelines, tools and processes to help promote fairness and inclusion in AI at Facebook, and make these resources widely available across the entire company so there is greater consistency in approaching questions of AI fairness.

During the Audit process, the Auditors were told about Facebook's four-pronged approach to fairness and inclusion in AI at Facebook: (1) creating guidelines and tools to identify and mitigate unintentional biases; (2) piloting a fairness consultation process; (3) participating in external engagement; and (4) investing in diversity of the Facebook AI team. Facebook's approach is described in more detail below, along with and observations from the Auditors.

1. **Creating guidelines and tools to identify and mitigate unintentional biases that can arise when the AI is built and deployed.**

There are a number of ways that bias can unintentionally appear in the predictions an AI model makes. One source of bias can be the underlying data used in building and training the algorithm; because algorithms are models for making predictions, part of developing an algorithm involves training it to accurately predict the outcome at issue, which requires running large data sets through the algorithm and making adjustments. If the data used to train a model is not sufficiently inclusive or reflects biased or discriminatory patterns, the model could be less accurate or effective for groups not sufficiently represented in the data, or could merely repeat stereotypes rather than make accurate predictions. Another source of potential bias are the decisions made and/or assumptions built in to how the algorithm is designed. To raise awareness and help avoid these pitfalls, Facebook has developed and continues to refine guidelines as well as a technical toolkit they call the Fairness Flow.

The Fairness Flow is a tool that Facebook teams use to assess one common type of algorithm. It does so in two ways: (1) it helps to flag potential gaps, skews, or unintended problems with the data the algorithm is trained on and/or instructions the algorithm is given; and (2) it helps to identify undesired or unintended differences in how accurate the model's predictions are for different groups or subgroups and whether the algorithms settings (*e.g.*, margins of error) are in the right place. The guidelines Facebook has developed include guidance used in applying the Fairness Flow.

The Fairness Flow and its accompanying guidelines are new processes and resources that Facebook has just begun to pilot. Use of the Fairness Flow and guidelines is voluntary, and they are not available to all teams. While the Fairness Flow has been in development longer than the guidelines, both are still works in progress and have only been applied a limited number of times. That said, Facebook hopes to expand the pilot and extend the tools to more teams in the coming months.

Facebook identified the following examples of how the guidelines and Fairness Flow have been initially used:

- When Facebook initially built a camera for its Portal product that automatically focuses the camera around people in the frame, it realized the tracking did not work as well for certain genders and skin tones. In response, Facebook relied on its guidelines to build representative test datasets across different skin tones and genders. Facebook then used those data sets on the algorithm guiding the camera technology to improve Portal's effectiveness across genders and skin tones.

- During the 2019 India general elections, in order to assist human reviewers in identifying and removing political interference content, Facebook built a model to identify high risk content (for example, content that discussed civic or political issues). Facebook used the Fairness Flow tool to ensure that the model's predictions as to whether content was civil/political were accurate across languages and regions in India. (This is important because systematically underestimating risk for content in a particular region or language, would result in fewer human review resources being allocated to that region or language than necessary.)

**2. Piloting a fairness consultation process.**

Facebook has also begun to explore ways to connect the teams building Facebook's AI tools and products to those on Facebook's Responsible AI team with more expertise in fairness in machine learning, privacy, and civil rights. Beginning in December 2019, Facebook began piloting a fairness consultation process, by which product teams who have identified potential fairness, bias, or privacy-related concerns associated with a product they are developing can reach out to a core group of employees with more expertise in these areas for guidance, feedback, or a referral to other employees with additional subject matter expertise in areas such as law, policy, ethics, and machine learning.

As part of this pilot effort, a set of issue-spotting questions was developed to help product teams and their cross-functional partners identify potential issues with AI fairness or areas where bias could seep in, and flag them for additional input and discussion by the consultative group. Once those issues are discussed with the core group, product teams either proceed with development on their own or continue to engage with the core group or others on the Responsible AI team for additional support and guidance.

This emerging fairness consultation process is currently only a limited pilot administered by a small group of employees, but is one way Facebook has begun to connect internal subject matter experts with product teams to help issue spot fairness concerns and subsequently direct them to further resources and support. (Part of the purpose of the pilot is to also identify those areas where teams need support but where internal guidance and expertise is lacking or underdeveloped so that the company can look to bring-in or build such expertise.) As a pilot, this is a new and voluntary process, rather than something that product teams are required to complete. But, Facebook asserts that its goal is to take lessons from these initial consultations and use them to inform the development of longer-term company processes and provide more robust guidance for product teams. In other words, part of the purpose of the pilot is to better understand the kinds of questions product teams have, and the kind of support that would be most effective in assisting teams to identify and resolve potential sources of bias or discrimination during the algorithm development process.

**3. Participating in external engagement.**

Because AI and machine learning is an evolving field, questions are constantly being raised about how to ensure fairness, non-discrimination, transparency, and accountability in AI systems and tools. Facebook recognizes that it is essential to engage with multiple external stakeholders and the broader research communities on questions of responsible AI.

Facebook reports that it has been engaging with external experts on AI fairness issues in a number of ways, including:

- Facebook co-founded and is deeply involved in the Partnership on AI (PAI), a multistakeholder organization that seeks to develop and share AI best practices. Facebook states that it is active in PAI working groups around fair, transparent, and accountable AI and initiatives including developing documentation guidelines to enable greater transparency of AI systems, exploring the role of gathering sensitive user data to enable testing for algorithmic bias and discrimination, and engaging in dialogue with civil society groups about facial recognition technologies.

- Facebook reports that in January 2020 that it sent a large delegation including engineers, product managers, researchers, and policy staff to the Fairness, Transparency, and Accountability Conference, the leading conference on fairness in machine learning, in order to connect with multidisciplinary academic researchers, civil society advocates, and industry peers and discuss challenges and best practices in the field.

- Facebook is part of the expert group that helped formulate the Organization for Economic Cooperation & Development’s (OECD) AI principles which include a statement that “AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity.” Facebook states that it is now working with the OECD Network of Experts on AI to help define what it means to implement these principles in practice.

- Trust Transparency and Control (TTC) Labs is an industry collaborative created to promote design innovation that helps give users more control of their privacy. TTC Labs includes discussion of topics like algorithmic transparency, but Facebook states that it is exploring whether and how to expand these conversations to include topics of fairness and algorithmic bias.

Through these external engagements, Facebook reports that it has begun exploring and debating a number of important topics relating to AI bias and fairness. For example, Facebook has worked with, and intends to continue to seek input from, experts to ensure that its approaches to algorithmic fairness and transparency are in line with industry best practices and guidance from the civil rights community. Even where laws are robust, and even among legal and technical experts, there is sometimes disagreement on what measures of algorithmic bias should be adopted—and approaches can sometimes conflict with one another. Experts are proposing ways to apply concepts like disparate treatment and disparate impact discrimination, fairness, and bias to evaluate machine learning models at scale, but consensus has not yet been reached on best practices that can be applied across all types of algorithms and machine-learning models.

Similarly, Facebook has been considering questions about whether and how to collect or estimate sensitive data. Methods to measure and mitigate bias or discrimination issues in algorithms that expert researchers have developed generally require collecting or estimating data about people’s sensitive group membership. In this way, the imperative to test and address bias and discrimination in machine learning models along protected or sensitive group lines can trigger the need to have access to, or estimate, sensitive or demographic data in order to perform those measurements. Indeed, this raises privacy, ethical, and representational questions like:

- Who should decide whether this sensitive data should be collected?
- What categories of data should private companies collect (if any)?
- When is it appropriate to infer or estimate sensitive data about people for the purpose of testing for discrimination?
- How should companies balance privacy and fairness goals?

These questions are not unique to Facebook: they apply to any company or organization that has turned to machine learning, or otherwise uses quantitative techniques to measure or mitigate bias or discrimination. In some other industries laws, regulations, or regulatory guidance, and/or the collective efforts of industry members answer these questions and guide the process of collecting or estimating sensitive information to enable industry players and regulators to measure and monitor discrimination. Facebook asserts that for social media companies like it, answering these questions requires broad conversations with stakeholders and policymakers about how to chart a responsible path forward. Facebook states that it has already been working with the Partnership on AI to initiate multi-stakeholder conversations (to include civil rights experts) on this important topic, and plans to consult with a diverse group of stakeholders on how to make progress in this area. Facebook also reports that it is working to better understand the cutting edge work being done by companies like Airbnb and determine if similar initiatives are applicable and appropriate for companies that are the size and scale of Facebook.

**4. Investing in the Diversity of the Facebook AI team.**

A key part of driving fairness in algorithms in ensuring companies are focused on increasing the diversity of the people working on and developing FB’s algorithms. Facebook reports that it has created a dedicated Task Force composed of employees in AI, Diversity and HR who are focused on increasing the number of underrepresented minorities and women in the AI organization and building an inclusive AI organization.

The AI Task Force has led initiatives focused on increasing opportunities for members of underrepresented communities in AI. These initiatives include:

**(i) Co-teaching and funding a deep learning course at Georgia Tech.** In this pilot program, Facebook developed, co-taught and led a 4 month program for 250+ graduate students with the aim to build a stronger pipeline of diverse candidates. Facebook states that its hope is that a subset of participating students will interview for future roles at Facebook. Facebook intends to scale this program to thousands of underrepresented students by building a consortium with 5-6 other universities, including minority-serving institutions.

**(ii) Northeastern’s Align Program.** Facebook also recently provided funding for Northeastern University’s Align program, which is focused on creating pathways for non-computer science majors to switch over to a Master’s Degree in Computer Science, with the goal of increasing the pipeline of underrepresented minority and female students who earn degrees in Computer Science. Facebook reports that its funding enabled additional universities to join the Align consortium, including: Georgia Tech, University of Illinois at Urbana–Champaign, and Columbia.

In addition to focusing on increasing diversity overall in AI, Facebook states that it has also increased hiring from civil society including nonprofits, research, and advocacy organizations that work closely with major civil rights institutions on emerging technology-related challenges — and these employees are actively engaged in the Responsible AI organization.

## B. Auditor Observations

It is important that Facebook has publicly acknowledged that AI can be biased and discriminatory and that deploying AI and machine learning models brings with it a responsibility to ensure fairness and accountability. The Auditors are encouraged that Facebook is devoting resources to studying responsible AI methodologies and engaging with external experts regarding best practices.

When it comes to Facebook's own algorithms and machine learning models, the Auditors cannot speak to the effectiveness of any of the pilots Facebook has launched to better identify and address potential sources of bias or discriminatory outcomes. (Both because the pilots are still in nascent stages and the Auditors have not had full access to the full details of these programs.) The Auditors do, however, credit Facebook for taking steps to explore ways to improve Facebook's AI infrastructure and develop processes designed to help spot and correct biases, skews, and inaccuracies in Facebook's models.

That being said, the Auditors strongly believe that processes and guidance designed to prompt issue-spotting and help resolve fairness concerns must be mandatory (not voluntary) and company-wide. That is, all teams building models should be required to follow comprehensive best practice guidance and existing algorithms and machine-learning models should be regularly tested. This includes both guidance in building models and systems for testing models.

And while the Auditors believe it is important for Facebook to have a team dedicated to working on AI fairness and bias issues, ensuring fairness and non-discrimination should also be a responsibility for all teams. To that end, the Auditors recommend that training focused on understanding and mitigating against sources of bias and discrimination in AI should be mandatory for all teams building algorithms and machine-learning models at Facebook and part of Facebook's initial onboarding process.

Landing on a set of widely accepted best practices for identifying and correcting bias or discrimination in models or for handling sensitive data questions is likely to take some time. Facebook can and should be a leader in this space. Moreover, Facebook cannot wait for consensus (that may never come) before building an internal infrastructure to ensure that the algorithms and machine learning models it builds meet minimum standards already known to help avoid bias pitfalls (*e.g.*, use of inclusive data sets, critical assessment of model assumptions and inferences for potential bias, *etc*.). Facebook has an *existing* responsibility to ensure that the algorithms and machine learning models that can have important impacts on billions of people do not have unfair or adverse consequences. The Auditors think Facebook needs to approach these issues with a greater sense of urgency. There are steps it can take now — including mandatory training, guidance on known best practices, and company-wide systems for ensuring that AI fairness guidance are being followed — that would help reduce bias and discrimination concerns even before expert consensus is reached on the most challenging or emergent AI fairness questions.

Given the vast amount of data Facebook has and the reach of its platform, the civil rights community has repeatedly raised concerns about user privacy. These concerns were only exacerbated by the Cambridge Analytica scandal in which the data of up to 87 million Facebook users was obtained by Cambridge Analytica without the express consent of the majority of those users.

While the larger digital privacy discourse has focused on issues such as transparency, data collection minimization, consent, and private rights of action, the civil rights and privacy communities are increasingly focused on the tangible civil rights and civil liberties harms that flow from social media data collection practices. Groups are concerned about the targeting of individuals for injurious purposes that can lead to digital redlining, discriminatory policing and immigration enforcement, retail discrimination, the targeting of advocates through doxxing and hate speech, identity theft, voter suppression, and a litany of other harms. In the wake of the COVID-19 pandemic and massive racial justice protests, these concerns are at an all-time high as people are more reliant on social media and digital platforms for civic activity and basic needs.

In recent years, the civil rights community has focused on the use of Facebook and Facebook data for law enforcement purposes. More specifically, civil rights and civil liberties groups have expressed concern about use of the platform to monitor or surveil people without their knowledge or consent by obtaining and scraping Facebook data, using facial recognition technology on Facebook users, or misrepresenting themselves to "investigate" people. There is particular concern that these tactics could be used to focus on communities of color.

Also, collection of personal social media data can also have enormous consequences for lawful and undocumented immigrants and the people they connect with on Facebook. For example, in a program starting in 2019, the State Department began collecting and reviewing social media accounts for most visa applicants and visitors entering the United States, affecting some 15 million travelers per year. The Department of Homeland Security (DHS) is building upon this. Although Facebook continues to push back on governments (and this use of social media data specifically), the use of public social media data by law enforcement and immigration authorities is seemingly ever-expanding in ways that can have significant privacy (and civil rights) implications.

Facebook's announcements regarding its planned adoption of end-to-end encryption for all of its messaging products have been praised by some privacy, human rights and civil liberties groups as an important step to protect the privacy, data security and freedom of expression rights for billions of users. However, the issue cuts both ways. Civil rights and anti-hate groups have also raised questions, given that encryption can prevent Facebook and law enforcement from proactively accessing or tracing harmful content such as hate speech, viral misinformation, efforts to engage in human trafficking or child exploitation.

This chapter provides an overview of the changes Facebook has recently implemented to provide increased privacy protections, including those adopted in connection with its 2019 settlement with the Federal Trade Commission. It also shines a light on Facebook's current policies with respect to the use of facial recognition technology, law enforcement's use of Facebook and access to Facebook data, data scraping, end-to-end encryption and COVID-tracing.

By providing transparency on these issues, the Auditors' goal is to inform future conversations between Facebook and advocates on the company's current policies and practices. While intervening events (such as time-sensitive Census and election-related issues and the COVID-19 crisis) prevented the Auditors from conducting the kind of comprehensive analysis of Facebook's privacy policies and practices necessary to make detailed recommendations, the Auditors hope that this chapter helps lay the groundwork for future engagement, analysis, and advocacy on privacy issues at Facebook.

## A. Privacy Changes from FTC Settlement

In July 2019, Facebook entered into a $5 billion settlement with the Federal Trade Commission (FTC) to resolve claims stemming from allegations that Facebook violated a prior agreement with the FTC by giving entities access to data that users had not agreed to share. That settlement was formally approved in court in April 2020. The agreement requires a fundamental shift in the way Facebook approaches building products and provides a new framework for protecting people's privacy and the information they give Facebook.

Through the settlement, Facebook has agreed to significant changes to its privacy policies and the infrastructure it has built for flagging and addressing privacy risks. Specifically, under the settlement Facebook will, among other things:

- Develop a process for documenting and addressing identified privacy risks during the product development process;
- Conduct a privacy review of every new or modified product, service, or practice before it is implemented and document its decisions about user privacy;
- Create a committee on its Board of Directors responsible for independently reviewing Facebook's compliance with its privacy commitments under the settlement;
- Designate privacy compliance officer(s) responsible for implementing Facebook's compliance program who are removable solely by the Board committee
- Engage an independent privacy assessor whose job will be to review Facebook's privacy program on an ongoing basis and report to the Board committee and the FTC, if they see compliance breakdowns or opportunities for improvement;
- Provide to the FTC quarterly and annual certifications signed by Mark Zuckerberg attesting to the compliance of the Privacy Program; and
- Report to the FTC any incidents in which Facebook has verified or otherwise confirmed that the personal information of 500 or more users was likely to have been improperly accessed, collected, used, or shared by a third party in a manner that violates the terms under which Facebook shared the data with them.

Facebook is working on implementing these new commitments. The company announced the membership of the Privacy Committee of the Board of Directors. The company also reports that it has added new members to its privacy leadership team, created dozens of technical and non-technical teams that are dedicated only to privacy, and currently have thousands of people working on privacy-related projects with plans to hire many more. Facebook reports that it has also updated the process by which they onboard every new employee at Facebook to make sure they think about their role through a privacy lens, design with privacy in mind and work to proactively identify potential privacy risks so that mitigations can be implemented. All new and existing employees are required to complete annual privacy training. Facebook further reports that it is looking critically at data use across its operations, including assessing how data is collected, used, and stored.

It is worth noting that despite these commitments, critics of the settlement contend that it did not go far enough because it did not impose any penalties on Facebook leadership and does not do enough to change the incentives and data gathering practices that led to the underlying privacy violations.

## B. Law Enforcement's Use of Facebook & Access to Facebook Data

When it comes to sharing user information or data with law enforcement, Facebook states that it provides such access only in accordance with applicable law and its terms of service. According to Facebook, that means that except in cases of emergency, its policy is to provide data to US law enforcement entities only upon receipt of a valid subpoena, court order, or warrant. Law enforcement officials may submit requests for information through Facebook's Law Enforcement Online Request System, which requires certification that the requesting person is a member of law enforcement and uses a government-issued email address. Facebook indicates that it provides notice to the person whose data is being sought unless it is prohibited by law from doing so or in exceptional circumstances, such as child exploitation cases or emergencies.

Facebook defines "emergency circumstances" as those involving imminent risk of harm to a child or risk of death or serious physical injury to anyone. In those cases, Facebook states that it will allow disclosure of information without the delay associated with obtaining a warrant, subpoena, or court order. According to Facebook's most recent Transparency Report, these emergency requests for user data make up approximately 11% of the data requests Facebook receives, and Facebook provides at least some requested data in response to such emergency requests approximately 74% of the time.

Facebook's authenticity policies prohibit users from misrepresenting who they are, using fake accounts, or having multiple accounts. Facebook does not have any exceptions to those policies for law enforcement. Accordingly, it is against Facebook policy for members of law enforcement to pretend they are someone else or use a fake or "undercover" alias to hide their law enforcement identities. Facebook states that it takes action against law enforcement that violate these policies. In 2018, Facebook learned that the Memphis Police Department set up fake accounts as part of a criminal investigation; in response, Facebook disabled the fake accounts it identified and wrote a public letter to the Department calling out the policy violations and directing it to cease such activities.

That being said, Facebook does not restrict law enforcement's ability (or anyone's ability) to access the public information users post on Facebook, including public posts, photos, profiles, likes, and friend networks — so long as law enforcement personnel do not misrepresent their identities in doing so. Further, Facebook's current policy does not prohibit law enforcement from posting on police or other law enforcement department Facebook pages images of or allegations about alleged suspects, persons of interest, arrestees, or people the department thinks might have connections to criminal or gang organizations — including those who have not been convicted (or even charged) with anything. (The only limitation on law enforcement's ability to use Facebook this way are Facebook's other policies, such as those prohibiting the posting of personal identifying information like social security numbers or home addresses, or Facebook's bullying and harassment policy.)

## C. Facial Recognition Technology

Facebook has several products and features that rely on facial recognition technology. One example is Facebook's "Photo Review" feature that is part of the Face Recognition setting. When that setting is turned on, a user is notified if they appear in photos uploaded by other users, even if they are not tagged, as long as the user has permission to see the photo based on the photo's privacy setting. This gives the user the option to tag themselves in the photo, leave the photo as is, reach out to the person who posted the photo or report the photo if the user has concerns. Facial recognition also allows Facebook to describe photos to people who use screen-reading assistive technology.

In 2017 and 2018, Facebook sent a notice to all users explaining the face recognition setting, how it works, and how users can enable or disable the setting. New users receive a similar notice. Facebook also includes in its Help Center an explanation of how the company uses their face profile or "template" and how users can turn that setting on or off. According to Facebook, facial recognition is disabled by default, and users would have to affirmatively turn the feature on in order for the technology to be activated. If a user turns the facial recognition setting off, Facebook automatically deletes the face template it has which allows Facebook to recognize that user based on images. (That said, where a user has already been tagged in a photo, turning off facial recognition does not untag the photo.)

In addition to on-platform uses, the civil rights community has sought clarity on whether/how Facebook makes facial recognition technology or data available off platform to government agencies, law enforcement entities, immigration officials, or private companies. Facebook maintains that it does not share facial recognition information with third parties, nor does it sell or provide its facial recognition technology to other entities. Facebook further indicates that it built its facial recognition technology to be unique to Facebook, meaning that even if someone were to gain access to the data, they would not be able to use it with other facial recognition systems because it (intentionally) does not work with other systems.

New or proposed uses of facial recognition are required to go through the privacy review described above and obtain approval before they can be launched.

Because facial recognition relies on algorithms, it necessarily raises the same questions of bias, fairness, and discrimination associated with AI more broadly. Facebook reports that it has been testing the algorithms that power its facial recognition system for accuracy when applied to people of different ages and genders since before 2017.

Facebook asserts that it began testing those algorithms for accuracy when applied to different skin tones starting in 2018. As a result of those tests, Facebook made adjustments to its algorithms in an effort to make them more accurate and inclusive. Facebook's testing of its facial recognition algorithms is in line with its new Inclusive AI initiative (announced in 2019 and described more fully in the Algorithmic Bias Chapter), through which the company is adopting guidelines to help ensure that the teams developing algorithms are using inclusive datasets and measuring accuracy across different dimensions and subgroups.

## D. Data Scraping

In the past few years (including in the wake of the Cambridge Analytica scandal) civil rights and privacy advocates have become increasingly concerned with data scraping (using technology to extract data from apps, websites, or online platforms without permission).

Since 2004, Facebook has prohibited data scraping and other efforts to collect or access data using automated technology from Facebook products or tools without prior permission from Facebook.

Facebook reports that in recent years it has continued to enhance its enforcement against scraping, including creating a team in 2019 that is dedicated to both proactively detecting (and preventing) scraping and conducting investigations in response to allegations of scraping. According to Facebook, it enforces its no-scraping policy through various means, including barring violators from using Facebook, cease and desist letters, and in some cases litigation. Last year, for example, Facebook sued two developers based in Ukraine who operated malicious software designed to scrape data from Facebook and other social networking sites. Recently, Facebook filed lawsuits against unauthorized automated activity — specifically data scraping and building software to distribute fake likes and comments on Instagram.

## E. End-to-End Encryption

End-to-end encryption is a system in which messages or communications between users are encrypted throughout the communication process such that the entity providing the communication service (such as WhatsApp or Messenger) cannot access or review the content of the messages. Advocates for such encryption maintain that it protects user privacy and security by ensuring that their private messages cannot be surveilled or accessed by third parties, whether those be government entities, criminal hackers, advertisers, or private companies. These protections against access can be critical for whistleblowers, protest organizers, individuals subject to government surveillance or suppressive regimes, public figures subject to targeted hacking, those who handle sensitive information, and many others. However, critics of end-to-end encryption have expressed concern that it may make it harder to identify and take action against individuals whose communications violate laws or Facebook policies, such as those running financial scams or seeking to harm or exploit children.

Although WhatsApp is already end-to-end encrypted and Messenger offers an opt-in end-to-end encrypted service, Facebook announced in 2019 that it plans to make its communication services, namely Messenger and Instagram Direct, fully end-to-end encrypted by default. To address concerns about shielding bad actors, Facebook indicates that alongside encryption, it is investing in new features that use advanced technology to help keep people safe without breaking end-to-end encryption and other efforts to facilitate increased reporting from users of harmful behavior/content communicated on encrypted messaging systems.

More specifically, Facebook states that it is using data from behavioral signals and user reports to build and train machine-learning models to identify account activity associated with specific harms such as child exploitation, impersonation, and financial scams. When these potentially harmful accounts interact with other users, a notice will surface to educate users on how to spot suspicious behavior and avoid unwanted or potentially harmful interactions so that wrongdoers can be detected and people can be protected even without breaking end-to-end encryption. In addition, Facebook reports that it is improving its reporting options to make them more easily accessible to users by, for example, inserting prompts asking users if they want to report a person or content.

Regardless of whether the content is end-to-end encrypted, Facebook permits users to report content that’s harmful or violates Facebook’s policies, and, in doing so, provide Facebook with the content of the messages. In other words, end-to-end encryption means that Facebook cannot proactively access message content on its own, but users are still permitted to voluntarily provide Facebook with encrypted content. This allows Facebook to continue to review and determine whether it is violating and then impose penalties and/or report the matter to law enforcement, if necessary.

## F. COVID-19 Tracing

In an effort to track the spread of COVID-19 and warn those who may have been exposed, contact tracing has been increasingly advanced as an important tool for containing the virus. Given the amount of data Facebook has and the number of Facebook users, some have called for Facebook to directly participate in contact tracing efforts. Others, however, have expressed concern that sharing information about the locations or contacts of those who have contracted the virus would be an unacceptable invasion of privacy.

Facebook has not participated in the development of contact tracing apps, but has received requests from government and private entities asking Facebook to promote contact tracing apps on Facebook through ad credits or News Feed notifications to users. Facebook states that it has not granted any such requests. If it were to do so, the apps would need to undergo a privacy review. Facebook has, however, promoted voluntary surveys conducted by third-party academic research institutions to track and study COVID-19 through users self-reported symptoms. (The research institutions do not share any individual survey responses with Facebook and Facebook does not share individual user information with the research institutions.)

Through its Data for Good initiative, Facebook also makes aggregate data available to researchers to assist them in responding to humanitarian crises, including things like the COVID-19 pandemic. Facebook has released to researchers (and the public) mobility data comparing how much people are moving around now versus before social distancing measures were put in place, and indicating what percentage of people appear to stay within a small area for the entire day. Only users who have opted in to providing Facebook with their location history and background location collection are included in the data set and the data shared is only shared on an aggregate level. Facebook asserts it has applied a special privacy protocol to protect people’s privacy in mobility datasets shared publicly and ensure that aggregated data cannot be disaggregated to reveal individual information .

Facebook is also taking steps to support manual contact tracing efforts — that is, efforts which promote awareness and understanding of off-platform tracing initiatives that do not involve the sharing of Facebook data. For example, through its COVID Information Center and advertising, Facebook is helping to disseminate information about contact tracing. Facebook states that it intends to continue to support such manual tracing efforts.

Facebook plans to continue with the work outlined above and will continue to assess where it can play a meaningful role in helping address the evolving health problems that society is facing related to COVID-19 with privacy in mind.

## G. Further Research

The specific issues discussed in this chapter are set against a larger digital privacy discourse that centers around transparency, data collection minimization, consent, the impacts of inaccurate data, and myriad potential civil rights implications depending on how captured data is used. The volume of data collected by technology companies on users, non-users associated with them, and both on- and off-platform activity requires that companies, including Facebook, be fully transparent about the ways data is collected and used. Without this transparency, users have no way of knowing whether the information collected on them is accurate, let alone any way to correct errors — and those inaccuracies can have significant consequences, especially for marginalized communities.

While beyond the capacity of this Audit, these privacy issues and their interconnectivity with topics like advertising, discriminatory policing and immigration enforcement, employment and lending discrimination, and algorithmic bias, are important issues with serious potential civil rights implications that are worth further study and analysis.