AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO

S. HRG. 116–639

DOES SECTION 230'S SWEEPING IMMUNITY ENABLE BIG TECH BAD BEHAVIOR?

HEARING

BEFORE THE

COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION UNITED STATES SENATE

ONE HUNDRED SIXTEENTH CONGRESS

SECOND SESSION

OCTOBER 28, 2020

Printed for the use of the Committee on Commerce, Science, and Transportation



Available online: http://www.govinfo.gov

DOES SECTION 230'S SWEEPING IMMUNITY ENABLE BIG TECH BAD BEHAVIOR?

S. HRG. 116–639

DOES SECTION 230'S SWEEPING IMMUNITY ENABLE BIG TECH BAD BEHAVIOR?

HEARING

BEFORE THE

COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION UNITED STATES SENATE

ONE HUNDRED SIXTEENTH CONGRESS

SECOND SESSION

OCTOBER 28, 2020

Printed for the use of the Committee on Commerce, Science, and Transportation



Available online: http://www.govinfo.gov

U.S. GOVERNMENT PUBLISHING OFFICE

54–131 PDF WASHINGTON : 2023

SENATE COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION

ONE HUNDRED SIXTEENTH CONGRESS

SECOND SESSION

ROGER WICKER, Mississippi, *Chairman*

JOHN THUNE, South Dakota
ROY BLUNT, Missouri
TED CRUZ, Texas
DEB FISCHER, Nebraska
JERRY MORAN, Kansas
DAN SULLIVAN, Alaska
CORY GARDNER, Colorado
MARSHA BLACKBURN, Tennessee
SHELLEY MOORE CAPITO, West Virginia
MIKE LEE, Utah
RON JOHNSON, Wisconsin
TODD YOUNG, Indiana
RICK SCOTT, Florida

MARIA CANTWELL, Washington, *Ranking*
AMY KLOBUCHAR, Minnesota
RICHARD BLUMENTHAL, Connecticut
BRIAN SCHATZ, Hawaii
EDWARD MARKEY, Massachusetts
TOM UDALL, New Mexico
GARY PETERS, Michigan
TAMMY BALDWIN, Wisconsin
TAMMY DUCKWORTH, Illinois
JON TESTER, Montana
KYRSTEN SINEMA, Arizona
JACKY ROSEN, Nevada

JOHN KEAST, *Staff Director*
CRYSTAL TULLY, *Deputy Staff Director*
STEVEN WALL, *General Counsel*
KIM LIPSKY, *Democratic Staff Director*
CHRIS DAY, *Democratic Deputy Staff Director*
RENAE BLACK, *Senior Counsel*

CONTENTS

	Page
Hearing held on October 28, 2020	1
Statement of Senator Wicker	1
Document entitled "Social Media Companies Censoring Prominent Conservative Voices"	66
Statement of Senator Cantwell	3
Statement of Senator Peters	18
Statement of Senator Gardner	21
Statement of Senator Klobuchar	23
Statement of Senator Thune	26
Statement of Senator Blumenthal	29
Statement of Senator Cruz	30
Statement of Senator Schatz	33
Statement of Senator Fischer	35
Statement of Senator Moran	40
Statement of Senator Markey	42
Statement of Senator Blackburn	44
Statement of Senator Udall	46
Statement of Senator Capito	48
Statement of Senator Baldwin	50
Statement of Senator Lee	52
Statement of Senator Duckworth	54
Statement of Senator Johnson	56
Statement of Senator Tester	59
Statement of Senator Scott	61
Statement of Senator Rosen	63

WITNESSES

Jack Dorsey, Chief Executive Officer, Twitter, Inc.	6
Prepared statement	7
Sundar Pichai, Chief Executive Officer, Alphabet Inc.	9
Prepared statement	11
Mark Zuckerberg, Chief Executive Officer, Facebook, Inc.	12
Prepared statement	14

APPENDIX

Letter dated October 27, 2020 to Senator Roger Wicker and Senator Maria Cantwell from Vanita Gupta, President and CEO, and LaShawn Warren, Executive Vice President for Government Affairs, The Leadership Conference on Civil and Human Rights	81
Response to written questions submitted to Jack Dorsey by:	
Hon. Roger Wicker	85
Hon. John Thune	87
Hon. Roy Blunt	92
Hon. Jerry Moran	93
Hon. Mike Lee	96
Hon. Ron Johnson	99
Hon. Maria Cantwell	100
Hon. Richard Blumenthal	109
Hon. Edward Markey	111
Hon. Gary Peters	112
Hon. Kyrsten Sinema	113
Hon. Jacky Rosen	115

Page

Response to written questions submitted to Sundar Pichai by:
Hon. Roger Wicker 118
Hon. John Thune 125
Hon. Jerry Moran 132
Hon. Mike Lee 137
Hon. Ron Johnson 141
Hon. Maria Cantwell 141
Hon. Amy Klobuchar 159
Hon. Richard Blumenthal 159
Hon. Edward Markey 163
Hon. Gary Peters 164
Hon. Kyrsten Sinema 167
Hon. Jacky Rosen 170
Response to written questions submitted to Mark Zuckerberg by:
Hon. Roger Wicker 176
Hon. John Thune 181
Hon. Jerry Moran 188
Hon. Mike Lee 192
Hon. Ron Johnson 198
Hon. Maria Cantwell 202
Hon. Amy Klobuchar 213
Hon. Richard Blumenthal 214
Hon. Edward Markey 217
Hon. Gary Peters 218
Hon. Kyrsten Sinema 225
Hon. Jacky Rosen 227
Publication dated July 8, 2020 entitled, "Facebook's Civil Rights Audit—Final Report" 233
Joint Publication dated October 21, 2020 entitled, "Complicit—The Human Cost of Facebook's Disregard for Muslim Life" by Muslim Advocates and the Global Project Against Hate and Extremism (GPAHE) 292

DOES SECTION 230'S SWEEPING IMMUNITY ENABLE BIG TECH BAD BEHAVIOR?

WEDNESDAY, OCTOBER 28, 2020

U.S. SENATE,
COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION,
Washington, DC.

The Committee met, pursuant to notice, at 10 a.m., in room SR–253, Russell Senate Office Building, Hon. Roger Wicker, Chairman of the Committee, presiding.

Present: Senators Wicker [presiding], Thune, Cruz, Fischer, Moran, Gardner, Blackburn, Capito, Lee, Johnson, Scott, Cantwell, Klobuchar, Blumenthal, Schatz, Markey, Udall, Peters, Baldwin, Duckworth, Tester, and Rosen.

OPENING STATEMENT OF HON. ROGER WICKER, U.S. SENATOR FROM MISSISSIPPI

The CHAIRMAN. This hearing will come to order. Senator Cantwell is going to join us in person, but she joins us remotely at the beginning of the hearing. We have convened this morning to continue the work of this committee to ensure that the Internet remains a free and open space and that the laws that govern it are sufficiently up to date.

The Internet is a great American success story, thanks in large part to the regulatory and legal structure our Government put in place. But we cannot take that success for granted. The openness and freedom of the Internet are under attack. Soon, we will hear from the CEOs of three of the most prominent Internet platforms Facebook, Google, and Twitter. Our witnesses include Mr. Jack Dorsey of Twitter, Mr. Sundar Pichai of Alphabet Incorporated and its subsidiary Google, and Mr. Mark Zuckerberg of Facebook. On October 1, this committee voted on a bipartisan and unanimous basis to approve the issuance of subpoenas. After discussions among representatives of the companies and the Committee, the witnesses agreed to attend the hearing voluntarily and remotely.

There is strong agreement on both sides of the aisle that hearing from these witnesses is important to deliberations before this committee, including deliberations on what legislative reforms are necessary to ensure a free and open Internet. For almost 25 years, the preservation of Internet freedom has been the hallmark of a thriving digital economy in the United States. This success has largely been attributed to a light touch regulatory framework and to Section 230 of the Communications Decency Act, often referred to as the 26 words that created the Internet. There is little dispute that Section 230 played a critical role in the early development and

growth of online platforms. Section 230 gave content providers protection from liability to remove and moderate content that they or their users considered to be, "obscene, lewd, lascivious, filthy, excessively violent, harassing or otherwise objectionable."

This liability shield has been pivotal in protecting online platforms from endless and potentially ruinous lawsuits. But it has also given these Internet platforms the ability to control, stifle, and even censor content in whatever manner meets their respective standards. The time has come for that free pass to end. After 24 years of Section 230 being the law of the land, much has changed. The Internet is no longer an emerging technology. Companies before us today are no longer scrappy startups operating out of a garage or a dorm room. They are now among the world's largest corporations, wielding immense power in our economy, culture, and public discourse. Immense power. The applications they have created are connecting the world in unprecedented ways, far beyond what lawmakers could have imagined three decades ago.

These companies are controlling the overwhelming flow of news and information that the public can share and access. One noteworthy example occurred just 2 weeks ago after our subpoenas were unanimously approved. The *New York Post,* the country's fourth largest newspaper ran a story revealing communications between Hunter Biden and a Ukrainian official. The report alleged that Hunter Biden facilitated a meeting with his father, Joe Biden, who was then Vice President of the United States. Almost immediately, both Twitter and Facebook took steps to block or limit access to the story. Facebook, according to its policy communications manager, began, "reducing its distribution on the platform," pending a third party fact check.

Twitter went beyond that, blocking all users, including the House Judiciary Committee, from sharing the article on feeds and through direct messages. Twitter even locked the *New York Post*'s account entirely claiming the story included hacked materials and was potentially harmful. It is worth noting that both Twitter and Facebook's aversion to hacked materials has not always been so stringent. For example, when the President's tax returns were illegally leaked, neither company acted to restrict access to that information. Similarly, the now discredited Steele dossier was widely shared without fact checking or disclaimers. This apparent double standard would be appalling under normal circumstances, but the fact that selective censorship is occurring in the midst of the 2020 election cycle dramatically amplifies the power wielded by Facebook and Twitter. Google recently generated its own controversy when it was revealed that the company threatened to cutoff several conservative websites, including the Federalist, from their ad platform.

Make no mistake, for sites that rely heavily on advertising revenue for their bottom line, being blocked from Google services or demonetized can be a death sentence. According to Google, the offense of these websites was posting user submitted comment sections that included objectionable content. But Google's own platform, YouTube hosts user submitted comment sections for every video uploaded. It seems that Google is far more zealous in policing conservative sites than its own YouTube platform for the same

types of offensive and outrageous language. It is ironic that when the subject is net neutrality, technology companies, including Facebook, Google, and Twitter, have warned about the grave threat of blocking or throttling the flow of information on the Internet.is

Meanwhile, these same companies are actively blocking and throttling the distribution of content on their own platforms and are using protections under Section 30 to do it. Is it any surprise that voices on the right are complaining about hypocrisy or even worse, anti-democratic election interference? These recent incidents are only the latest in a long trail of censorship and suppression of conservative voices on the Internet. Reasonable observers are left to wonder whether big tech firms are obstructing the flow of information to benefit one political ideology or agenda. My concern is that these platforms have become powerful arbiters of what is true and what content users can access. The American public gets little insight into the decisionmaking process when content is moderated and users have little recourse when they are censored or restricted. I hope we can all agree that the issues the Committee will discuss today are ripe for thorough examination and action.

I have introduced legislation to clarify the intent of Section 230s liability protections and increase the accountability of companies who engage in content moderation. The Online Freedom and Viewpoint Diversity Act would make important changes to right size the liability shield and make clear what type of content moderation is protected. This legislation would address the challenges we have discussed while still leaving fundamentals of Section 230 in place. Although some of my colleagues on the other side of the aisle have characterized this as a purely partisan exercise, there is strong bipartisan support for reviewing Section 230. In fact, both Presidential candidates, Trump and Biden, have proposed repealing Section 230 in its entirety, a position I have not yet embraced.

I hope we can focus today's discussion on the issues that affect all Americans. Protecting a true diversity of viewpoints and free discourse is central to our way of life. I look forward to hearing from today's witnesses about what they are doing to promote transparency, accountability, and fairness in their content moderation processes.

And I thank each of them for cooperating with us in the scheduling of this testimony. I now turn to my friend and Ranking Member, Senator Cantwell, for her opening remarks. Senator Cantwell.

STATEMENT OF HON. MARIA CANTWELL,
U.S. SENATOR FROM WASHINGTON

Senator CANTWELL. [No audio] . . . beautiful State of Washington in my Senate office here in Washington, D.C. that shows the various ecosystems of the State of Washington, which we very much appreciate. I bring that up because just recently the Seattle area was named the number one STEM economy in the United States. That is the largest STEM workforce in the United States of America. So, this issue about how we harness the information age to work for us and not against us is something that we deal with every day of the week, and we want to have discussion and discourse. I believe that discussion and discourse today should be broader than just 230.

There are issues of privacy that our committee has addressed and issues of how to make sure there is a free and competitive news market. I noticed today we are not calling in the NAB or the Publishers Association, asking them why they haven't printed or reprinted information that you allude to in your testimony, that you wish was more broadly distributed. To have competition in the news market is to have a diversity of voices and a diversity of opinion. And in my report just recently released, we show that true competition really does help perfect information both for our economy and for the health of our democracy. So, I do look forward to discussing these issues today. I do not want today's hearing to have a chilling effect on the very important aspects of ensuring that hate speech or misinformation related to health and public safety are not allowed to remain on the Internet. We all know what happened in 2016, and we had reports from the FBI, our intelligence agencies, and a bipartisan Senate committee that concluded in 2016 that Russian operatives did, masquerading as Americans, use targeted advertisements, intentionally falsified news articles, self-generated content, and social media platform tools, to interact with and attempt to deceive tens of millions of social media users in the United States.

Director of National Intelligence, and former Republican Senator Dan Coats said, in July 2018, that "the warning lights are blinking red," that the digital infrastructure that serves our country is literally under attack. So, I take this issue very seriously, and I have for many years. As Special Counsel Mueller indicated, 12 Russian intelligence officers hacked the DNC and various information detailing phishing attacks into our state election boards, online personas, and stolen documents. So, when we had a subcommittee hearing and former Bush Homeland Security Director Michael Chertoff testified, I asked him point blank, because there were some of our colleagues who were saying that everybody participates in election interference, if election interference was something that we did encourage or should be encouraging. He responded that he agreed, "interfering with infrastructure or elections is completely off limits and unacceptable." That is why I believe that we should be working aggressively, internationally to sanction anybody that interferes in our elections.

So, I hope today that we will get a report from the witnesses on exactly what they have been doing to clamp down on election interference. I hope that they will tell us what kind of hate speech and misinformation that they have taken off the books. It is no secret that there are various state actors who are doing all they can to take a whack at democracy, to try to say that our way of Government, that our way of life, that our way of freedom of speech and information is somehow not as good as we have made it, despite being the beacon of democracy around the globe. I am not going to tolerate people continuing to whack at our election process, our vote by mail system, or the ability of tech platforms, security companies, or law enforcement entities and the collective community to speak against misinformation and hate speech.

We have to show that the United States of America stands behind our principles and that our principles also transfer to the responsibility of communication online. As my colleagues will note,

we have all been through this in the past. That is why you, Mr. Chairman, and I and Senators Rosen and Thune sponsored the HACKED Act to help increase the security and cyber security of our Nation and create a workforce that can fight against that. That is why I joined with Senators Van Hollen and Rubio on the DETER Act, in establishing sanctions against Russian election interference and continuing to make sure that we build the infrastructure of tomorrow. So, I know that some people think that these issues are out of sight and out of mind. I guarantee you they are not. There are actors who have been at this for a long time. They wanted to destabilize Eastern Europe, and we became the second act when they tried to destabilize our democracy here by sowing disinformation. I want to show them that we, in the United States, do have fair elections and that we do have a fair process.

We are going to be that beacon of democracy. So, I hope that as we talk about 230 today and we hear from the witnesses on the progress that they have made in making sure that disinformation is not allowed online, we will also consider ways to help build and strengthen that. As some witnesses are testifying today, we will consider what we can do on transparency, on reporting, and on analysis. And yes, I think you are going to hear a lot about algorithms today, and the kind of oversight that we all want, to make sure that we can continue to have a diversity of voices in the United States of America, both online and offline. I do want to say though, Mr. Chairman, I am concerned about the vertical nature of news and information. Today, I expect to ask the witnesses about the fact that I believe they create a choke point for local news. The local news media have lost 70 percent of their revenue over the last decade, and we have lost thousands and thousands of journalistic jobs that are important. It was even amazing to me that someone at a newspaper who was funded by a joint group of the Knight Foundation and probably Facebook interviewed me about the fact that the news media and broadcast are on such a decline because of loss of revenue as they have made the transition to the digital age.

Somehow, we have to come together to show that the diversity of voices that local news represents needs to be dealt with fairly when it comes to the advertising market, and that too much control in the advertising market puts a foot on their ability to continue to move forward and grow in the digital age. Just as other forms of media have made the transition, and are still making the transition, we want to have a very healthy and dynamic news media across the United States of America. So, I plan to ask the witnesses today about that. I wish we had time to go into depth on privacy and privacy issues. But, Mr. Chairman, you know, and so does Senator Thune and other colleagues of the Committee on my side, how important it is that we protect American consumers on privacy issues. But we are not done with this work.

There is much to do to bring consensus in the United States on this important issue, and I hope that if we do have time in the follow up to these questions, we can ask the witnesses about that today. But make no mistake, gentlemen, thank you for joining us, but this is probably one of many, many, many conversations that we will have about all of these issues. Let's harness the informa-

tion age as you are doing, but let's also make sure that consumers are fairly treated and that we are making it work for all of us to guarantee our privacy, our diversity of voices, and our democratic principles. And solidify the fact that we, the United States of America, stand for freedom of information and freedom of the press. Thank you.

The CHAIRMAN. Thank you, Senator Cantwell. And certainly you are correct that this will not be the last hearing with regard to this subject matter. And I also appreciate you mentioning your concerns, which I share, about local journalism. At this point, we are about to receive testimony from our witnesses. Before we begin that, let me remind members that today's hearing will provide Senators with a round of 7 minute questioning rather than the usual 5 minutes that we have done in the past. At 7 minutes, the gavel will, let's say that a few seconds after 7 minutes the gavel will go down. Even so, this hearing could last some 3 hours and 42 minutes at that rate. So this will be an extensive and lengthy hearing. Members are advised that we will adhere closely to that 7 minute limit and also shortly before noon at the request of one of our witnesses, we will take a short 10 minute break. With that, we welcome our panel of witnesses, thank them for their testimony, and ask them to give their opening statements, summarizing them in some 5 minutes. The entire statement will be added at this point in the record. And we will begin with Mr. Jack Dorsey of Twitter. Sir, are you here? Do you hear us and do we have contact with you?

Mr. DORSEY. Yes. Can you hear me?

The CHAIRMAN. Yes, yes. So thank you for being with us. And you are now recognized for five minutes, sir.

STATEMENT OF JACK DORSEY, CHIEF EXECUTIVE OFFICER, TWITTER, INC.

Mr. DORSEY. OK. Well thank you members of the Commerce committee for the opportunity to speak with the American people about Twitter and Section 230. My remarks will be brief so we can get to questions. Section 230 is the most important law protecting Internet speech, and removing Section 230 will remove speech from the Internet. Section 230 gave Internet services two important tools. The first, provides immunity from liability for users content.

The second, provides Good Samaritan protections for content moderation and removal even of constitutionally protected speech as long as it is done in good faith. That concept of good faith is what is being challenged by many of you today. Some of you don't trust we are acting in good faith. That is the problem I want to focus on solving. Other services like Twitter earn your trust. How do we ensure more choice in the market if we don't? There are three solutions we would like to propose to address the concerns raised, all focused on services that decide to moderate or remove content. It could be expansions to Section 230, new legislative frameworks, or commitment to industry wide self-regulation best practices.

The first is requiring a services moderation process to be published. How are cases reported and reviewed? How are decisions made? What tools are used to enforce? Publishing answers to ques-

tions like these will make our process more robust and accountable to the people we serve. The second is requiring a straightforward process to appeal decisions made by humans or by algorithms. This ensures people can let us know when we don't get it right so we can fix any mistakes and make our processes better in the future. And finally, much of the content people see today is determined by algorithms with very little visibility into how they choose what they show.

We took a first step in making this more transparent by building a button to turn off our home timeline algorithms. It is a good start, but we are inspired by the market approach suggested by Dr. Stephen Wolfram before this committee in June 2019, enabling people to choose algorithms created by third parties to rank and filter the content is an incredibly energizing idea that is in reach. Requiring one, moderation process and practices to be published, two, a straightforward process to appeal decisions, and three, best efforts around algorithmic choice, are suggestions to address the concerns we all have going forward.

And they are all achievable in short order. It is critical as we consider these solutions, we optimize for new startups and independent developers. Doing so ensures a level playing field that increases the probability of competing ideas to help solve problems. We mustn't entrench the largest companies any further. Thank you for the time and I look forward to a productive discussion to dig into these and other ideas.

[The prepared statement of Mr. Dorsey follows:]

PREPARED STATEMENT OF JACK DORSEY, CHIEF EXECUTIVE OFFICER, TWITTER, INC.

Chairman Wicker, Ranking Member Cantwell, and Members of the Committee: Thank you for the opportunity to appear before the Committee and speak with the American people. Section 230 is the Internet's most important law for free speech and safety. Weakening Section 230 protections will remove critical speech from the Internet.

Twitter's purpose is to serve the public conversation. People from around the world come together on Twitter in an open and free exchange of ideas. We want to make sure conversations on Twitter are healthy and that people feel safe to express their points of view. We do our work recognizing that free speech and safety are interconnected and can sometimes be at odds. We must ensure that all voices can be heard, and we continue to make improvements to our service so that everyone feels safe participating in the public conversation—whether they are speaking or simply listening. The protections offered by Section 230 help us achieve this important objective.

As we consider developing new legislative frameworks, or committing to self-regulation models for content moderation, we should remember that Section 230 has enabled new companies—small ones seeded with an idea—to build and compete with established companies globally. Eroding the foundation of Section 230 could collapse how we communicate on the Internet, leaving only a small number of giant and well-funded technology companies.

We should also be mindful that undermining Section 230 will result in far more removal of online speech and impose severe limitations on our collective ability to address harmful content and protect people online. I do not think anyone in this room or the American people want less free speech or more abuse and harassment online. Instead, what I hear from people is that they want to be able to trust the services they are using.

I want to focus on solving the problem of how services like Twitter earn trust. And I also want to discuss how we ensure more choice in the market if we do not. During my testimony, I want to share our approach to earn trust with people who use Twitter. We believe these principles can be applied broadly to our industry and build upon the foundational framework of Section 230 for how to moderate content online. We seek to earn trust in four critical ways: (1) transparency, (2) fair proc-

esses, (3) empowering algorithmic choice, and (4) protecting the privacy of the people who use our service. My testimony today will explain our approach to these principles.

I. ENSURING GREATER TRANSPARENCY

We believe increased transparency is the foundation to promote healthy public conversation on Twitter and to earn trust. It is critical that people understand our processes and that we are transparent about what happens as a result. Content moderation rules and their potential effects, as well as the process used to enforce those rules, should be simply explained and understandable by anyone. We believe that companies like Twitter should publish their moderation process. We should be transparent about how cases are reported and reviewed, how decisions are made, and what tools are used to enforce. Publishing answers to questions like these will make our process more robust and accountable to the people we serve.

At Twitter, we use a combination of machine learning and humans to review reports and determine whether they violate the Twitter Rules. We take a behavior-first approach, meaning we look at how accounts behave before we review the content they are posting. Twitter's open nature means our enforcement actions are plainly visible to the public, even when we cannot reveal the private details of individual accounts that have violated our Rules. We have worked to build better in-app notices where we have removed Tweets for breaking our Rules. We also communicate with both the account that reports a Tweet and the account that posted it with additional detail on our actions. That said, we know we can continue to improve to further earn the trust of the people using Twitter.

In addition, regular reporting of outcomes in aggregate would help us all to increase accountability. We do this currently through the Twitter Transparency Center. This site provides aggregate content moderation data and other information for the individuals who use Twitter, academics, researchers, civil society groups, and others who study what we do to understand bigger societal issues. We believe it is now more important than ever to be transparent about our practices.

II. ADVANCING PROCEDURAL FAIRNESS

As a company, Twitter is focused on advancing the principle of procedural fairness in our decision-making. We strive to give people an easy way to appeal decisions we make that they think are not right. Mistakes in enforcement, made either by a human or algorithm, are inevitable, and why we strive to make appeals easier. We believe that all companies should be required to provide a straightforward process to appeal decisions made by humans or algorithms. This makes certain people can let us know when we do not get it right, so that we can fix any mistakes and make our processes better in the future.

Procedural fairness at Twitter also means we ensure that all decisions are made without using political viewpoints, party affiliation, or political ideology, whether related to automatically ranking content on our service or how we develop or enforce the Twitter Rules. Our Twitter Rules are not based on ideology or a particular set of beliefs. We believe strongly in being impartial, and we strive to enforce our Twitter Rules fairly.

III. EMPOWERING ALGORITHMIC CHOICE

We believe that people should have choices about the key algorithms that affect their experience online. At Twitter, we want to provide a useful, relevant experience to all people using our service. With hundreds of millions of Tweets every day on Twitter, we have invested heavily in building systems that organize content to show individuals the most relevant information for that individual first. With 186 million people last quarter using Twitter each day in dozens of languages and countless cultural contexts, we rely upon machine learning algorithms to help us organize content by relevance.

In December 2018, Twitter introduced an icon located at the top of everyone's timelines that allows individuals using Twitter to easily switch to a reverse chronological order ranking of the Tweets from accounts or topics they follow. This improvement gives people more control over the content they see, and it also provides greater transparency into how our algorithms affect what they see. It is a good start. We believe this points to an exciting, market-driven approach where people can choose what algorithms filter their content so they can have the experience they want. We are inspired by the approach suggested by Dr. Stephen Wolfram, Founder and Chief Executive Officer of Wolfram Research, in his testimony before the Subcommittee on Communications, Technology, Innovation, and the Internet in June 2019. Enabling people to choose algorithms created by third parties to rank and filter their content is an incredibly energizing idea that is in reach.

We also recognize that we can do even more to improve our efforts to provide greater algorithmic transparency and fair machine learning. The machine learning teams at Twitter are studying these techniques and developing a roadmap to ensure our present and future machine learning models uphold a high standard when it comes to algorithmic transparency and fairness. We believe this is an important step in ensuring fairness in how we operate and we also know that it is critical that we be more transparent about our efforts in this space.

IV. PROTECTING THE PRIVACY OF PEOPLE ON TWITTER

In addition to the principles I have outlined to address content moderation issues in order to better serve consumers, it is also critical to protect the privacy of the people who use online services. At Twitter, we believe privacy is a fundamental human right, not a privilege. We offer a range of ways for people to control their privacy experience on Twitter, from offering pseudonymous accounts to letting people control who sees their Tweets to providing a wide array of granular privacy controls. Our privacy efforts have enabled people around the world using Twitter to protect their own data.

That same philosophy guides how we work to protect the data people share with Twitter. Twitter empowers the people who use our service to make informed decisions about the data they share with us. We believe individuals should know, and have meaningful control over, what data is being collected about them, how it is used, and when it is shared.

Twitter is always working to improve transparency into what data is collected and how it is used. We believe that individuals should control the personal data that is shared with companies and provide them with the tools to help them control their data. Through the account settings on Twitter, we give people the ability to make a variety of choices about their data privacy, including limiting the data we collect, determining whether they see interest-based advertising, and controlling how we personalize their experience. In addition, we provide them the ability to access information about advertisers that have included them in tailored audiences to serve them ads, demographic and interest data about their account from ad partners, and information Twitter has inferred about them.

* * *

As you consider next steps, we urge your thoughtfulness and restraint when it comes to broad regulatory solutions to address content moderation issues. We must optimize for new startups and independent developers. In some circumstances, sweeping regulations can further entrench companies that have large market shares and can easily afford to scale up additional resources to comply. We are sensitive to these types of competition concerns because Twitter does not have the same breadth of interwoven products or market size as compared to our industry peers. We want to ensure that new and small companies, like we were in 2006, can still succeed today. Doing so ensures a level playing field that increases the probability of competing ideas to help solve problems going forward. We must not entrench the largest companies further.

I believe the best way to address our mutually-held concerns is to require the publication of moderation processes and practices, a straightforward process to appeal decisions, and best efforts around algorithmic choice. These are achievable in short order. We also encourage Congress to enact a robust Federal privacy framework that protects consumers while fostering competition and innovation.

We seek to earn trust from the people who use our service every day, and I hope the principles I describe and my responses to your questions can better inform your efforts. Thank you for the opportunity to appear. We look forward to continuing this dialogue with the Committee.

The CHAIRMAN. Thank you very much, Mr. Dorsey. We now call on Mr. Sundar Pichai. You are recognized for five minutes, sir.

STATEMENT OF SUNDAR PICHAI, CHIEF EXECUTIVE OFFICER, ALPHABET INC.

Mr. PICHAI. Chairman Wicker, Ranking Member Cantwell, and distinguished members of the Committee, thank you for the opportunity to appear before you today. The Internet has been a powerful force for good for the past three decades, has radically improved access to information, whether it is connecting Americans to jobs,

getting critical updates to people in times of crisis, or helping a parent find answers to questions like how can I get my baby to sleep through the night? At the same time, people everywhere can use their voices to share new perspectives, express themselves, and reach broader audiences than ever before.

Whether you are a barber in Mississippi or a home renovator in Indiana, you can share a video and build a global fanbase and a successful business right from your living room. In this way, the Internet has been one of the world's most important equalizers: information can be shared and knowledge can flow from anyone anywhere. The same low barriers to entry also make it possible for bad actors to cause harm. As a company whose mission is to organize the world's information and make it universally accessible and useful, Google is deeply conscious of both the opportunities and risks the Internet creates. I am proud that Google's information services like search, Gmail, maps, and photos provide thousands of dollars a year in value to the average American for free. We feel a deep responsibility to keep the people who use our products safe and secure and have long invested in innovative tools to prevent abuse of our services. When it comes to privacy, we are committed to keeping your information safe, treating it responsibly, and putting you in control.

We continue to make privacy improvements like the changes I announced earlier this year to keep less data by default and support the creation of comprehensive Federal privacy laws. We are equally committed to protecting the quality and integrity of information on our platforms and supporting our democracy in a nonpartisan way. That is just one timely example. Our information panels on Google and YouTube inform users about right to vote and how to register. We have also taken many steps to raise up high quality journalism, from sending 24 billion visits to news websites globally every month, to our recent $1 billion investment in partnerships with news publishers. Since our founding, we have been deeply committed to the freedom of expression.

We also feel a responsibility to protect people who use our products from harmful content and to be transparent about how we do that. That is why we said and publicly disclose clear guidelines for our products and platforms which we enforce impartially. We recognize that people come to our services for the broad spectrum of perspectives, and we are dedicated to building products that are helpful to users of all backgrounds and viewpoints. Let me be clear, we approach our work without political bias, full stop. To do otherwise would be contrary to both our business interests and our mission, which compels us to make information accessible to every type of person, no matter where they live or what they believe. Of course, our ability to provide access to a wide range of information is only possible because of existing legal frameworks like Section 230.

The United States adopted Section 230 early in the Internet's history, and it has been foundational to U.S. leadership in the tech sector. It protects the freedom to create and share content while supporting the ability of platforms and services of all sizes to responsibly address harmful content. We appreciate that this com-

mittee has put great thought into how platforms should address content. We look forward to having these conversations.

As you think about how to shape policy in this important area, I would urge the Committee to be very thoughtful about any changes to Section 230, and to be very aware of the consequences those changes might have on businesses and customers. At the end of the day, we all share the same goal, free access to information for everyone and responsible productions for people and their data. We support legal frameworks that achieve these goals. I look forward to engaging with you today about these important issues and answering your questions. Thank you.

[The prepared statement of Mr. Pichai follows:]

PREPARED STATEMENT OF SUNDAR PICHAI, CHIEF EXECUTIVE OFFICER, ALPHABET INC.

Chairman Wicker, Ranking Member Cantwell, and distinguished members of the Committee, thank you for the opportunity to appear before you today.

The Internet has been a powerful force for good over the past three decades. It has radically improved access to information, whether it's connecting Americans to jobs, getting critical updates to people in times of crisis, or helping a parent find answers to questions like "How can I get my baby to sleep through the night?"

At the same time, people everywhere can use their voices to share new perspectives, express themselves and reach broader audiences than ever before. Whether you're a barber in Mississippi or a home renovator in Indiana, you can share a video and build a global fanbase—and a successful business—right from your living room.

In this way, the Internet has been one of the world's most important equalizers. Information can be shared—and knowledge can flow—from anyone, to anywhere. But the same low barriers to entry also make it possible for bad actors to cause harm.

As a company whose mission is to organize the world's information and make it universally accessible and useful, Google is deeply conscious of both the opportunities and risks the Internet creates.

I'm proud that Google's information services like Search, Gmail, Maps, and Photos provide thousands of dollars a year in value to the average American—for free. We feel a deep responsibility to keep the people who use our products safe and secure, and have long invested in innovative tools to prevent abuse of our services.

When it comes to privacy we are committed to keeping your information safe, treating it responsibly, and putting you in control. We continue to make privacy improvements—like the changes I announced earlier this year to keep less data by default—and support the creation of comprehensive Federal privacy laws.

We are equally committed to protecting the quality and integrity of information on our platforms, and supporting our democracy in a non-partisan way.

As just one timely example, our information panels on Google and YouTube inform users about where to vote and how to register. We've also taken many steps to raise up high-quality journalism, from sending 24 billion visits to news websites globally every month, to our recent $1 billion investment in partnerships with news publishers.

Since our founding, we have been deeply committed to the freedom of expression. We also feel a responsibility to protect people who use our products from harmful content and to be transparent about how we do that. That's why we set and publicly disclose clear guidelines for our products and platforms, which we enforce impartially.

We recognize that people come to our services with a broad spectrum of perspectives, and we are dedicated to building products that are helpful to users of all backgrounds and viewpoints.

Let me be clear: We approach our work without political bias, full stop. To do otherwise would be contrary to both our business interests and our mission, which compels us to make information accessible to every type of person, no matter where they live or what they believe.

Of course, our ability to provide access to a wide range of information is only possible because of existing legal frameworks, like Section 230. The United States adopted Section 230 early in the internet's history, and it has been foundational to U.S. leadership in the tech sector. Section 230 protects the freedom to create and

share content while supporting the ability of platforms and services of all sizes to responsibly address harmful content.

We appreciate that this Committee has put great thought into how platforms should address content, and we look forward to having these conversations.

As you think about how to shape policy in this important area, I would urge the Committee to be very thoughtful about any changes to Section 230 and to be very aware of the consequences those changes might have on businesses and consumers.

At the end of the day, we all share the same goal: free access to information for everyone and responsible protections for people and their data. We support legal frameworks that achieve these goals, and I look forward to engaging with you today about these important issues, and answering your questions.

The CHAIRMAN. Thank you very much, Mr. Pichai. Members should be advised at this point that we are unable to make contact with Mr. Mark Zuckerberg. We are told by Facebook staff that he is alone and attempting to connect with this hearing, and that they are requesting a five-minute recess at this point to see if that connection can be made. I think this is a most interesting development. But we are going to accommodate the request of the Facebook employees and see if within five minutes we can make contact and proceed. So at this point, declare a five-minute recess.

[Recess.]

The CHAIRMAN. Call the hearing back into order, and we are told it that in less than 5 minutes we have success. So, Mr. Zuckerberg, I am told that we have both video and audio connection. Are you there, sir?

Mr. ZUCKERBERG. Yes, I am. Can you hear me?

The CHAIRMAN. Can hear you fine. And you are now recognized for five minutes to summarize your testimony. Welcome.

Mr. ZUCKERBERG. All right. Thank you, Chairman. I was able to hear the other opening statements. I was just having a hard time connecting myself. All right, so——

The CHAIRMAN. I know the feeling. Mr. Zuckerberg.

STATEMENT OF MARK ZUCKERBERG, CHIEF EXECUTIVE OFFICER, FACEBOOK, INC.

Mr. ZUCKERBERG. Chairman Wicker, Ranking Member Cantwell and members of the Committee, every day millions of Americans use the Internet to share their experiences and discuss issues that matter to them. Setting the rules for online discourse is an important challenge for our society, and there are principles at stake that go beyond any one platform. How do we balance free expression and safety? How do we define what is dangerous? Who should decide? I don't believe that private companies should be making so many decisions about these issues by themselves. And at Facebook, we often have to balance competing equities.

Sometimes the best approach from a safety or security perspective isn't the best for privacy or free expression. So we work with experts across society to strike the right balance. We don't always get it right, but we try to be fair and consistent. The reality is that people have very different ideas and views about where the line should be. Democrats often say that we don't remove enough content. And Republicans often say we remove too much. I expect that we will hear some of those criticisms today. And the fact that both sides criticize us doesn't mean that we are getting this right. But it does mean that there are real disagreements about where the

limits of online speech should be. And I think that is understandable.

People can reasonably disagree about where to draw the lines. That is a hallmark of democratic societies, especially here in the U.S. with our strong First Amendment tradition. But it strengthens my belief that when a private company is making these calls, we need a more accountable process that people feel is legitimate and that gives platform certainty. At Facebook, we publish our standards and issue quarterly reports on the content that we take down. We launch an independent oversight board that can overturn our decisions, and we have committed to an audit of our content reports. But I believe Congress has a role to play too in order to give people confidence that the process is carried out in a way that balances society's deeply held values appropriately. And that is why I have called for regulation.

Right now the discussion is focused on Section 230. Some say that ending 230 would solve all of the Internet's problems. Others say that would end the Internet as we know it. From our perspective, Section 230 does two basic things. First, it encourages free expression, which is fundamentally important. Without 230, platforms could potentially be held liable for everything that people say. They would face much greater pressure to take down more content to avoid legal risk. Second, it allows platforms to moderate content. Without 230, platforms could face liability for basic moderation, like removing harassment that impacts the safety of their communities. Now there is a reason why America leads in technology.

Section 230 helped create the Internet as we know it. It has helped new ideas get built and our companies to spread American values around the world and we should maintain this advantage. But the Internet has also evolved. And I think that Congress should update the law to make sure that it is working as intended. One important place to start would be making content moderation systems more transparent. Another would be to separate good actors from bad actors by making sure that companies can't hide behind Section 230 to avoid responsibility for intentionally facilitating illegal activity on their platforms. We are open to working with Congress on these ideas and more. I hope the changes that you make will ring true to the spirit and intent of 230.

There are consequential choices to make here, and it is important that we don't prevent the next generation of ideas from being built. Now, although this hearing is about content policy, I also want to cover our election preparedness work. Voting ends in 6 days. We are in the midst of a pandemic, and there are ongoing threats to the integrity of this election. Since 2016, Facebook has made major investments to stop foreign interference. We have hired more than 35,000 people to work on safety and security. We have disrupted more than 100 networks coming from Russia, Iran and China and more. They were misleading people about who they are and what they are doing, including three just this week. This is an extraordinary election and we have updated our policies to reflect that. We are showing people reliable information about voting and results, and we have strengthened our ads and misinformation policies.

We are also running the largest voting information campaign in U.S. history. We estimate that we have helped more than 4.4 million people register to vote and 100,000 people volunteer to be poll workers. Candidates on both sides continue to use our platforms to reach voters. People are rightly focused on the role that technology companies play in our elections. I am proud of the work that we have done to support our democracy. This is a difficult period, but I believe that America will emerge stronger than ever, and we are focused on doing our part to help.

[The prepared statement of Mr. Zuckerberg follows:]

PREPARED STATEMENT OF MARK ZUCKERBERG, CHIEF EXECUTIVE OFFICER, FACEBOOK, INC.

I. Introduction

Chairman Wicker, Ranking Member Cantwell, and members of the Committee, thank you for the opportunity to be here today.

Facebook's mission is to give people the power to build community and bring the world closer together. Our products enable more than 3 billion people around the world to share ideas, offer support, and discuss important issues. We know we have a responsibility to make sure people using our products can do so safely, and we work hard to set and enforce policies that meet this goal.

II. CDA Section 230s Role in Giving People a Voice and Keeping Them Safe

Section 230 of the Communications Decency Act is a foundational law that allows us to provide our products and services to users. At a high level, Section 230 does two things:

- First, it encourages free expression. Without Section 230, platforms could potentially be held liable for everything people say. Platforms would likely censor more content to avoid legal risk and would be less likely to invest in technologies that enable people to express themselves in new ways.
- Second, it allows platforms to moderate content. Without Section 230, platforms could face liability for doing even basic moderation, such as removing hate speech and harassment that impacts the safety and security of their communities.

Thanks to Section 230, people have the freedom to use the Internet to express themselves. At Facebook, this is one of our core principles. We believe in giving people a voice, even when that means defending the rights of people we disagree with. Free expression is central to how we move forward together as a society. We've seen this in the fight for democracy around the world, and in movements like Black Lives Matter and #MeToo. Section 230 allows us to empower people to engage on important issues like these—and to provide a space where non-profits, religious groups, news organizations, and businesses of all sizes can reach people.

Section 230 also allows us to work to keep people safe. Facebook was built to enable people to express themselves and share, but we know that some people use their voice to cause harm by trying to organize violence, undermine elections, or otherwise hurt people. We have a responsibility to address these risks, and Section 230 enables us to do this more effectively by removing the threat of constant litigation we might otherwise face.

We want Facebook to be a platform for ideas of all kinds, but there are specific types of harmful content that we don't allow. We publish our content policies in our Community Standards, and we update them regularly to address emerging threats. To address each type of harmful content, we've built specific systems that combine sophisticated technology and human judgment. These systems enabled us to take down over 250 million pieces of content that violated our policies on Facebook and Instagram in the first half of 2020, including almost 25 million pieces of content relating to terrorism and organized hate, almost 20 million pieces of content involving child nudity or sexual exploitation, and about 8.5 million pieces of content identified as bullying or harassment. We report these numbers as part of our Transparency Reports, and we believe all other major platforms should do the same so that we can better understand the full picture of online harms.

However, the debate about Section 230 shows that people of all political persuasions are unhappy with the status quo. People want to know that companies are taking responsibility for combatting harmful content—especially illegal activity—on

their platforms. They want to know that when platforms remove content, they are doing so fairly and transparently. And they want to make sure that platforms are held accountable.

Section 230 made it possible for every major Internet service to be built and ensured important values like free expression and openness were part of how platforms operate. Changing it is a significant decision. However, I believe Congress should update the law to make sure it's working as intended. We support the ideas around transparency and industry collaboration that are being discussed in some of the current bipartisan proposals, and I look forward to a meaningful dialogue about how we might update the law to deal with the problems we face today.

At Facebook, we don't think tech companies should be making so many decisions about these important issues alone. I believe we need a more active role for governments and regulators, which is why in March last year I called for regulation on harmful content, privacy, elections, and data portability. We stand ready to work with Congress on what regulation could look like in these areas. By updating the rules for the internet, we can preserve what's best about it—the freedom for people to express themselves and for entrepreneurs to build new things—while also protecting society from broader harms. I would encourage this Committee and other stakeholders to make sure that any changes do not have unintended consequences that stifle expression or impede innovation.

III. Preparing for the 2020 Election and Beyond

The issues of expression and safety are timely as we are days away from a presidential election in the midst of a pandemic. With COVID–19 affecting communities around the country, people will face unusual challenges when voting. Facebook is committed to doing our part to help ensure everyone has the chance to make their voice heard. That means helping people register and vote, clearing up confusion about how this election will work, and taking steps to reduce the chances of election-related violence and unrest.

This election season, Facebook has run the largest voting information campaign in American history. Based on conversion rates we calculated from a few states we partnered with, we've helped an estimated 4.4 million people register to vote across Facebook, Instagram, and Messenger. We launched a Voting Information Center to connect people with reliable information on deadlines for registering and voting and details about how to vote by mail or vote early in person, and we're displaying links to the Voting Information Center when people post about voting on Facebook. We've directed more than 39 million people so far to the Voting Information Center, and we estimate we've helped about 100,000 people sign up to be poll workers.

We're also working to tackle misinformation and voter suppression. We've displayed warnings on more than 150 million pieces of content that have been debunked by our third-party fact-checkers. We're partnering with election officials to remove false claims about polling conditions, and we've put in place strong voter suppression policies that prohibit explicit or implicit misrepresentations about how or when to vote, as well as attempts to use threats related to COVID–19 to scare people into not voting. We're removing calls for people to engage in poll watching that use militarized language or suggest that the goal is to intimidate, exert control, or display power over election officials or voters. In addition, we're blocking new political and issue ads during the final week of the campaign, as well as all political and issue ads after the polls close on election night.

Since many people will be voting by mail, and since some states may still be counting valid ballots after election day, many experts are predicting that we may not have a final result on election night. It's important that we prepare for this possibility in advance and understand that there could be a period of uncertainty as the final results are counted, so we've announced a variety of measures to help in the days and weeks after voting ends:

- We'll use the Voting Information Center to prepare people for the possibility that it may take a while to get official results. This information will help people understand that there is nothing illegitimate about not having a result on election night.
- We're partnering with Reuters and the National Election Pool to provide reliable information about election results. We'll show this in the Voting Information Center so it's easily accessible, and we'll notify people proactively as results become available. Importantly, if any candidate or campaign tries to declare victory before the results are in, we'll add a label to their post stating that official results are not yet in and directing people to the official results.
- We'll attach an informational label to content that seeks to delegitimize the outcome of the election or discuss the legitimacy of voting methods, for example,

by claiming that lawful methods of voting will lead to fraud. This label will provide basic reliable information about the integrity of the election and voting methods.

- We'll enforce our violence and harm policies more broadly by expanding our definition of high-risk targets to include election officials in order to help prevent any attempts to pressure or harm them, especially while they're fulfilling their critical obligations to oversee the vote counting.
- We've strengthened our enforcement against militias, conspiracy networks, and other groups that could be used to organize violence or civil unrest in the period after the election. We have already removed thousands of these groups from our platform, and we will continue to ramp up enforcement over the coming weeks.

It's important to recognize that there may be legitimate concerns about the electoral process over the coming months. We want to make sure people can speak up if they encounter problems at the polls or have been prevented from voting, but that doesn't extend to spreading misinformation.

Four years ago we encountered a new threat: coordinated online efforts by foreign governments and individuals to interfere in our elections. This threat hasn't gone away. We've invested heavily in our security systems and now have some of the most sophisticated teams and systems in the world to prevent these attacks, including the teams working in our dedicated Election Operations Center. Since 2017, we've removed more than 100 networks worldwide engaging in coordinated inauthentic behavior, including ahead of major democratic elections, and we've taken down 30 networks so far this year. We're also blocking ads from state-controlled media outlets in the U.S. to provide an extra layer of protection against various types of foreign influence in the public debate ahead of the election.

IV. Supporting a Healthy News Ecosystem

Facebook also supports our democracy by supporting journalism—particularly local journalism, which is vital for helping people be informed and engaged citizens. Facebook has invested hundreds of millions of dollars across a variety of initiatives to support a healthy news and journalism ecosystem. We launched Facebook News in October 2019, making a $300 million commitment to help publishers invest in building their readership and subscription models. We now have multi-year partnerships with ABC News, The New York Times, The Wall Street Journal, The Washington Post, BuzzFeed, Fox News, The Dallas Morning News, and many more.

Among other benefits, Facebook provides publishers with free organic distribution of news and other content, which grows audience and revenue for news publishers; customized tools and products to help publishers monetize their content; and initiatives to help them innovate with online news content. We've also built tools to help publishers increase their subscribers by driving people from Facebook links to publisher websites. Helping publishers reach new audiences has been one of our most important goals, and we have found that over 95 percent of the traffic Facebook News now delivers to publishers is in addition to the traffic they already get from News Feed.

The Facebook Journalism Project is another initiative to create stronger ties between Facebook and the news industry. Over the past three years, we've invested more than $425 million in this effort, including developing news products; providing grants, training, and tools for journalists; and working with publishers and educators to increase media literacy. Since launching the Facebook Journalism Project, we have met with more than 2,600 publishers around the world to understand how they use our products and how we can make improvements to better support their needs.

This investment includes support for organizations like the Pulitzer Center, Report for America, the Knight-Lenfest Local News Transformation Fund, the Local Media Association and Local Media Consortium, the American Journalism Project, and the Community News Project. We've seen how important it is that people have information they can rely on, and we're proud to support organizations like these that play a critical role in our democracy.

V. Conclusion

I'd like to close by thanking this Committee, and particularly Chairman Wicker and Ranking Member Cantwell, for your leadership on the issue of online privacy. Facebook has long supported a comprehensive Federal privacy law, and we have had many constructive conversations with you and your staffs as you have crafted your proposals. I understand that there are still difficult issues to be worked out, but I am optimistic that legislators from both parties, consumer advocates, and industry all agree on many of the fundamental pieces. I look forward to continuing

to work with you and other stakeholders to ensure that we provide consumers with the transparency, control, and accountability they deserve.

I know we will be judged by how we perform at this pivotal time, and we're going to continue doing everything we can to live up to the trust that people have placed in us by making our products a part of their lives.

The CHAIRMAN. Well, thank you. Thank you very much, Mr. Zuckerberg, and thanks to all of our witnesses. We will now—I think we are supposed to set the clock to 7 minutes and I see 5 minutes up there. But somehow we will keep time. So there we are. OK. Well, thank you all. Let me start then with Mr. Dorsey. Mr. Dorsey, the Committee has compiled dozens and dozens of examples of conservative content being censored and suppressed by your platforms over the last 4 years. I entered these examples into the record on October 1 when the Committee voted unanimously to issue the subpoenas. And thank you all three again for working with us on the scheduling, alleviating the necessity for actually exercising the subpoenas.

Mr. Dorsey, your platform allows foreign dictators to post propaganda, typically without restriction, yet you routinely restrict the President of the United States. And here is an example. In March, a spokesman for the Chinese Communist Party falsely accused the U.S. Military of causing the coronavirus epidemic. He tweeted, "CDC was caught on the spot. When did patient zero begin in the U.S.? How many people are infected? What are the names of the hospitals? It might be the U.S. Army who brought the epidemic to Wuhan." And on and on. After this tweet was up for some 2 months, Twitter added a fact check label to this tweet—after being up for 2 months.

However, when President Trump tweeted about how mail-in ballots are vulnerable to fraud, a statement that I subscribe to and agree with, and a statement that is in fact true, Twitter immediately imposed a fact check label on that tweet. Mr. Dorsey, how does a claim by Chinese communists that the U.S. Military is to blame for COVID remain up for 2 months without a fact check and the President's tweet about security in mail-in ballots get labeled instantly?

Mr. DORSEY. Well, first and foremost, we as you mentioned, we did label that tweet. As we think about enforcement, we consider severity of potential offline harm, and we act as quickly as we can. We have taken action against tweets from world leaders all around the world, including the President. And we did take action on that tweet because we saw it. We saw the confusion it might encourage and we labeled it accordingly. And the goal——

The CHAIRMAN. You are speaking of the President's tweet?

Mr. DORSEY. Yes.

The CHAIRMAN. OK.

Mr. DORSEY. The goal of our labeling is to provide more context, to connect the dots so that people can have information so they can make decisions for themselves. We, you know, we have created these policies recently. We are enforcing them. There are certainly things that we can do much faster. But generally, we believe that the policy was enforced in a timely manner and in the right regard.

The CHAIRMAN. And yet you seem to have no objection to a tweet by the Chinese Communist Party saying the U.S. Army brought the epidemic to Wuhan?

Mr. DORSEY. Well, we did, and we labeled that tweet——

The CHAIRMAN. Too much to do so, is that correct?

Mr. DORSEY. I am not sure of the exact timeframe, but we can get back to you on that.

The CHAIRMAN. So you are going to get back to us as to how a tweet from the Chinese Communist Party falsely accusing the U.S. Military of causing the coronavirus epidemic was left up for 2 months with no comment from Twitter while the President of the United States making a statement about being careful about voter—about ballot security with the mail was labeled immediately. I have a tweet here from Mr. Ajit Pai. Mr. Ajit Pai is the Chairman of the Federal Communications Commission. And he recounts some four tweets by the Iranian dictator, Ayatollah Ali Khamenei, which Twitter did not place a public label on. All four of them glorify violence. The first tweet says this, and I quote each time, "the Zionist regime is a deadly cancerous growth and a detriment to the region. It will undoubtedly be uprooted and destroyed." That is the first tweet.

The second tweet, "the only remedy until the removal of the Zionist regime is firm, armed resistance." Again, left up without comment by Twitter. The third, "the struggle to free Palestine is jihad in the way of God." I quote that in part for the sake of time. And number four, "we will support and assist any Nation or any group anywhere who opposes and fights the Zionist regime." I would simply point out that these tweets are still up, Mr. Dorsey. And how is it that they are acceptable to be there? I will ask unanimous consent to enter this tweet from Ajit Pai in the record at this point. That will be done without objection. How is, Mr. Dorsey, is that acceptable based on your policies at Twitter?

Mr. DORSEY. We believe it is important for everyone to hear from global leaders. And we have policies around world leaders. We want to make sure that we are respecting their right to speak and to publish what they need. But if there is a violation of our terms of service, we want to label it——

The CHAIRMAN. They are still up. Do they violate your terms of service, Mr. Dorsey?

Mr. DORSEY. We did not find those to violate our terms of service because we considered them saber rattling, which is part of the speech of world leaders in concert with other countries. Speech against our own people or a country's own citizens, we believe, is different and can cause more immediate harm.

The CHAIRMAN. Very telling information, Mr. Dorsey. Thank you very much. Senator Cantwell, you are recognized.

Senator CANTWELL. I think I am deferring to our colleague, Senator Peters, just because of the timing and situation for him.

The CHAIRMAN. All right, Senator Peters, are you there?

STATEMENT OF HON. GARY PETERS,
U.S. SENATOR FROM MICHIGAN

Senator PETERS. I am here. I am here.

The CHAIRMAN. You are recognized for seven minutes.

Senator PETERS. Well, thank you, Mr. Chairman and Ranking Member Cantwell. I appreciate your deferral to me. I certainly appreciate that consideration a great deal. I also want to thank each of our panelists here today for coming forward and being a witness. And I appreciate all of you accommodating your schedules so that we can have this hearing. My first question is for Mr. Zuckerberg, as—and I want to start off by saying how much I appreciated our opportunity last night to speak at length on a number of issues. And as I told you last night, I appreciate Facebook's efforts to assist the law enforcement to disrupt a plot to kidnap and hold a sham trial and kill our Governor, Governor Whitmer. The individuals in that case apparently used the Facebook for a broad recruiting effort. But then they actually planned the specifics of that operation off of your platform.

My question is, when users reach the level of radicalization that violates your community standards, you often will ban those groups and you will then drive them off to other platforms, those platforms tend to have less transparency and oversight. But the issue that I would like you to address is for those individuals that remain on your platform, they are often far down the path of radicalization, but they are definitely looking for an outlet.

And I understand that Facebook has recently adopted a strategy to redirect users who are searching, for example, for election misinformation. But it doesn't seem that that policy applies to budding violent extremists. Mr. Zuckerberg, do believe that the platform is—your platform has a responsibility to offramp users who are on the path to radicalization by violent extremist groups?

Mr. ZUCKERBERG. Senator, thanks for the question. I think this is a very important and my understanding is that we actually do a little of what you are talking about here. If people are searching for, and I think for example white supremacist organizations of which we ban those, we treat them as terrorist organizations, not only we are not going to show that content, but I think we try to where we can highlight information that would be helpful. And I think we try to work with experts on that. I can I can follow up and get you more information on the scope of those activities and when we invoke that. But I certainly agree with the spirit of the question that this is a good idea and something that we should continue pursuing and perhaps expand.

Senator PETERS. Well, I appreciate those comments. I am the Ranking Member on Senate Homeland Security committee, and what we are seeing is a rise of violent extremist groups, which is very troubling. And certainly we need to work very closely with you as to how do we disrupt this kind of radicalization especially from folks that are using your platform.

So I appreciate the opportunity to work further. And as we talked about last night, you asserted that Facebook is proactively working with law enforcement now to disrupt some of these real world violent attempts that stem from some of that activity that originated in your platform. But could you tell me specifically how many threats that you have proactively referred to, local or state law enforcement, prior to being approached for a preservation request?

Mr. ZUCKERBERG. Senator, I don't know the number off the top of my head, so I can follow up with you on that, but it is increasingly common that our systems are able to detect when there is potential issues. And over the last 4 years in particular, we have built closer partnerships with law enforcement and the intelligence community to be able to share those kind of signals. So what we are doing more of that, including in the case that you mentioned before, around the attempted kidnapping of Governor Witmer, we identified that as a signal to the FBI, I think was about 6 months ago when we started seeing some suspicious activity on our platform. And there certainly—that is part of our routine and how we operate.

Senator PETERS. Well, Mr. Zuckerberg, discovery tools and recommendation algorithms that your platforms use have served up potentially extremist content based on the user profiles of folks. As we seek to understand why membership in these extremist groups is rising, I would hope that your companies are right now engage again some forensic analysis of membership. Once you take down an extremist group, to take a look at how that happened on your platform is certainly going to better inform us as to how we can disrupt this type of recruitment into extremist groups. My question for you, though, is that in 2016, you said and this was apparently an internal Facebook, internal document that was reported by *The Wall Street Journal* that said that "64 percent of members of violent groups became members because of your platform's recommendation."

And I will quote from that report that was reported in *The Wall Street Journal* that said, "our recommendation systems grow the problem." That is clearly very concerning. And I know in response to that report in 2016, you had made changes to your policies. You made changes to some of the algorithms that existed at that time. Our question is, have you seen a reduction in your platform's facilitation of extremist group recruitment since those policies were changed?

Mr. ZUCKERBERG. Senator, I am not familiar with that specific study, but I agree with the concern and making sure that our recommendation systems for what groups people are given the opportunity to join is certainly one important vector for addressing this issue. And we have taken a number of steps here, including disqualifying groups from being included in our recommendation system at all, if they routinely are being used to share misinformation or if they have content violations or a number of other criteria. So I am quite focused on this. I agree with where you are going with that question. I don't have any data today on the real world impact of that yet. But I think that addressing this upstream is very important.

Senator PETERS. So I appreciate you agreeing with that and that we need more data. Is it that you don't have the data just at the top of your head or that it doesn't exist?

Mr. ZUCKERBERG. Well, Senator, certainly the former and then potentially the latter as well. I think it probably takes some time before—after we make these changes to be able to measure the impact of it. And I am not aware of what studies are going on into this. This is—this seems like the type of thing that one would want

not just internal Facebook researchers to work on, but also potentially a collaboration with independent academics as well.

The CHAIRMAN. Thank you, Mr. Zuckerberg. And thank you, Senator Peters.

Senator PETERS. Thank you.

The CHAIRMAN. Senator Gardner has also asked to go out of order and Senator Thune has graciously deferred to him. So, Senator Gardner, you are recognized for seven minutes, sir.

STATEMENT OF HON. CORY GARDNER,
U.S. SENATOR FROM COLORADO

Senator GARDNER. Well, thank you, Mr. Chairman. And thank you, Senator Thune, for sharing your time or at least deferring your time to me. And thank you, Mr. Zuckerberg. Thank you very much. And, Mr. Dorsey, thank you for being here. Mr. Dorsey, I am going to direct these first questions to you. Mr. Dorsey, do you believe that the Holocaust really happened? Yes or no?

Mr. DORSEY. Yes.

Senator GARDNER. So you would agree that someone who says the Holocaust may not have happened is spreading misinformation? Yes or no?

Mr. DORSEY. Yes.

Senator GARDNER. I appreciate your answers on this but they surprise me and probably a lot of other Coloradoans and Americans. After all, Iran's Ayatollah has done exactly this, questioning the Holocaust. And yet his tweets remain unflagged on Twitter's platform. You and I agree that moderating your platform makes sense in certain respects. We don't want the next terrorist finding inspiration on Twitter or any platform for that matter. But you have also decided to moderate certain content from influential world leaders. And I would like to understand your decisions to do so a little bit better. Can you name any other instance of Twitter hiding or deleting a tweet from heads of state?

Mr. DORSEY. Not none off the top of my head, but we have many examples across world leaders around the world.

Senator GARDNER. Would you be willing to provide a list of those?

Mr. DORSEY. Absolutely.

Senator GARDNER. I know we have established many free content moderation can have certain upsides like combating terrorism but Twitter has chosen to approach content moderation from the standpoint of combating misinformation as well. So it is strange to me that you have flagged the tweets from the President, but haven't hidden the Ayatollah's tweets on Holocaust denial or calls to wipe Israel off the map and that you can't recall off the top of your head hidden or deleted tweets from other world leaders. I would appreciate that list. I think it is important that we all hear that. So that brings my next question to the front. Does Twitter maintain a formal list of certain accounts that you actively monitor for misinformation?

Mr. DORSEY. No. And we don't have a policy against misinformation. We have a policy against misinformation in three categories which are manipulated media, public health, specifically COVID, and civic integrity, election interference and voter suppression.

That is all we have policy on for misleading information. We do not have policy or enforcement for any other types of misleading information that you are mentioning.

Senator GARDNER. So somebody denying the murder of millions of people or instigating violence against a country as a head of state does not categorically fall in any of those three misinformation or other categories perhaps?

Mr. DORSEY. Not misinformation. But we do have other policies around incitement to violence, which some of the tweets you mentioned of the examples that you are mentioning may follow, but for misleading information, we are focused on those three categories only.

Senator GARDNER. So somebody denies the Holocaust has happened, is not misinformation?

Mr. DORSEY. It is misleading information, but we don't have a policy against that type of misleading information——

Senator GARDNER. Millions of people died. And that is not a violation of Twitter—again, I just don't understand how you can label a President of the United States—have you ever taken a tweet down from the Ayatollah?

Mr. DORSEY. I believe we have but we can get back to you on that. We have certainly labeled tweets and I believe we have taken one down as well.

Senator GARDNER. You know, did you say you do not have a list, is that correct? Do you not maintain a list?

Mr. DORSEY. We don't maintain a list of accounts we watch. We look for reports and issues brought to us, and then we weigh it against our policy and then force if needed.

Senator GARDNER. You look for reports from your employees or from the—.

Mr. DORSEY. No, from the people using the service.

Senator GARDNER. Right. And then they turn that over to your Board of Review. Is that correct?

Mr. DORSEY. Well, so in some cases algorithms take action. In other cases humans do. In some cases it is a pairing of the two.

Senator GARDNER. There are numerous examples of blue checkmarks that are spreading false information that aren't flagged. So, and Twitter must have some kind of list of priority accounts that it maintains. You have the blue checkmark list. How do you decide when to flag a tweet that we just—just got into that a little bit? Is there a formal threshold of tweets or likes that must be met before a tweet is flagged?

Mr. DORSEY. No.

Senator GARDNER. Twitter can't claim that—I just with your answers on the Ayatollah and others, I just don't understand how Twitter can claim to want a world of less hate and misinformation while you simultaneously let the kind of content that the Ayatollah has tweeted out to flourish on the platform, including from other world leaders. I just—it is no wonder that Americans are concerned about politically motivated content moderation at Twitter given what we have just said. I don't like the idea of a group of unelected elites in San Francisco or Silicon Valley deciding whether my speech is permissible on their platforms, but I like even less the idea of unelected Washington, D.C. bureaucrats trying to enforce

some kind of politically neutral content moderation. So just as we have heard from other panelists, as we have we are going to hear throughout the day, we have to be very careful and not rush to legislate in ways that stifle speech. You can delete Facebook, turn off Twitter or try to ditch Google, but you cannot unsubscribe from Government censors. Congress should be focused on encouraging speech, not restricting it.

The Supreme Court has tried teaching us that lesson time and time again, and the Constitution demands that we remember it. I am running short on time. Something very quickly. I will go through another question. One of the core ideas of Section 230s liability protections is this, you shouldn't be responsible for what someone else says on your platform. Conversely, you should be liable for what you say or do on your own platform. I think that is pretty common sense. But courts have not always agreed with this approach.

Even Rep. Chris Cox opined in a recent *Wall Street Journal* op ed that Section 230 has sometimes been interpreted by courts more broadly than I expected, for example, allowing some websites to escape liability for content they helped create. Mr. Zuckerberg, I have a simple question for you and each of the panelists today. Quickly, to be clear, I am not talking about technical tools or operating the platform itself here. I am purely talking about content. Do you agree that Internet platforms should be held liable for the specific content that you yourself create on your own platforms, yes or no?

The CHAIRMAN. Very quickly.

Mr. ZUCKERBERG. Senator, I think that that is reasonable.

Senator GARDNER. Yes or no, Mr. Dorsey, if Twitter creates specific content, should Twitter be liable for that content?

Mr. DORSEY. Twitter does as well.

Senator GARDNER. Mr. Pichai, same question to you, yes or no, should Google be liable for the specific content that it creates?

Mr. PICHAI. If we are acting as a publisher, I would say yes.

Senator GARDNER. The specific content that you create on your own platform, yes.

Mr. PICHAI. That seems reasonable.

Senator GARDNER. Thank you. I think what the other side's liability questions in regard to the good faith removal provision in Section 230 and we will get into a little bit more on the private questions. I know I am out of time. So, Mr. Chairman, thank you for giving me this time. Senator Thune, thank you as well. Thanks to the witnesses.

The CHAIRMAN. Thank you, Senator Gardner. The Ranking Member has now deferred to Senator Klobuchar. So, Senator, you are now recognized.

STATEMENT OF HON. AMY KLOBUCHAR,
U.S. SENATOR FROM MINNESOTA

Senator KLOBUCHAR. Thank you, Chairman. I want to note first that this hearing comes 6 days before Election Day, and it makes, I believe, we are politicizing, and the Republican majority is politicizing, what should actually not be a partisan topic. And I do want to thank the witnesses here for appearing, but also for the work that they are doing to try to encourage voting and to put out that

correct information when the President and others are undermining vote by mail, something we are doing in every state in the country right now. Second point, Republicans failed to pass the bipartisan Honest Ads Act and the White House blatantly blocked the bipartisan election security bill that I have with Senator Lankford, as well as several other Republicans.

And it is one of the reasons I think we need a new President. Third, my Republican colleagues in the Senate, many of them I work with very well on this committee, but we have had 4 years to do something when it comes to antitrust, privacy, local news, a subject that briefly came up, and so many other things. So I am going to use my time to focus on what I consider in Justice Ginsburg's words to be a "blueprint for the future." I will start with you, Mr. Zuckerberg. How many people log into Facebook every day?

Mr. ZUCKERBERG. Senator, it is more than 2 billion.

Senator KLOBUCHAR. OK. And how much money have you made on political advertisements in the last two years?

Mr. ZUCKERBERG. Senator, I do not know off the top of my head. It is a relatively small part of our revenue.

Senator KLOBUCHAR. OK. Small for you, but I think it is $2.2 billion, over ten 10,000 ads sold since May 2018. Those are your numbers and we can check them out later. Do you require Facebook employees to review the content of each of the political ads that you sell in order to ensure that they comply with the law and your own internal rules?

Mr. ZUCKERBERG. Senator, we require all political advertisers to be verified before they could run ads. And I believe we do review advertising as well.

Senator KLOBUCHAR. But does a real person actually read the political ads that you are selling, yes or no?

Mr. ZUCKERBERG. Senator, I imagine that a person does not look at every single ad. Our systems are a combination of artificial intelligence systems and people. We have 35,000 people who do content and security review for us. But the massive amount——

Senator KLOBUCHAR. I really just had a straightforward question because I don't think they do. I think the algorithms hidden—because I think the ads instantly are placed. Is that correct?

Mr. ZUCKERBERG. Senator, my understanding of the way the system works is we have computers and artificial intelligence scan everything, and if we think that there are potential violations, then either the AI system will act or it will flag it to the tens of thousands of people who do content review.

Senator KLOBUCHAR. With all the money you have, you could have a real person review like a lot of the other traditional media organizations do. So another question, when John McCain and I and Senator Warner introduced the Honest Ads Act, we got pushback from your company, others, and you were initially against it. But then we discussed this at a hearing. You are for it. I appreciate that. And have you spent any of the money? I know you spent the most money, Facebook spent the most money over lobbying last year. Have you spent any of the money trying to change or block the bill?

Mr. ZUCKERBERG. Senator, no. In fact, I have endorsed it publicly and we have implemented it into our systems, even though it hasn't become law. I am a big supporter——

Senator KLOBUCHAR.—tried to change it. No, have you done anything to get it passed because we are at a roadblock on it? And I do appreciate that you voluntarily implemented some of it, but have you voluntarily implemented the part of the Honest Ads Act where you fully disclose which groups of people are being targeted by political ads?

Mr. ZUCKERBERG. Senator, we have, I think, industry leading transparency around political ads, and part of that is showing which audiences in broad terms ended up seeing the ads. Of course, getting the right resolution on that is challenging without it becoming a privacy issue. But we have tried to do that and provide as much transparency as we can. I think we are currently leading in that area. And to your question about how we—.

Senator KLOBUCHAR. I still have concerns, and I don't mean to interrupt you, but I have such limited time. One of the things that I—last thing I want to ask you about is divisiveness on the platform. And I know there has been a recent—studies have shown that part of your algorithms that push people toward more polarized content, left, right, whatever. In fact, one of your researchers warned senior executives that our algorithms exploit the human brains attraction to divisiveness. The way I look at it, more divisiveness, more time in the platform, more time on the platform, the company makes more money. Does that bother you what it has done to our politics?

Mr. ZUCKERBERG. Senator, I respectfully disagree with that characterization of how the systems work. We design our systems to show people the content that is going to be the most meaningful to them, which is not trying to be as divisive as possible. Most of the content on the systems is not political. It is things like making sure that you can see when your cousin had her baby or——

Senator KLOBUCHAR. OK. I am going to move on to a quick Google here and Mr. Pichai, but I am telling you right now that that is not what I am talking about, the cousins and the babies here. I am talking about conspiracy theories and all the things that I think the Senators on both sides of the aisle know what I am talking about. And I think it has been corrosive. Google, Mr. Pichai, I have not really liked your response to the lawsuit and what has been happening. I think we need a change in competition policy for this country. I hope, I believe to ask you more about it at the Judiciary committee. And I think your response isn't just defensive, it has been defiant to the Justice Department and suits all over the world. You control almost 90 percent of all general search engine queries, 70 percent of the search advertising market. Don't you see these practices as anti-competitive?

Mr. PICHAI. Well, Senator, we are a popular general purpose search engine. We do see robust competition, many categories of information, and, you know, we invest significantly in R&D. We are innovating. We are lowering prices in all the markets we are operating in and happy to, you know, engage and discuss it further.

Senator KLOBUCHAR. Well, one of your employees testified before the antitrust subcommittee last month, and he suggested that

Google wasn't dominant in ad tech, that it was only one of many companies in a highly competitive ad tech landscape. Google has 90 percent of the publisher ad server market, a product of its double click acquisition. Does the market sound highly competitive to you when you have 90 percent of it?

The CHAIRMAN. Very brief—very brief answer.

Mr. PICHAI. Many publishers can use simultaneously made tools. Amazon created this alone of growing significantly in the last 2 years. You know, we, this is a market in which we share a majority of our revenue. Our margins are low. We are happy to take feedback here. We are trying to support the publishing industry, but definitely open to feedback and happy to engage in this.

The CHAIRMAN. Thank you. Thank you, Mr. Pichai. Thank you, Senator Klobuchar.

Senator KLOBUCHAR.—so I am looking forward to our next hearing to discuss it more. Thank you.

The CHAIRMAN. Thank you very much. Senator Thune, you are now recognized.

STATEMENT OF HON. JOHN THUNE,
U.S. SENATOR FROM SOUTH DAKOTA

Senator THUNE. Thank you, Mr. Chairman. And I appreciate you convening the hearing today, which is an important follow up to the Subcommittee hearing that we convened in July on Section 230. Many of us here today and many of those we represent are deeply concerned about the possibility of political bias and discrimination by large Internet social media platforms. Others are concerned that even if your actions aren't skewed, that they are hugely consequential for our public debate. Yet you operate with limited accountability. Such distrust is intensified by the fact that the moderation practices used to suppress or amplify content remain largely a black box to the public. Moreover, the public explanations given by the platforms for taking down or suppressing content too often seem like excuses that have to be walked back after scrutiny. And due to exceptional secrecy with which platforms protect their algorithms and content moderation practices, it has been impossible to prove one way or another whether political bias exists, so users are stuck with anecdotal information that frequently seems to confirm their worst fears. Which is why I have introduced two bipartisan bills the Platform Accountability and Consumer Transparency, or the PACT Act, and the Filter Bubble Transparency Act to give users, the regulators, and the general public meaningful insight into online content moderation decisions and how algorithms may be amplifying or suppressing information. And so I look forward to continuing that discussion today. My Democrat colleagues suggest that when we criticize the bias against conservatives, that we are somehow working the refs. But the analogy of working the refs assumes that it is legitimate even to think of you as refs. It assumes that you three Silicon Valley CEOs get to decide what political speech gets amplified or suppressed. And it assumes that you are the arbiters of truth or at the very least the publishers making editorial decisions about speech. So yes or no, I would ask this of each of the three of you, are the Democrats correct that you

all are the legitimate referees over our political speech? Mr. Zuckerberg, are you the ref?

Mr. ZUCKERBERG. Senator, I certainly think not. And I do not want us to have that role.

Senator THUNE. Mr. Dorsey, are you the ref?

Mr. DORSEY. No.

Senator THUNE. Mr. Pichai, are you the ref?

Mr. PICHAI. Senator, I do think we make content moderation decisions, but we are transparent about it and we do it to protect users. But we really believe and support maximizing freedom of expression.

Senator THUNE. I will take that as three noes, and I agree with that. You are not the referees of our political speech. That is why all three of you have to be more transparent and fair with your content moderation policies and your content selection algorithms, because at the moment it is, as I said, largely a black box. There is real mistrust among the American people about whether you are being fair or transparent. And this extends to concerns about the kinds of amplification and suppression decisions your platforms may make on Election Day and during the post-election period if the results of the election are too close to call.

And so I just want to underscore again, for my Democratic friends who keep using this really bad referee analogy, Google, Facebook and Twitter are not the referees over our democracy. Now, a second question, the PACT Act, which I referenced earlier, includes provisions to give users due process and an explanation when content they post is removed. So this is, again, a yes or no question. Do you agree that users should be entitled to due process and an explanation when content they post has been taken down? Mr. Zuckerberg?

Mr. ZUCKERBERG. Senator, I think that that would be a good principle to have.

Senator THUNE. Thank you. Mr. Dorsey.

Mr. DORSEY. Absolutely. We believe in a fair and straightforward appeals process.

Senator THUNE. Right. Mr. Pichai?

Mr. PICHAI. Yes, Senator.

Senator THUNE. Alright. Thank you. Mr. Zuckerberg, Mr. Dorsey, your platforms knowingly suppressed or limited the visibility of this *New York Post* article about the content on Hunter Biden's abandoned laptop. Many in the country are justifiably concerned how often the suppression of major newspaper articles occurs online. And I would say, Mr. Zuckerberg, would you commit to provide, for the record, a complete list of newspaper articles that Facebook suppressed or limited the distribution of over the past 5 years, along with an explanation of why each article was suppressed or the distribution was limited?

Mr. ZUCKERBERG. Senator, I can certainly follow up with you and your team to discuss that. We have an independent fact checking program, as you are as you are saying. You know, we try not to be arbiters of what is true ourselves, but we have partnered with fact checkers around the world to help assess that, to prevent misinformation and viral hoaxes from becoming widely distributed on our platform. And I believe that the information that they fact

check and the content that they fact check is public. So I think that there is probably already a record of this that can be reviewed.

Senator THUNE. Yes. But if you could do that, as it applies to newspapers, that would be very helpful. And Mr. Dorsey, would you commit to doing the same on behalf of Twitter?

Mr. DORSEY. We would absolutely be open to it and we are suggesting going a step further, which is aligned with what you are introducing in the PACT Act, which is much more transparency around our process, content moderation process, and also the results, the outcomes and doing that on a regular basis. I do agree and think that builds more accountability and ultimately that that lends itself to more trust.

Senator THUNE. Great. Thank you. All right. Very quickly, I have a lot of time either, but I often hear from conservative and religious Americans who look at the public statements of your companies, the geographic concentration of your companies and the political donations of your employees, which often are in the 80 to 90 percent to Democrat politicians. And you can see why this lack of ideological diversity among the executives and employees of your company could be problematic and may be contributing to some of the distrust among conservatives and Republican users.

And so I guess the question that I would ask is, and Mr. Zuckerberg, my understanding is that the person that is in charge of election integrity and security at Facebook is a former Joe Biden staffer. Is there someone that is closely associated with President Trump who is in the same sort of election integrity role at Facebook? And what—how do you all respond to that argument that there isn't sufficient balance in terms of the political ideology or diversity in your companies? And how do you deal with the lack of, sort of, trust that creates among conservatives?

The CHAIRMAN. Let's see if we can have three brief answers there.

Mr. ZUCKERBERG. Senator, I think having balances is valuable. We try to do that. I am not aware of the example that you say of someone in charge of this process who worked for Biden in the past. So we can follow up on that if that is right.

The CHAIRMAN. Follow up on the record for the rest of this answer, please, Mr. Zuckerberg. Thank you.

Mr. ZUCKERBERG. Alright.

The CHAIRMAN. Mr. Dorsey.

Mr. DORSEY. This is why I do believe it is important to have more transparency around our process and our practices, and it is independent of the viewpoints that our employees hold. We need to make sure that we are showing people that we have objective policies and enforcement.

The CHAIRMAN. And Mr. Pichai.

Mr. PICHAI. In these teams, there are people who are liberal, Republican, libertarian and so on. We are committed. We consult widely with important third party organizations across both sides when we develop our policies. And as the CEO, I am committed to running it without any political bias, but happy to engage more.

The CHAIRMAN. Thank you, gentlemen, and thank you, Senator Thune. The Ranking Member has now deferred to Senator Blumenthal. Sir, you are recognized.

STATEMENT OF HON. RICHARD BLUMENTHAL,
U.S. SENATOR FROM CONNECTICUT

Senator BLUMENTHAL. Thanks, Mr. Chairman, and thank you to the Ranking Member. I want to begin by associating myself with the very thoughtful comments made by the Ranking Member as to the need for broader consideration of issues of privacy and competition and local news. They are vitally important. And also with comments made by my colleague, Senator Klobuchar, about the need for antitrust review and I soon will be examining some of these topics in November before the Judiciary committee. You know, I have been an advocate of reform of Section 230 for literally 15 years. When I was Attorney General of the State of Connecticut, I raised this issue of the absolute immunity that no longer seems appropriate, so I really welcome the bipartisan consensus that we are seeing now that there needs to be constructive review. But frankly, I am appalled that my Republican colleagues are holding this hearing literally days before an election when they seem to want to bully and browbeat the platforms here to try to tilt them toward President Trump's favor.

The timing seems inexplicable, except to game the ref, in effect. I recognize the referee analogy is not completely exact, but that is exactly what they are trying to do, namely, to bully and browbeat these platforms to favor Senator—President Trump's tweets and posts. And frankly, President Trump has broken all the norms and he has put on your platforms potentially dangerous and lethal misinformation and disinformation. I am going to hold up one of them. This one, as you could see, pertains to COVID. We have learned to live with it, he says, just like we are learning to live with COVID, talking about the flu. We have learned to live with it.

In most populations, far less lethal. He has said that children, I would say almost definitely, are almost immune from this disease. He has said about the elections, big problems and discrepancies with mail in ballots all over the USA. Must have final total on November 3rd. Fortunately, the platforms are acting to label or take down these kinds of posts but my Republican colleagues have been silent. They lost their phones or their voices and the platforms, in my view——

The CHAIRMAN. We just lost your voice there in midsentence, Richard. Let's suspend for just a minute till we get.

Senator BLUMENTHAL. I hope you can hear me now.

The CHAIRMAN. There we are. OK. We can hear you now, Senator Blumenthal. Just start back one sentence before—we had you until then.

Senator BLUMENTHAL. I just want to say about this disinformation from the President, there has been deafening silence from my Republican colleagues. And now we have a hearing that is, in effect, designed to intimidate and browbeat the platforms that have labeled this disinformation for exactly what it is. We are on the verge of a massive onslaught on the integrity of our election. President Trump has indicated that he will potentially interfere by posting disinformation on Election Day or the morning after. The Russians have begun already interfering in our elections. We have all received briefings that are literally chilling about what they are doing and the FBI and the CSIS have recently issued public alerts

that, "foreign actors and cyber criminals likely to spread disinformation regarding 2020 results." They are making 2016 look like child's play in what they are doing.

So, President Trump and the Republicans have a plan which involves disinformation and misinformation. The Russians have a plan. I want to know whether you have a plan, Facebook, Twitter, Google, a plan if the President uses your platforms to say on the day of the election that there is rigging or fraud without any basis in evidence or attempts to say that the election is over and the voting—the counting of votes must stop either on November 4 or someday subsequent. And I would like as to this question about whether you have a plan, a yes or no.

Mr. ZUCKERBERG. Senator to start, we do. We have policies related to all of the areas that you just mentioned. Candidates or campaigns trying to delegitimize methods of voting or the election. Candidates trying to prematurely declare victory. And candidates trying to spread voter suppression material that is misleading about how, when or where to vote. So we are—we have taken a number of steps on that front.

The CHAIRMAN. Perhaps we could take Mr. Pichai next and then Mr. Dorsey. Mr. Pichai.

Mr. PICHAI. Senator, yes, we definitely are robustly—we have been planning for a while and we rely on racing up on new sources through moments like that, as well as we have closely partnered with the *Associated Press* to make sure we can provide users the most accurate information possible.

Mr. DORSEY. Yes, we also we also have a plan. So, you know, our plan and our enforcement around these issues is pointing to more information and specifically state election officials. So we want to give the people using the service as much information as possible.

The CHAIRMAN. Thank you, Senator Blumenthal. Senator Cruz.

STATEMENT OF HON. TED CRUZ,
U.S. SENATOR FROM TEXAS

Senator CRUZ. I want to thank you, Mr. Chairman, for holding this hearing. The three witnesses we have before this Committee today collectively pose, I believe, the single greatest threat to free speech in America and the greatest threat we have to free and fair elections. Yesterday, I spent a considerable amount of time speaking with both Mr. Zuckerberg and Mr. Pichai. I have concerns about the behavior of both of their companies. I would note that Facebook is at the minimum, at least trying to make some efforts in the direction of defending free speech.

I appreciate their doing so. Google, I agree with the concerns that Senator Klobuchar raised. I think Google has more power than any company on the face of the planet. And the antitrust concerns are real. The impact of Google is profound. And I expect we will have continued and ongoing discussions about Google's abuse of that power and its willingness to manipulate search outcomes to influence and change election results. But today, I want to focus my questioning on Mr. Dorsey and on Twitter, because of the three players before us, I think Twitter's conduct has by far been the most egregious. Mr. Dorsey, does Twitter have the ability to influence elections?

Mr. DORSEY. No.

Senator CRUZ. You don't believe Twitter has any ability to influence elections?

Mr. DORSEY. No, we are one part of a spectrum of communication channels that people have.

Senator CRUZ. So you are testifying to this committee right now that Twitter, when it silences people, when it censors people, when it blocks political speech, that has no impact on elections?

Mr. DORSEY. People have a choice of other communication channels——

Senator CRUZ. Not if they don't hear information. If you don't think you have the power to influence elections, why do you block anything?

Mr. DORSEY. Well, we have policies that are focused on making sure that more voices on the platform are possible. We see a lot of abuse and harassment, which turns up silencing people and helping them leave from the platform.

Senator CRUZ. Alright. Mr. Dorsey, I find your opening questions, your opening answers absurd on their face. Let's talk about the last two weeks in particular. As you know, I have long been concerned about Twitter's pattern of censoring and silencing individual Americans with whom Twitter disagrees. But two weeks ago, Twitter and to a lesser extent, Facebook, crossed a threshold that is fundamental in our country.

Two weeks ago, Twitter made the unilateral decision to censor the *New York Post* in a series of two blockbuster articles, both alleging evidence of corruption against Joe Biden. The first concerning Ukraine. The second concerning communist China. And Twitter made the decision, number one, to prevent users, any user from sharing those stories. And number two, you went even further and blocked the *New York Post* from sharing on Twitter its own reporting. Why did Twitter make the decision to censor the *New York Post*?

Mr. DORSEY. We had a hack materials policy——

Senator CRUZ. When was policy adopted?

Mr. DORSEY.in 2018, I believe.

Senator CRUZ. In 2018, go ahead. What was the policy?

Mr. DORSEY. So the policy is around limiting the spread of materials that are hackable. We do not want Twitter to be a distributor for hacked materials. We found that the *New York Post,* because it showed the direct materials, screenshots of the direct materials, and it was unclear how these were obtained, that it fell under this policy. Now——

Senator CRUZ. So in your view, if it is unclear the source of a document, in this instance, the *New York Post* documented what it said the source was, which it said it was a laptop owned by Hunter Biden that had been turned into a repair store. So they weren't hiding what they claimed to be the source. Is it your position that Twitter, when you can't tell the source, blocks press stories?

Mr. DORSEY. No, not at all. Our team made a fast decision. The enforcement action over blocking URLs, both in tweets and in DMs, in direct messages, we believe was incorrect. And we changed it—
.

Senator CRUZ. Today, the *New York Post* is still blocked from tweeting two weeks later.

Mr. DORSEY. Yes, they have to log into their account, which they can do at this minute. Delete the original tweet, which fell under our original enforcement actions, and they can tweet the exact same material, the exact same article, and it would go through.

Senator CRUZ. And so, Mr. Dorsey, your ability is you have the power to force a media outlet—and let's be clear, the *New York Post* isn't just some random guy tweeting. The *New York Post* has the fourth highest circulation of any newspaper in America. The *New York Post* is over 200 years old. The *New York Post* was founded by Alexander Hamilton. And your position is that you can sit in Silicon Valley and demand of the media, that you can tell them what stories they can publish. You can tell the American people what reporting they can hear, is that right?

Mr. DORSEY. No, every person, every account, every organization that signs up to Twitter, agrees to a terms of service. And the terms of service——

Senator CRUZ. The media outlets must genuflect and obey your dictates if they wish to be able to communicate with readers. Is that right?

Mr. DORSEY. No, not at all. You know, we recognized an error in this policy and specifically the enforcement——

Senator CRUZ. You are still blocking their post. You are still blocking their post. Right now, today, you are blocking their posts.

Mr. DORSEY. We are not blocking the post. Anyone can tweet——

Senator CRUZ. Can the *New York Post* on their Twitter account?

Mr. DORSEY. If they go into their account—.

Senator CRUZ. No is your answer to that, no, unless they take back and agree with your dictates. Let me ask you something, you claimed it was because of a hacked materials policy. I find that facially highly dubious and clearly employed in a deeply partial way. Did Twitter block the distribution of the *New York Times* story a few weeks ago that purported to be based on copies of President Trump's tax returns?

Mr. DORSEY. We didn't find that to be a violation of our terms of service because his reporting about the material, it wasn't distributing the material.

Senator CRUZ. OK. Well, that's actually not true. They posted what they reported to be original source materials and Federal law, Federal statute makes it a crime, a Federal felony to distribute someone's tax returns against their knowledge. So that material was based on something that was distributed in violation of Federal law, and yet Twitter gleefully allowed people to circulate that. But when the article was critical of Joe Biden, Twitter engaged in rampant censorship and silencing.

Mr. DORSEY. And again, we recognize there is that policy. We changed it within 24 hours.

Senator CRUZ. But you still blocked the *New York Post.* You haven't changed it.

Mr. DORSEY. We have changed that. They can log into their account, delete the original tweet——

Senator CRUZ. You forced a *Politico* reporter to take down his post about the *New York Post* as well. Is that correct?

Mr. DORSEY. Within that 24 hour period, yes. But we know as the policy has changed, anyone can tweet that article out.

The CHAIRMAN. Thank you, Senator Cruz.

Senator CRUZ. So you censor the *New York Post.* You can censor *Politico.* Presumably you can censor the *New York Times* or any other media outlet. Mr. Dorsey, who the hell elected you and put you in charge of what the media are allowed to report and what the American people are allowed to hear? And why do you persist in behaving as a Democratic super PAC, silencing views to the contrary of your political beliefs?

The CHAIRMAN. Let's give Mr. Dorsey a few seconds to answer that, and then we will have to conclude this segment.

Mr. DORSEY. We are not doing that. And this is why I opened this hearing with calls for more transparency. We realize we need to earn trust more. We realize that more accountability is needed to show our intentions and to show the outcomes.

The CHAIRMAN. Thank you, Senator.

Mr. DORSEY. So I hear the concerns and acknowledge them. We want to fix it with more transparency.

The CHAIRMAN. Thank you, Senator Cruz. The Ranking Member has deferred now to Senator Schatz, who joins us remotely. Sir, you are recognized.

STATEMENT OF HON. BRIAN SCHATZ,
U.S. SENATOR FROM HAWAII

Senator SCHATZ. Thank you, Mr. Chairman. Thank you, Ranking Member. You know, this is an unusual hearing at an unusual time. I have never seen a hearing so close to an election on any topic, let alone on something that is so obviously a violation of our obligation under the law and the rules of the Senate to stay out of electioneering. We never do this and there is a very good reason that we don't call people before us to yell at them for not doing our bidding during an election. It is a misuse of taxpayer dollars. What is happening here is a scar on this committee and the U.S. Senate. What we are seeing today is an attempt to bully the CEOs of private companies into carrying out a hit job on a Presidential candidate by making sure that they push out foreign and domestic misinformation meant to influence the election. To our witnesses today, you and other tech leaders need to stand up to this immoral behavior.

The truth is that because some of my colleagues accuse you, your companies and your employees of being biased or liberal, you have institutionally bent over backward and overcompensated. You have hired Republican operatives, posted private dinners with Republican leaders and in contravention of your terms of service, given special dispensation to right wing voices and even throttled progressive journalism. Simply put, the Republicans have been successful in this play. And so during one of the most consequential elections in American history, my colleagues are trying to run this play again and it is an embarrassment. I have plenty of questions for the witnesses on Section 230, on antitrust, on privacy, on anti-Semitism, on their relationship with journalism, but we have to call this hearing what it is, it is a sham.

And so for the first time in my 8 years in the U.S. Senate, I am not going to use my time to ask any questions because this is nonsense and it is not going to work this time. This play my colleagues are running did not start today. And it is not just happening here in the Senate. It is a coordinated effort by Republicans across the Government. Last May, President Trump issued an Executive Order designed to narrow the protections of Section 230 to discourage platforms from engaging in content moderation on their own sites. After it was issued, President Trump started tweeting that Section 230 should be repealed, as if he understands Section 230. In the last 6 months, President Trump has tweeted to repeal Section 230 five times, in addition to other tweets in which he has threatened the tech companies.

A few weeks later, President Trump withdrew the nomination of FCC Commissioner Michael O'Rielly. Republican Commissioner O'Rielly questioned the FCC's authority to regulate under Section 230 and the statute is not unclear on this. President Trump then nominated Nathan Simington, who was the drafter of NTIA's petition to the FCC regarding Section 230. And Republican Senators have enthusiastically participated. Since June of this year, six Republican only bills have been introduced, all of which threaten platforms' ability to moderate content on their site. And as the election draws closer, this Republican effort has become more and more aggressive. September 23, DOJ unveiled its own Section 230 draft legislation that would narrow the protections under the current law and discourage platforms from moderating content on their own site. September 14 and October 1, respectively, Senators Hawley and Kennedy tried to pass their Republican-only Section 230 bills by unanimous consent.

Now, what that means is they went down to the floor and without a legislative hearing, without any input from Democrats at all, they tried to pass something so foundational to the Internet unanimously without any discussion and any debate. On the same day as Senator Kennedy's UC attempt, Senator Wicker forced the Commerce Committee, without any discussion or negotiation beforehand, to vote on subpoenaing the CEOs of Twitter, Facebook and Google to testify. That is why we are here today. Two weeks later, on October 14, Justice Clarence Thomas, on his own, issued a statement that appeared to support the narrowing of the court's interpretation on Section 230. The very next day, the FCC Chairman, Ajit Pai, announced that the FCC would seek to clarify the meaning of Section 230.

On that day, Senator Graham announced that the Judiciary Committee would vote to subpoena the tech companies over the content moderation. And the context of this, in addition to everything, Senator Cruz is on Maria Bartiromo talking about a blockbuster story from *The New York Post.* Senator Hawley is on Fox and on the Senate floor. And the Commerce Committee itself is tweeting out a campaign style video that sort of alarmingly says Hunter Biden's e-mails, text censorship. On October 21, Senator Hawley reattempted to pass his bill on Section 230 via UC, again, without going through any committee markup or vote.

And on Friday, Senator Graham announced that the CEOs of Facebook and Twitter would testify before the Senate Judiciary

Committee on November 17. This is bullying and it is for electoral purposes. Do not let the U.S. Senate bully you into carrying the water for those who want to advance misinformation. And don't let the specter of removing Section 230 protections or an amendment to antitrust law or any other kinds of threats cause you to be a party to the subversion of our democracy. I will be glad to participate in good faith, bipartisan hearings on these issues when the election is over. But this is not that. Thank you.

The CHAIRMAN. Thank you, Senator Schatz. Next is Senator Fischer.

STATEMENT OF HON. DEB FISCHER,
U.S. SENATOR FROM NEBRASKA

Senator FISCHER. Thank you, Mr. Chairman. Gentlemen, I am not here to bully you today. And I am certainly not here to read any kind of political statement right before an election. To me, this hearing is not a sham. I am here to gain some clarity on the policies that you use. I am here to look at your proposals for more transparency because your platforms have become an integral part of our democratic process for both candidates, but also more importantly, for our citizens as well. Your platforms also have enormous power to manipulate user behavior and to direct content and to shape narratives. Mr. Dorsey, I heard your opening statement. I have read it. You also tweeted that the concept of good faith is what is being challenged by many of you here today. Some of you don't trust we are acting in good faith. That is the problem I want to focus on solving. Mr. Dorsey, why should we trust you with so much power? In other words, why shouldn't we regulate you more?

Mr. DORSEY. Well, the suggestions we are making around more transparency is how we want to build that trust. We do agree that we should be publishing more of our practice of content moderation. We have made decisions to moderate content. We have made decisions to moderate content to make sure that we are enabling as many voices on our platform as possible. And I acknowledge and completely agree with the concerns that it feels like a black box and anything that we can do to bring transparency to it, including publishing our policies, our practices, answering very simple questions around how content is moderated, and then doing what we can around the growing trend of algorithms moderating more of this content. As I said, this one is a tough one to actually bring transparency to. Explainability in AI is a field of research, but it is far out. And I think a better opportunity is giving people more choice around the algorithms they use, including people to turn off the algorithms completely, which is what we are attempting to do. So——

Senator FISCHER. Right. You can understand the concerns that the people have when they see that what many consider you are making value judgments on what is going to be on your platforms. You say users can report content and then you take action. But certainly you can understand that people are very concerned, they are very worried about what they see as manipulation on your part. And to say you are going to have more transparency and—yes, that is—sir, I would say with respect, I don't think that is enough just to say you are you are going to have that transparency there and

you are not influencing people, because as any time a free press is blocked on both sides with what we would view in the political world as both sides here, when views aren't able to be expressed, that does have a huge amount of influence.

Mr. DORSEY. I completely understand. I agree that it is not enough. I don't think transparency alone addresses these concerns. I think we have to continue to push for a more straightforward and fast and efficient appeals process. And I do believe we need to look deeply at algorithms and how they are used and how people have choice on how to use those algorithms or whether they use them.

Senator FISCHER. But ultimately, somebody makes a decision. Where does the buck stop? With the algorithms? Where does the buck stop? Who is going to make a value judgment? Because in my opinion, it is a value judgment.

Mr. DORSEY. Well, ultimately, I am accountable to all the decisions that the company makes. But we want to make sure that we are providing clear frameworks that are objective and that can be tested. And that we have multiple checkpoints associated with them so that we can learn quickly if we are doing something in error.

Senator FISCHER. And when your company amplifies some content over others, is it fair for you to have legal protections for your actions?

Mr. DORSEY. We believe so. Keep in mind, a lot of our algorithms recommending content is focused on saving people time. So we are ranking things that the algorithms believe people would find most relevant and most valuable in the time——

Senator FISCHER. That is your value judgment on what those people would find most relevant.

Mr. DORSEY. No, it is not a value judgment. It is based on engagement metrics, it is based on who you follow. It is based on activity you take on on the network.

Senator FISCHER. Mr. Zuckerberg, with your ever expanding content moderation policies, are you materially involved in that content?

Mr. ZUCKERBERG. Senator, yes, I spend a meaningful amount of time on making sure that we get our content policies and enforcement right.

Senator FISCHER. OK. Thank you. What, if any, changes do you think should be made to Section 230 to address the specific concerns regarding content moderation that you have heard so far this morning?

Mr. ZUCKERBERG. Senator, I would outline a couple. First, I agree with Jack that increasing transparency into the content moderation process would be an important step for building trust and accountability. One thing that we already do at Facebook is every quarter, we issue a transparency report where for each of the 20 or so categories of harmful content that we are trying to address, so terrorism, child exploitation, incitement of violence, pornography, different types of content, we issue a report on how we are doing, what the prevalence of that content is on our network, and what percent of it our systems are able to take down before someone even has to report it to us, and what the precision is and basically how accurate our systems are in dealing with it.

And getting to the point where everyone across the industry is reporting on a baseline like that I think would be valuable for people to have these discussions, not just about anecdotes, OK, I saw a piece of content tonight. I am not necessarily sure I agree with how that was moderated. It would allow the conversation to move to data so that we can understand how these platforms are performing overall and hold them accountable.

The CHAIRMAN. Thank you.

Senator FISCHER. At issue with your answer, I think, would be the time involved. That it wouldn't be an immediate response to have that conversation, as you call it. I hope that all three of you gentlemen can answer that question and written questions. So my time is up. Thank you.

The CHAIRMAN. Thank you, Senator Fischer. I appreciate that. We are going to take now Senator Cantwell's questioning after which we are going to accommodate our witnesses with a 5-minute recess. So, Senator Cantwell, you are recognized.

Senator CANTWELL. Thank you, Mr. Chairman. Can you hear me?

The CHAIRMAN. Surely can.

Senator CANTWELL. And can you see me this time?

The CHAIRMAN. We can now see you, yes.

Senator CANTWELL. OK. Well, thank you, Mr. Chairman. And this is such an important hearing. I agree with many of the statements my colleagues have made, that this hearing didn't need to take place at this moment, that the important discussion about how we maintain a thriving Internet economy and how we continue to make sure that hate speech and misinformation is taken down from the web is something that would probably better been done in January than now. But here we are today and we have heard some astounding things that I definitely must refute.

First of all, I am not going to take lightly anybody who tries to undermine mail-in voting. Mail-in voting in the United States of America is safe. The State of Washington and the State of Oregon have been doing it for years. There is nothing wrong with our mail-in system. So, I think that there will be secretaries of state and law enforcement agencies who have worked hard with state election officials, and others who will be talking about how this process works and how we are going to fight to protect it.

I am also not going to demean an organization just because they happen to be headquartered in the State of Washington or happen to have business there. I seriously doubt that the geography of a company somehow makes it more political for one side of the aisle or another. I know that because I see many of you coming to the State of Washington for Republican fundraisers with these officials. I know ythat ou know darn well that there are plenty of Republicans that work in high tech firms.

So, the notion that somehow these people are crossing the aisle because of something and creating censorship, the notion that free speech is about the ability to say things and it doesn't take—well, maybe we need to have a history lesson from high school again. But, yes, free speech means that people can make outrageous statements about their beliefs. So, I think that the CEOs are telling us here what their process is for taking down healthcare information

that is not true, that is a threat to the public, and information that is a threat to our democracy.

That is what they are talking about. So, I want to make it clear that this hearing could have happened at a later date, and I don't appreciate the misinformation that is coming across today, that is trying to undermine our election process. It is safe. It is the backbone of what distinguishes America from other countries in the world. We do know how to have a safe and fair election. And one of the ways that we are doing that is to have these individuals work with our law enforcement entities.

My colleague Gary Peters made it very clear that they successfully helped stop a threat targeting the Governor of Michigan. And why? Because they were working with them to make sure that information was passed on. So, this is what we are talking about. We are talking about whether we are going to be on the side of freedom and information and whether we are going to put our shoulder to the wheel to continue to make sure that engine is there or whether we are going to prematurely try to get rid of 230 and squash free speech. And so, I want to make sure that we continue to move forward. So, Mr. Zuckerberg, I would like to turn to you because there was a time when there was great concern about what happened in Myanmar, about the Government using information against a Muslim minority. And you acted and reformed the system.

Just recently in September, Facebook and Twitter announced they had suspended networks and accounts linked to various organizations and used for laundering Russian backed websites and accounts and derisive propaganda that we associated with state-run attempts to interfere in our elections. So, could you please, Mr. Zuckerberg, talk about what you were doing to make sure state-run entities don't interfere in U.S. elections?

Mr. ZUCKERBERG. Yes. Thank you, Senator. Since 2016, we have been building up some very sophisticated systems to make sure that we can stop foreign interference in elections, not just in the U.S., but all around the world. And a lot of this involves building up AI systems to identify when clusters of accounts aren't behaving in the way that a normal person would. They are behaving as fake accounts in some coordinated way. A lot of this is also about forming partnerships. The tech companies here today work more closely together to share signals about what is happening on the different platforms to be able to combat these threats, as well as working more closely with law enforcement and intelligence communities around the world.

And the net result of that is that over the last few years, we have taken down more than 100 networks that were potentially attempting to spread misinformation or interfere. A lot of them were coming from Russia or Iran. A growing number from China as well. And at this point, I am proud that our company and as well as the others in the industry, I think have built systems that are very effective at this. We can't stop countries like Russia from trying to interfere in an election. Only the U.S. Government can really push back with the appropriate leverage to do that. But we have built up systems to make sure that we can identify much faster when they are attempting to do that, and I think that that should give

the American people a good amount of confidence leading into this election.

Senator CANTWELL. And is it true that those entities are trying to find domestic sources to help with that misinformation?

Mr. ZUCKERBERG. Senator, yes. The tactics of these different Governments are certainly evolving, including trying to find people outside of their country and in some cases we are seeing domestic interference operations as well. And the systems have had to evolve to be able to identify and take those down as well. Of the hundred or so networks that I just cited that we took down, about half were domestic operations at this point. And that is in various countries around the world, not primarily in the U.S., but this is a global phenomenon that we need to make sure that we continue pushing forward aggressively on.

Senator CANTWELL. Thank you.

Mr. Pichai, I would like to turn to you for a second, because I do want information from Facebook on this point, but I would like to turn to you. There is information now from media organizations that broadcasters and newsprint are losing somewhere between 30 and 50 percent of their revenue that they could be getting to newspapers and broadcasting to the formats that Google has as it relates to their platform and ad information. Can you confirm what information you have about this? And do you think that Google is taking ad revenue from these new sources in an unfair way?

Mr. PICHAI. Senator, it is an important topic. It is a complex topic. I do think journalism, as you rightfully call it, attention to it, particularly local journalism, is very important. The Internet has been a tremendously disrupting force and the pandemic has exacerbated it. I would have to say that Google, you know, I would make the case that we believe in raising news across our products because we realize the importance of journalism. We send a lot of traffic to news publishers, all the ad technology questions I am getting asked today. We invest in ad technology.

We share the majority of revenue back to publishers. We are investing in subscription products. We have committed $2 billion in new licensing over the next 2 years to news organizations. We have set up local emergency fund to code for local journalistic institutions. I could give plenty of examples, but the underlying forces which are impacting the industry, which is the Internet and whether it is school or if not Google, advertisers are——

Senator CANTWELL. Yes, I don't have a clock——

The CHAIRMAN. You are a minute and a half over, but so let's——

Senator CANTWELL. OK. I would just leave it with this, that Mr. Pichai, you hit on the key word, majority. I don't think that you are returning in the majority of the revenue to these broadcast entities. I do think it is a problem. Yes, they have had to make it through the transformation, which is a rocky transformation. The message from today's hearing is the free press needs to live and be supported by all of us. And we look forward to discussing how we can make sure that they get fair return on their value. Thank you, Mr. Chairman.

The CHAIRMAN. Thank you, Senator Cantwell. We will now take a five-minute recess and then we will begin with—most of our

members have not yet had a chance to ask questions. The Committee is in recess for five minutes.

[Recess.]

The CHAIRMAN. OK. This hearing will return to order and we understand that Senator Moran is next. Sir, you are recognized.

STATEMENT OF HON. JERRY MORAN,
U.S. SENATOR FROM KANSAS

Senator MORAN. Chairman Wicker, thank you very much. And thank you for you and Senator Cantwell hosting this hearing. Let me address initially the topic that seems to be primary today and the time, data privacy. Let me ask all three witnesses, how much money does your company spend annually on content moderation? How many people work in general in the area of content moderation, including by private contract? Let me just start with those two questions. I also want to ask you, how much money does your company spend in defending lawsuits stemming from user content on the platform?

The CHAIRMAN. OK, Mr. Zuckerberg, you want to go first there?

Mr. ZUCKERBERG. Senator, we have more than 35,000 people who work on content and safety review. And I believe our budget is multiple billions of dollars a year on this. I think upwards of three or maybe even more billion dollars a year, which is a greater amount in revenue that we are spending on this than the whole revenue of our company was the year before we filed to go public in 2012.

The CHAIRMAN. Mr. Pichai.

Mr. PICHAI. Senator, we use both a combination of human reviewers and AI moderation systems. We have well over an available 10,000 reviewers and we are investing there significantly. And, you know, I would again, not sure of the exact numbers, but I would say it is in the order of four billion dollars we spend on these things.

The CHAIRMAN. Thank you. Mr. Dorsey.

Mr. DORSEY. I don't have the specific numbers, but we want to maintain agility between the people that we have working on this and also just building better technology to automate it. So our goal is flexibility here.

Senator MORAN. Let me ask that question again about how much would you estimate that your company is currently spending on defending lawsuits related to user content?

The CHAIRMAN. In the same order. OK?

Mr. ZUCKERBERG. Senator, I don't know the answer to that off the top of my head, but I can get back to you.

Senator MORAN. Thank you.

Mr. PICHAI. Senator, we do spend a lot on legal lawsuits, but not sure what of it applies to content related issues. But will be happy to follow up.

Senator MORAN. Thank you.

Mr. DORSEY. And I don't have those numbers.

Senator MORAN. Let me use your answers to highlight something that I want to be a topic of our conversation as we debate this legislation. Whatever the numbers are, you indicate that they are significant. It is an enormous amount of money and an enormous

amount of employee time, contract labor time in dealing with a modification of content. These efforts are expensive and I would highlight for my colleagues on the Committee that they will not be any less expensive, perhaps less in scale, but not less in cost for startups and small businesses. And as we develop our policies in regard to this topic, I want to make certain that entrepreneurship, startup businesses and small business are considered in what it would cost in their efforts to meet the kind of standards that operate in this sphere. Let me quickly turn to Federal privacy.

I chaired the Consumer Data Privacy Security Act. We tried for months, Senator Blumenthal and I, to develop a bipartisan piece of legislation. We were close, but unsuccessful in doing so. Let me ask Mr. Zuckerberg, Facebook entered into a Consent Order with the FTC in July 2012 for violations of the FTC Act and later agreed to pay a $5 billion penalty, along with a robust settlement order in 2018, following the Cambria Analytica incident that violated the 2012 order. My legislation would provide the FTC with first time civil penalty authority. Do you think this type of enforcement tool for the FTC would better deter unfair and deceptive practices than the current enforcement regime?

Mr. ZUCKERBERG. Senator, I would need to understand it a little bit more detail before weighing in on this. But I think that the settlement that we have with the FTC, what we are going to be setting up is an industry leading privacy program. We have, I think, more than 1,000 engineers working on the privacy program now and we are basically implementing a program which is sort of the equivalent of Sarbanes-Oxley's financial regulation around kind of internal auditing and controls around privacy and protecting people's data as well. So I think that that settlement will be quite effective in ensuring that people's data and privacy are protected.

Senator MORAN. Mr. Pichai, Google, YouTube's $170 million settlement with the FTC in the State of New York for alleged violations of COPPA involve persistent identifiers. How should Federal legislation address persistent identifiers for consumers over the age of 13?

Mr. PICHAI. Senator, we today have invested—we have done two things as a company, we have invested in wonderful special product called YouTube Kids. Fair content can be safe for kids. Obviously, on a YouTube main product today, how the Internet gets used, families to view content and part of our settlement was adapting so that we can accommodate for those use cases as well. You know, privacy is one of the most important areas being missed and as a company, have thousands of engineers working on it. We believe in giving users control of choice and transparency.

And any time we associate data with users, we are transparent, they can go see what data is there. We give them delete controls. We give data portability options. And just last year, we announced an important change by which for all new users, we delete that data automatically without them needing to do anything. And the end stages is to go through privacy checkup or a billion people have gone through their privacy checkouts and, you know, it is an area where we are investing significantly.

Senator MORAN. Thank you. Chairman. I don't see my time clock. Do I have time for one more?

The CHAIRMAN. You really don't. Your time has just expired. But thank you very much for—.

Senator MORAN. Mr. Chairman, thank you.

The CHAIRMAN. Thank you so much. Senator Markey.

STATEMENT OF HON. EDWARD MARKEY,
U.S. SENATOR FROM MASSACHUSETTS

Senator MARKEY. Thank you, Mr. Chairman, very much. Today from his Republican allies in Congress and his propaganda parrots on Fox News are peddling a myth. And today, my Republican colleagues on the Senate Commerce Committee are simply doing the President's bidding. Let's be clear, Republicans can and should join us in addressing the real problems posed by big tech. But instead, my Republican colleagues are determined to feed a false narrative about anti-conservative bias meant to intimidate big tech so it will stand idly by and allow interference in our elections again. Here is the truth. Violence and hate speech online are real problems. Anti-bias is a problem. Foreign attempts to influence our election with disinformation are real problems. The anti-conservative bias is not a problem. The big tech business model which puts profits ahead of people is a real problem. Anti-conservative bias is not a problem.

The issue is not that the companies before us today are taking too many posts down, the issue is that they are leaving too many dangerous posts up. In fact, they are amplifying harmful content so that it spreads like wildfire and torches our democracy. Mr. Zuckerberg, when President Trump posted on Facebook that when the looting starts, the shooting starts, you failed to take down that post. Within a day, the post had hundreds of thousands of shares and likes on Facebook. Since then, the President has gone on national television and told a hate group to, "stand by." And he has repeatedly refused to commit that he will accept the election results.

Mr. Zuckerberg, can you commit that if the President goes on Facebook and encourages violence after election results are announced, that you will make sure your company's algorithms don't spread that content and you will immediately remove those messages?

Mr. ZUCKERBERG. Senator, yes, incitement of violence is against our policy, and there are not exceptions to that, including for politicians.

Senator MARKEY. There are exceptions, did you say?

Mr. ZUCKERBERG. There are no exceptions.

Senator MARKEY. There are no exceptions, which is very important because obviously there could be messages that are sent that could throw our democracy into chaos and a lot of it can be and will be created if social media sites do not police what the President says. Mr. Zuckerberg, if President Trump shares Russian or Iranian disinformation lying about the outcome of the election, can you commit that you will make sure your algorithms do not amplify that content and that you will immediately take that content down?

Mr. ZUCKERBERG. Senator, we have a policy in place that prevents any candidate or campaign from prematurely declaring vic-

tory or trying to delegitimize the result of the election. And what we will do in that case is we will append some factual information to any posts that is trying to do that. So if someone says that they won the election when the result isn't in, for example, we will append a piece of information to that saying that official election results are not in yet.

So that way, anyone who sees that post will see that context in line. And also, if one of the candidates tries to prematurely declare victory or cite an incorrect result, we have a precaution that we have built in to the top of the Facebook app for everyone who signs in, in the U.S. information about the accurate U.S. election voting results. I think that this is a very important issue to make sure that people can get accurate information about the results of the election.

Senator MARKEY. It cannot be stated as being anything less than critically important. Democracy could be seriously challenged beginning next Tuesday evening and for several days afterwards, maybe longer. And a lot of responsibility is going to be on the shoulders of Facebook and our other witnesses today. Mr. Zuckerberg, if President Trump uses his Facebook account to call for armed private citizens to patrol the polls on Election Day, which would constitute illegal voter intimidation and violation of the Voting Rights Act, will you commit that your algorithms will not spread that content and that you will immediately take that content down?

Mr. ZUCKERBERG. Senator, my understanding is that content like what you are saying would violate our voter suppression policies and would come down.

Senator MARKEY. OK. Again, the stakes are going to be very high and we are going to take that as a commitment that you will do that because obviously we would otherwise have a serious question mark placed over our elections. We know Facebook cares about one thing using—keeping users glued to its platform. One of the ways you do that is with Facebook groups. Mr. Zuckerberg, in 2017, you announced the goal of one billion users joining Facebook groups. Unfortunately, these forum pages have become breeding grounds for hate, echo chambers of misinformation, and then used for coordination of violence. Again, Facebook is not only failing to take these pages down, it is actively spreading these pages and helping these groups' recruitment efforts. Facebook's own internal research found that 64 percent of all extremist group joins are due to Facebook's recommendation tools. Mr. Zuckerberg, will you commit to stopping all group recommendations on your platform until U.S. election results are certified, yes or no?

Mr. ZUCKERBERG. Senator, we have taken the step of stopping recommendations and groups for all political content or social issue groups as a precaution for this. Just to clarify one thing, the vast, vast majority of groups and communities that people are part of are not extremist organizations or even political. They are interest based in communities that I think are quite helpful and healthy for people to be a part of. I do think we need to make sure that our recommendation algorithm doesn't encourage people to join extremist groups. That is something that we have already taken a number

of steps on. And I agree with you it is very important that we continue to make progress on.

The CHAIRMAN. Thank you.

Senator MARKEY. Well, your algorithm are promoting online spaces that foster political violence. At the very least, you should disable those algorithms that are recruiting users during this most sensitive period of our democracy.

The CHAIRMAN. Thank you.

Senator MARKEY. Thank you, Mr. Chairman.

The CHAIRMAN. Thank you, Senator Markey. Mr. Zuckerberg, let me just ask you this. In these scenarios that Senator Markey was posing, the action of Facebook would not be a function of algorithms in those cases, would it?

Mr. ZUCKERBERG. Senator, I think that you are right and that that is a good clarification. A lot of this is more about enforcement of content policies. Some of the questions were about algorithms. I think group ranking is an algorithm. But broadly, I think a lot of it is content enforcement.

The CHAIRMAN. Thank you for clarifying that. Senator Blackburn, you are recognized.

STATEMENT OF HON. MARSHA BLACKBURN,
U.S. SENATOR FROM TENNESSEE

Senator BLACKBURN. Thank you, Mr. Chairman, and I want to thank each of you for coming to us voluntarily. We appreciate that. There are undoubtably benefits to using your platforms, as you have heard everyone mention today. There are also some concerns which you are also hearing. Privacy, free speech, politics, religion and I have kind of chuckled as I have sat here listening to you all. That book, *Valley of the Gods,* it reminds me that you all are kind of in control of what people are going to hear, what they are going to see, and therefore, you have the ability to dictate what is coming in, what information is coming into them. And I think it is important to realize, you know, you're set up as an information source, not as a news media. And so, therefore, censoring things that you all think unseemly may be something that is not unseemly to people in other parts of the country. But let me ask each of you very quickly, do any of you have any content moderators who are conservatives? Mr. Dorsey, first, yes or no?

Mr. DORSEY. We don't ask political ideology——

Senator BLACKBURN. OK, you don't. OK, Mr. Zuckerberg?

Mr. ZUCKERBERG. Senator, we don't ask for their ideology, but just statistically, they are 35,000 in cities and places all across the country and world so I imagine, yes.

Senator BLACKBURN. Mr. Pichai?

Mr. PICHAI. The answer would be yes, because we hired them, you know, through the United States.

Senator BLACKBURN. OK. Alright. And looking at some of your censoring, Mr. Dorsey, you all have censored Joe Biden zero times. You have censored Donald Trump 65 times. So I want to go back to Senator Gardner's question. You claimed earlier that the Holocaust denial and threats of Jewish genocide by Iran's terrorist Ayatollah don't violate Twitter's so called rules and that it is important

for world leaders like Iran's terrorist leader to have a platform on Twitter. So let me ask you this. Who elected the Ayatollah?

Mr. DORSEY. I don't know.

Senator BLACKBURN. You don't know? OK. I think this is called a dictatorship. So are people in Iran allowed to use Twitter, or does the country whose leader you claim deserves a platform ban them from doing so?

Mr. DORSEY. Ideally, we would love for the people of Iran to use Twitter.

Senator BLACKBURN. Well, Iran banned Twitter, and Mr. Zuckerberg, I know you are aware they banned Facebook also. So, Mr. Dorsey, is Donald Trump a world leader?

Mr. DORSEY. Yes.

Senator BLACKBURN. OK. So it would be important for world leaders to have access to your platform, correct?

Mr. DORSEY. Correct.

Senator BLACKBURN. And so why did you deny that platform via censorship to the U.S. President?

Mr. DORSEY. We haven't censored the U.S. President.

Senator BLACKBURN. Oh, yes, you have. How many posts from Iran's terrorist Ayatollah have you censored? How many posts from Vladimir Putin have you censored?

Mr. DORSEY. We have we have label tweets of world leaders. We have a policy around not taking down the content, but simply finding more context around it.

Senator BLACKBURN. OK. And the U.S. President, you have censored 65 times. You testified that you are worried about disinformation and election interference. That is something we all worry about. And, of course, for about 100 years, foreign sources have been trying to influence U.S. policy and U.S. elections. Now they are onto your platforms. They see this as a way to get access to the American people. So given your refusal to censor or ban foreign dictators while regularly censoring the President, aren't you at this very moment personally responsible for flooding the Nation with foreign disinformation?

Mr. DORSEY. Just to be clear, we have not censured the President. We have not taken the tweets down that you are referencing. They have more context in the label applied to them. And we do the same for leaders around the world.

Senator BLACKBURN. OK. Let me ask you this. Do you share any of your data mining, and this is to each of the three of you, do you share any of your data mining with the Democrat National Committee?

Mr. DORSEY. I am not sure what you mean by the question, but we have a data platform that we have a number of customers. I am not sure of the customer list.

Senator BLACKBURN. OK. And you said you don't keep lists. I make that note.

Mr. DORSEY. Well we keep a list of accounts that we watch—we don't keep lists of accounts that we watch.

Senator BLACKBURN. OK. Alright. OK. Mr. Pichai, is a Blake Lemoine, one of your engineers still working with you?

Mr. PICHAI. Senator, I am familiar with this name as a name, as an employer—I am not sure that he is currently an employee.

Senator BLACKBURN. OK. Well, he has had very unkind things to say about me. And I was just wondering if you all had still kept him working there. Also, I want to mention with you, Mr. Pichai, the way you all have censored some things. Google searches for Joe Biden generated approximately 30,000 impressions for Breitbart links. This was on May 1. And after May 5, both the impressions and the clicks went to zero. I hope that what you all realize from this hearing is that there is a pattern. You may not believe it exist, but there is a pattern of subjective manipulation of the information that is available to people from your platforms.

What has driven additional attention to this is the fact that more of a family's functional life is now being conducted online. Because of this, more people are realizing that you are picking winners and losers. You are trying to—Mr. Zuckerberg, years ago you said Facebook functioned more like a Government than a company. And you are beginning to insert yourself into these issues of free speech. Mr. Zuckerberg, with my time that is left, let me ask you this. You mentioned early in your remarks that you saw some things as competing equities. Is the First Amendment a given right or is that a competing equity?

Mr. ZUCKERBERG. I believe strongly in free expression. Sorry if I was on mute there. But I do think that, like all equities, it is balanced against other equities like safety and privacy. And even the people who believe in the strongest possible interpretation of the First Amendment still believe that there should be some limits on speech when it could cause imminent risk of physical harm. That the kind of famous example that is used is that you can't shout fire in a crowded theater. So I think that getting those equities in the balance right is the challenge that we face.

Senator BLACKBURN. My time has expired.

The CHAIRMAN. The time has expired, Perhaps we can——

Senator BLACKBURN. Well, we believe in the First Amendment and we are going to—yes, we will have questions to follow up. Thank you, Mr. Chairman. I can't see the clock.

The CHAIRMAN. Thank you. Senator Udall.

STATEMENT OF HON. TOM UDALL,
U.S. SENATOR FROM NEW MEXICO

Senator UDALL. Mr. Chairman, thank you, and Senator Cantwell, really appreciate this hearing. I want to start by laying out three facts. The U.S. intelligence community has found that the Russian Government is intent on election interference in the United States. They did it in 2016. They are doing it in 2020. The intelligence also says they want to help President Trump. They did so in 2016. The President doesn't like this to be said but it is a fact. We also know that the Russian strategy this time around is going after Hunter Biden. So I recognize that the details of how to handle misinformation on the Internet are tough. But I think that companies like Twitter and Facebook that took action to not be a part of a suspected Russian election interference operation were doing the right thing. And let me be clear, no one believes these companies represent the law or represent the public.

When we say work the refs, the U.S. Government is the referee. The FCC, the Congress, the Presidency and the Supreme Court are

the referees. It is very dangerous for President Trump. Justice Thomas and Republicans in Congress and at the FCC to threaten new Federal laws in order to force social media companies to amplify false claims, two, conspiracy theories, and disinformation campaigns. And my question to all three of you, do the Russian Government and other foreign nations continue to attempt to use your company's platforms to spread disinformation and influence the 2020 election, can you briefly describe what you are seeing? Please start, Mr. Dorsey and then Mr. Pichai. And Mr. Zuckerberg, you gave an answer partially on this. I would like you to expand on that answer. Thank you.

Mr. DORSEY. Yes. So we do continue to see interference. We recently disclosed actions we took on both Russia and actions originating out of Iran. We made those disclosures public. We can share those with your team. But this remains, as you have heard from others on the panel, and as Mark has detailed, one of our highest priorities and something we want to make sure that we are focused on eliminating as much platform manipulation as possible.

Mr. PICHAI. Senator, we do continue to see coordinated influence operation at times. We have been very vigilant. We appreciate the cooperation we get from intelligence agencies and companies. We are sharing information, to give you an example. And we publish transparency reports. In June, we identified efforts, one from Iran, a group, PAD 35, targeting the Trump campaign, one from China, a group, PAD 31, targeting the Biden campaign. Most of those were phishing attempts by our spam filters. We are able to remove most of the e-mails out from reaching users, but we notified intelligence agencies and that is an example of the kind of activity we see. And, you know, I think it is an area where we would need strong cooperation with Government agencies moving forward.

Mr. ZUCKERBERG. Senator like Jack and Sundar, we also see continued attempts by Russia and other countries, especially Iran and China, to run these kind of information operations. We also see an increase in kind of domestic operations around the world. Fortunately, we have been able to build partnerships across the industry, both with the companies here today and with law enforcement and the intelligence community to be able to share signals to identify these threats sooner. And along the lines of what you mentioned earlier, you know, one of the threats that the FBI has alerted our companies and the public to, was the possibility of a hack and leak operation in the days or weeks leading up to this election.

So you have both public testimony from the FBI and in private meetings, alerts that were given to at least our company, I assume the others as well, that suggested that we be on high alert and sensitivity, that if a trove of documents appeared, that we should view that with suspicion that it might be part of a foreign manipulation attempt. So that is what we are seeing. And I am happy to go into more detail as well if that is helpful.

Senator UDALL. Thank you very much. And this one is a really simple question. I think a yes or no. Will you continue to push back against this kind of foreign interference, even if powerful Republicans threaten to take official action against your companies? Mr. Zuckerberg, why don't we start with you and work the other way back?

Mr. ZUCKERBERG. Senator, absolutely. This is incredibly important for our democracy and we are committed to doing this work.

Mr. PICHAI. Senator, absolutely. Protecting our civic and democratic process is fundamental to what we do and we will do everything we can.

Mr. DORSEY. Yes, and we will continue to work and push back on any manipulation of the platform.

Senator UDALL. Thank you for those answers. Mr. Zuckerberg, do Facebook and other social media networks have an obligation to prevent disinformation and malicious actors spreading conspiracy theories, dangerous health disinformation, and hate speech, even if preventing its spread means less traffic and potentially less advertising revenue for Facebook?

Mr. ZUCKERBERG. Senator, in general, yes. I think that for foreign countries trying to interfere in democracy, I think that that is a relatively clear cut question where I would hope that no one disagrees that we don't want foreign countries or Governments trying to interfere in our elections, whether through disinformation or fake accounts or anything like that. Around health misinformation, you know, we are in the middle of a pandemic. It is a health emergency. I certainly think that this is a high sensitivity time. So we are treating with extra sensitivity any misinformation that could lead to harm around COVID that would lead people to not get the right treatments or to not take the right security precautions. We do draw a distinction between harmful misinformation and information that's just wrong. And we take a harder line and more enforcement against harmful misinformation.

The CHAIRMAN. Thank you, Senator—.

Senator UDALL. Thank you, Mr. Chairman.

The CHAIRMAN. Thank you, Senator Udall. Senator Capito.

STATEMENT OF HON. SHELLEY MOORE CAPITO, U.S. SENATOR FROM WEST VIRGINIA

Senator CAPITO. Thank you, Mr. Chairman, and thank all of you for being with us today. I would say that any time that we can get the three of you in front of the American people, whether it is several days before the election or several days after, is extremely useful and can be very productive. So I appreciate the three of you coming in the Committee holding this hearing. As we have heard, Americans turn every day to your platforms for a lot of different information.

I would like to give a shout out to Mr. Zuckerberg, because the last time, he was in front of our Committee, I had asked him to share the plenty of Facebook into rural America and help us with our fiber deployments into rural America. And when we see in this COVID environment, we see how important that is. And he followed through with that. I would like to thank him and his company for helping partner with us in West Virginia to get more people connected. And I think that is an essential—I would make a suggestion as well. Maybe when we get to the end, when we talk about fines.

What I think we could do with these millions and billion dollar fines that some of your companies that have been penalized on, we could make a great jump and get to that last household. But the

topic today is on objectionable content and how you make those judgments. So quickly each one of you, I know that in the Section 230, it says that if the term is "objectionable content" or "otherwise objectionable," would you be in favor of redefining that more specifically? That is awful broad. And that is where I think some of these questions become very difficult to answer. So we will start with Mr. Dorsey on the how do you define "otherwise objectionable" and how can we improve that definition so that it is easier to follow?

Mr. DORSEY. Well, our interpretation of objectional is anything that is limiting potentially the speech of others. However, policies are focused on making sure that people feel safe to express themselves. And when we see abuse, harassment, misleading information, these are all threats against us and that makes people want to leave the Internet. It makes people want to leave these conversations online. So that is what we are trying to protect, is making sure that people feel safe enough and free enough to express themselves in whatever way they wish.

Senator CAPITO. So this is a follow up to that. Much has been talked about the blocking of the *New York Post.* Do you have an instance or, for instance, of when you have actually blocked somebody that would be considered politically liberal on the other side of the political realm in this country? Do you have an example of that to sort of offset where the *New York Post* criticism has come from?

Mr. DORSEY. Well, we don't have an understanding of the ideology of any one particular account, and that is also not our policies or our enforcement taken. So I am sure there are a number of examples. But that is not our focus. We are looking purely for violations of our policies, taking action against it.

Senator CAPITO. Yes. Mr. Zuckerberg, how would you define "otherwise objectionable"?—not how would you define it, but how would you define the definition of that to make it more objective than subjective?

Mr. ZUCKERBERG. Senator, thank you. When I look at the written language in Section 230 and the content that we think shouldn't be allowed on our services, some of the things that we bucket in otherwise objectionable content today include general bullying and harassment of people on the platform. So somewhat similar to what Jack was just talking about a minute ago. And I would worry that some of the proposals that suggest getting rid of the phrase otherwise objectionable from Section 230 would limit our ability to remove bullying and harassing content from our platforms, which I think would make them worse places for people. So I think we need to be very careful in how we think through that.

Senator CAPITO. Well, thank you. Mr. Pichai?

Mr. PICHAI. Senator. Maybe what I would add is that the content is so dynamic. YouTube gets 500 hours per minute of video uploaded on an average of any day search. 15 percent of queries we have never seen before. To give me an example, a few years ago there was an issue around teenagers consuming Tide Pods and it was a kind of issue which was causing real harm. When we had run into those situations, we were able to act with certainty and protect our users. The Christchurch shooting where there was a

live shooter, you know, live-streaming, horrific images. It was a learning moment for all our platforms. We were able to intervene again with certainty. And so that is what otherwise objectionable allows. And, you know, I think I think that flexibility is what allows us to focus. We always state with clear policies what we are doing. But I think it gives platforms of all sizes, flexibilities to protect our users.

Senator CAPITO. Thank you. I think I am hearing from all three of you, really, that the definition is fairly acceptable to you all. In my view, sometimes I think it can go too much to the eye of the beholder—the beholder being either a you all or your reviewers or your AI, and then it gets into the region where maybe you become so, so very subjective. I want to move to a different topic, because in my personal conversations with at least two of you, you have expressed the need to have the 230 protections because of the protections that it gives to the small innovators. Well, you sit in front of us and I think all of us are wondering how many small innovators and what kind of market share could they possibly have when we see the dominance of the three of you. So I understand you started as small innovators when you first started. I get that. How can a small innovator really break through? And what does 230 really have to do with the ability—I am skeptical on the argument, quite frankly. So whoever wants to answer that, Mr. Zuckerberg, do you want to start?

Mr. ZUCKERBERG. Sure, Senator. I do think that when we were getting started with building Facebook, if we were subject to a larger number of content lawsuits because 230 didn't exist, that would have likely made it prohibitive for me as a college student in a dorm room to get started with this enterprise. And I think that it may make sense to modify 230 at this point just to make sure that it is still working as intended, but I think it is extremely important that we make sure that that for smaller companies that are getting started, the cost of having to comply with any regulation is either waived until a certain scale or is, at a minimum, taken into account as a serious factor to make sure that we are not preventing the next set of ideas from getting built.

The CHAIRMAN. Thank you. Thank you, Senator.

Senator CAPITO. Thank you, Mr. Chairman.

The CHAIRMAN. Thank you. Senator Baldwin.

STATEMENT OF HON. TAMMY BALDWIN,
U.S. SENATOR FROM WISCONSIN

Senator BALDWIN. Thank you. I would like to begin by making two points. I believe the Republicans have called this hearing in order to support a false narrative fabricated by the President to help his reelection prospects. And number two, I believe that the tech companies here today need to take more action, not less, to combat misinformation, including misinformation on the election, misinformation on the COVID–19 pandemic, and misinformation and posts meant to incite violence. And that should include misinformation spread by President Trump on their platforms. So I want to start with asking the Committee Clerk to bring up my first slide. Mr. Dorsey, I appreciate the work that Twitter has done to flag or even take down false or misleading information about COVID–19,

such as this October 11th tweet by the President claiming he has immunity from the virus after contracting it and recovering, contrary to what the medical community tells us. Just yesterday morning, the President tweeted this, that the media is incorrectly focused on the pandemic and that our Nation is, "rounding the turn on COVID–19." In fact, according to Johns Hopkins University in the past week, the 7-day national average of new cases reached its highest level ever. And in my home State of Wisconsin, case counts continue to reach record levels. Yesterday, Wisconsin set a new record with 64 deaths and 5,462 new confirmed cases of COVID–19. That is not rounding the turn. But it is also not a tweet that was flagged or taken down. Mr. Dorsey, given the volume of misleading posts about COVID–19 out there, do you prioritize removal based on something like the reach or audience of a particular alert user of Twitter?

Mr. DORSEY. I could be mistaken, but it looks like the tweet that you showed actually did have a label pointing to both of them, pointing to our COVID resource in our interface. So we, with in regards to misleading information, we have policies against manipulating media in support of public health and COVID information. And election interference and civic integrity. And we take action on it. In some cases, it's labeling. In some cases, it is removal.

Senator BALDWIN. What additional steps are you planning to take to address dangerously misleading tweets like the President's rounding the turn tweet?

Mr. DORSEY. We want to make sure that we are giving people as much information as possible and that ultimately we are connecting the dots. When they see information like that, that they have an easy way to get an official resource or many more viewpoints on what they are seeing. So we will continue to refine our policy. We will continue to refine our enforcement around misleading information. And we are looking deeply at how we can evolve our product to do the same.

Senator BALDWIN. Mr. Zuckerberg, I want to turn to you to talk about the ongoing issue of right-wing militias using Facebook as a platform to organize and promote violence. Could the Committee Clerk please bring up my second slide? On August 25, a self-described militia group called Kenosha Guard created a Facebook event page entitled "Armed Citizens to Protect Our Lives and Property," encouraging armed individuals to go to Kenosha and, "defend the city during a period of civil unrest following the police shooting of Jacob Blake" That evening, a 17 year old from Illinois did just that and ended up killing two protesters and seriously injuring a third. Commenters in this group wrote that they wanted to kill looters and rioters and switch to real bullets and put a stop to these rioting impetuous children.

While Facebook has already had a policy in place banning militia groups, this page remained in place. According to press reports, Facebook received more than 450 complaints about this page, but your content moderators did not remove it, something you subsequently called an operational mistake. Recently, as you heard earlier in questions, the alleged plot to kidnap Michigan Governor Gretchen Whitmer and the potential for intimidation or even violence at voting locations show that the proliferation of the threat

of violence on Facebook remains a very real and urgent problem. Mr. Zuckerberg, in light of the operational mistake around Kenosha, what steps has Facebook taken to ensure that your platform is not being used to promote more of this type of violence?

Mr. ZUCKERBERG. Thank you, Senator. This is a big area of concern for me personally and for the company. We have strengthened our policies to prohibit any militarized social movement. So any kind of militia like this. We have also banned conspiracy network. So, Qanon being the largest example of that. That is completely prohibited on Facebook at this point, which, you know, in this period where I am personally—I am worried about the potential of increased civil unrest, making sure that those groups can't organize on Facebook. It may cutoff some legitimate uses, but I think that they will also preclude greater potential for organizing any harm. And by making the policy simpler, we will also make it so that there are fewer mistakes in content moderation. So I feel like we are in a much stronger place on the policies on this at this point.

The CHAIRMAN. Thank you, Senator Baldwin. Senator Lee.

STATEMENT OF HON. MIKE LEE,
U.S. SENATOR FROM UTAH

Senator LEE. Thank you very much, Mr. Chairman. I want to read a few quotes from each of you, each of our three witnesses, and from your companies. And then I may ask for a response. So, Mr. Zuckerberg, this one is from you. You said, "we have built Facebook to be a platform for all ideas. Our community's success depends on everyone feeling comfortable sharing what they want. It doesn't make sense for our mission or for our business to suppress political content or prevent anyone from seeing what matters most to them." You said that, I believe on May 18, 2016. Mr. Dorsey, on September 5, 2018, you said, "let me be clear about one important and foundational fact, Twitter does not use political ideology to make any decisions."

Mr. Pichai, on October 28, 2020, you said, "let me be clear, we approach our work without political bias." Now, these quotes make me think that there is a good case to be made that you are engaging in unfair or deceptive trade practices in violation of Federal law. I see these quotes where each of you tell consumers and the public about your business practices. But then you seem to do the opposite and take censorship related actions against the President, against members of his Administration, against the *New York Post,* the *Babylon Bee,* the Federalist, pro-life groups, and there are countless other examples. In fact, I think the trend is clear that you almost always censor and when I use the word censor here, I mean block content, fact check or label content or demonetize websites of conservative, Republican, or pro-life individuals or groups or companies contradicting your commercial promises. But I don't see this suppression of high profile liberal commentators.

So, for example. Have you ever censored a Democratic Senator? How about President Obama? How about a Democratic Presidential candidate? How about Planned Parenthood or NARAL or Emily's List? So Mr. Zuckerberg, Mr. Dorsey and Mr. Pachai, can any of you, and let's go in that order, Zuckerberg, Dorsey, and then Pichai, can you name for me one high profile person or entity from

a liberal ideology who you have censored and what particular action you took?

Mr. ZUCKERBERG. Senator, I can get you a list of some more of this, but there are certainly many examples that your Democratic colleagues object to when, you know, the fact checker might label something as false they disagree with or they are not able to—.

Senator LEE. I get that. I get that. I just want to be clear. I am just asking you if you can name for me one high-profile liberal person or company who you have censored. I understand that you are saying that there are complaints on both sides, but I just I just want one name of one person or one entity.

Mr. ZUCKERBERG. Senator, I need to think about it and get you more of a list. But there are certainly many, many issues on both sides of the aisle where people think we are we are making content moderation decisions that they disagree with.

Senator LEE. I got that. And I think everybody on this call could agree that they could identify at least five, maybe 10, maybe more high profile conservative examples. And what about you, Mr. Dorsey?

Mr. DORSEY. Well, we can give a more exhaustive list, but again, we don't have an understanding of political ideologies of our accounts——

Senator LEE. But I am not asking for an exhaustive list. I am asking for a single example. One, just one individual, one entity, anyone?

Mr. DORSEY. We have—we have taken action on tweets from members of the House for Election misinfo.

Senator LEE. Can you identify an example?

Mr. DORSEY. Yes, with two Democratic congresspeople.

Senator LEE. What are their names?

Mr. DORSEY. I will get those names to you.

Senator LEE. Great, great. Mr. Pichai, how about you?

Mr. PICHAI. Senator, I will give specific examples, but let me step back. We don't censor. We have moderation policies which apply equally. To give you an example——

Senator LEE. I get that. I used the word censor as a term and I defined that term. And again, I am not asking for a comprehensive list. I want a name——

Mr. PICHAI. We have done ads from Priorities USA, from Vice President Biden's campaign. We have had compliance issues with the World Socialist Review, which is a left leaning publication. I can give you several examples. But for example, when we have a graphic content policy, we don't allow for ads which show graphic violent content in those ads. And we have taken down ads on both sides of the campaign. And I gave you a couple of examples.

Senator LEE. OK. At least with respect to Mr. Zuckerberg and Mr. Dorsey. And I would point out that with respect to Mr. Pichai, those are not nearly as high profile. I don't know if I can identify anyone picked at random from the public even and picked at random from the public as far as members of the political active community in either political party who could identify those right off the top—of the bat. Look, there is a disparity between the censorship, and again, I am using that as a term of art, as I did find it a moment ago, between the censorship of conservative and liberal

points of view. It is an enormous disparity. Now you have the right, and I want to be very clear about this, you have every single right to set your own terms of service and to interpret them and to make decisions about violations. But given the disparate impact of who gets censored on your platforms, it seems that you are either one, not enforcing your terms of service equally or alternatively two, that you are writing your standards to target conservative viewpoints.

You certainly have the right to operate your own platform. But you also have to be transparent about your actions, at least in the sense that you can't promise certain corporate behavior, and then deceive customers through contradictory actions that just blatantly contradict what you have stated as your corporate business model or as your policy. So, Mr. Zuckerberg and Mr. Dorsey, if Facebook is still a platform for all ideas and if Twitter, "does not use political ideology to make decisions," then do you state before this committee, before the record, that you always apply your terms of service equally to all of your users?

Mr. ZUCKERBERG. Senator, our principle is to stand for free expression and to be a platform for all ideas. I certainly don't think we have any intentional examples where we are trying to enforce our policies in a way that is anything other than fair and consistent. But it's also a big company. So I get that there are probably mistakes that are made from time to time. But our North Star and what we intend to do is to be a platform for all ideas and to give everyone a voice.

Senator LEE. I appreciate that. I understand what you are saying, but intentional examples of a big company. But again, there is a disparate impact. There is a disparate impact that's unmistakable as evidenced by the fact that neither you nor Jack could identify a single example. Mr. Dorsey, how do you answer that question?

The CHAIRMAN. A brief answer, please, Mr. Dorsey.

Mr. DORSEY. Yes, we operate our enforcement and our policy without an understanding of political ideology. We don't—anytime we find examples of bias in how people operate our systems or our algorithms, we remove it. And as Mark mentioned, there are checkpoints in these companies and in these frameworks, and we do need more transparency around them and how they work. And we do need a much more straightforward and quick and efficient appeals process to give us a further checkpoint from the public.

The CHAIRMAN. Thank you, Senator Lee. Senator Duckworth.

STATEMENT OF HON. TAMMY DUCKWORTH, U.S. SENATOR FROM ILLINOIS

Senator DUCKWORTH. Thank you, Mr. Chairman. You know, I have devoted my life to public service, to upholding a sacred oath through support and defend the Constitution of the United States against all enemies foreign and domestic. And I have to be honest, it makes my blood boil and it also breaks my heart a little as I watch my Republican colleagues just days before an election sink down to the level of Donald Trump. By placing the selfish interests of Donald Trump ahead of the health of our democracy, Senate Re-

publicans, whether they realize it or not, are weakening our national security and providing aid to our adversaries.

As my late friend Congressman Cummings often reminded us, you know, we are better than this. Look, our democracy is under attack right now. Every American, every Member of Congress should be committed to defending the integrity of our elections from hostile foreign interference. Despite all the recent talk of great power competition, our adversaries know they still cannot defeat us on a conventional battlefield. Yet meanwhile, the members of the United States military and our dedicated civil servants are working around the clock in the cyber domain to counter hostile actors such as Iran, China and Russia. And they do this while the Commander in Chief cowers in fear of Russia and stubbornly refuses to take any action to criticize or warn Russia against endangering our troops.

I have confidence in the United States armed forces, intelligence community and civil servants. Their effective performance explains why our foreign adversaries have sought alternative avenues to attacking our Nation. Afraid to face us in conventional, military or diplomatic ways, they look for unconventional means to weaken our democracy, and they realize that social media could be the exhaust port of our democracy. Social media is so pervasive in the daily lives of Americans and traditional media outlets that it can be weaponized to manipulate the public discourse and destabilize our institutions.

You know, after Russia was incredibly successful in disrupting our democracy 4 years ago, all of our adversaries learned a chilling lesson, social media companies cannot be trusted to put patriotism above profit. Facebook and Twitter utterly failed to hinder Russia's sweeping and systemic interference in our 2016 election, which used the platforms to infiltrate our communities, spread disinformation and turn Americans against one another. Of course, the situation has grown far worse today, as evidenced by today's partisan sham hearing. While corporations may plead ignorance prior to the 2016 election, President Trump and his Republican enablers in the Senate have no such excuse.

Senate Republicans cut a deal to become the party of Trump, and now they find themselves playing a very dangerous game by encouraging Russia's illegal hacking, by serving as the spreaders and promoters of disinformation cooked up by foreign intelligence services, and by falsely claiming censorship. When responsible actors attempt to prevent hostile foreign adversaries from interfering in our elections, Senate Republicans insult the efforts of true patriots working to counter malign interference and weaken our security. This committee is playing politics at a time when responsible public officials should be doing everything to preserve confidence in our system of elections and system of Government. The reckless actions of Donald Trump and Senate Republicans do not let technology companies off the hook.

None of the companies testifying before our committee today are helpless in the face of threats to our democracy, small d democracy. Federal law provides you respective companies—Federal laws provides your respective companies with authority to counter foreign disinformation and counterintelligence propaganda. And I want to

be absolutely clear, gentlemen, that I fully expect each of you to do so. Each of you will be attacked by the President, Senate Republicans and right wing media for countering hostile foreign interference in our election. But you have to do a duty to do the right thing because facts still exist.

Facts still matter. Facts save lives. And there is no both sides when one side has chosen to reject truth and embrace poisonous false information. So in closing, I would like each witness to provide a personal commitment that your respective companies will proactively counter domestic disinformation that spreads the dangerous lie such as "masks don't work", while aggressively identifying and removing disinformation that is part of foreign adversaries efforts to interfere in our election or undermine our democracy. Do I have that commitment from each of you gentlemen?

The CHAIRMAN. OK, Will. We will take Dorsey, Pichai, and then Zuckerberg. Mr. Dorsey.

Mr. DORSEY. We make that commitment.

The CHAIRMAN. Mr. Pichai.

Mr. PICHAI. Senator, absolutely, yes.

The CHAIRMAN. And Mr. Zuckerberg.

Mr. ZUCKERBERG. Yes, Senator. I agree with that.

Senator DUCKWORTH. Thank you. Your industry success or failure in achieving this goal will have far reaching life or death consequences for the American people and the future of our democracy. Thank you and I yield back, Mr. Chairman.

The CHAIRMAN. The Senator yields back. Senator Johnson.

STATEMENT OF HON. RON JOHNSON,
U.S. SENATOR FROM WISCONSIN

Senator JOHNSON. I would like to start with a question for all three of the witnesses. You know, we have public reports that you have different chat forums in your companies and also public reports where, you know, the few conservatives that might work for your companies have certainly been harassed on those types of forums. I don't expect you to have taken a poll of your employees but I just want to get a kind of a sense, because I think it is pretty obvious, but would you say that the political ideology of the employees in your company is, you know, let's say 50/50, conservative versus liberal, progressive? Or do you think it's closer to 90 percent liberal, 10 percent conservative? We will start with Mr. Dorsey.

Mr. DORSEY. As you mentioned, I don't know the makeup of our employees because it is not something we ask or focus on.

Senator JOHNSON. Just what do you think off the top of your head based on your chat rooms and kind of people you talk to?

Mr. DORSEY. Not something I look for or——

Senator JOHNSON. Right. Mr. Pichai.

Mr. PICHAI. Senator, we have over 100,000 employees. For the past two years, we have hired greater than 50 percent of our workforce outside California. It does tend to be proportionate to the areas where we are in but we do have a have a million message boards at Google. We have groups like Republicans side, liberals side, conservative side, and so on. And we have definitely made an effort to make sure people of all viewpoints are welcome.

Senator JOHNSON. So, again, you and Mr. Zuckerberg, you answered the question honestly. Is it 90 percent or 50/50, which is closer to——

Mr. ZUCKERBERG. Senator, I don't know the exact number, but I would guess that our employee base skews left leaning.

Senator JOHNSON. Thank you for your honesty. Mr. Dorsey, you started your opening comments that, you know, you think that people don't trust you. I agree that, we don't trust you. You all say you are fair and you are consistent. You are neutral. You are unbiased. Mr. Dorsey, I think the most incredible answer I have seen so far in this hearing is when Senator Cruz asked, does Twitter have the ability to influence elections? Again, does Twitter have the ability to influence elections? You said no. Did you stick to that answer that you don't even believe—let's face it, you all believe that Russia has the ability to influence the elections or interfere by using your social platforms. Mr. Dorsey, you still deny that you don't have the ability to influence and interfere in our elections?

Mr. DORSEY. Yes, I mean my answer was around of people's choice around other communication channels.

Senator JOHNSON. No, your answer was—the question was, does Twitter have the ability to influence elections? You said no. Do you still stand by that? That doesn't translate?

Mr. DORSEY. Twitter is a company, no.

Senator JOHNSON. You don't think you have the ability by moderation policies, by what I would call censoring. You know, what you did to the *New York Post.* You don't think that censorship, that moderation policies, you don't think that influences elections by withholding what I believe is true information for American public? You don't think that interferes in elections?

Mr. DORSEY. Not our current moderation policies. Our current moderation policies are to protect the conversation and the integrity of the conversation around the elections.

Senator JOHNSON. OK. For both Mr. Zuckerberg and Dorsey, who censored, censored *New York Post* stories or throttled them back either way, did you have any evidence that the *New York Post* story is part of Russian disinformation or that those e-mails aren't authentic? Did any of you have any information whatsoever that they are not authentic or that they are Russian disinformation? Mr. Dorsey.

Mr. DORSEY. We don't.

Senator JOHNSON. You don't know? So why would you censor it? Why did you prevent that from being disseminated on your platform? It is supposed to be for the free expression of ideas and particularly true ideas.

Mr. DORSEY. We believe—of hacking materials policy. We judged——.

Senator JOHNSON. But what evidence of—it was hacked? They weren't hacked.

Mr. DORSEY. We judged at the moment that it looked like it was hacked material——

Senator JOHNSON. You were wrong.

Mr. DORSEY.—surfacing, and we updated our policy and our enforcement within 24 hours.

Senator JOHNSON. Mr. Zuckerberg?

Mr. ZUCKERBERG. Senator, as I testified before, we relied heavily on the FBI's intelligence and alert staffs, both through their public testimony and private briefings——

Senator JOHNSON. Did the FBI contact you and say the *New York Post* story was false?

Mr. ZUCKERBERG. Senator, not about that story specifically—.

Senator JOHNSON. Why did you throttle it back?

Mr. ZUCKERBERG. They alerted us to be on heightened alert around a risk of hack and leak operations around a release of information—and Senator, to be clear on this, we didn't censor the content. We flagged it for fact checkers to review and pending that review, we temporarily constrained its distribution to make sure that it didn't spread wildly while it was being reviewed. But it is not up to us either to determine whether it is Russian interference nor whether it is true. We rely on the FBI and intelligence and fact checkers to do that——

Senator JOHNSON. Fine. Mr. Dorsey, you talked about your policies toward misinformation and that you will block misinformation if it is about civic integrity, election interference or voter suppression. Let me give you a tweet that was put up on Twitter.

It says, "Senator Ron Johnson is my neighbor and strangled our dog, Buttons, right in front of my 4 year old son and 3 year old daughter." The police refused to investigate. This is a complete lie but important to retweet and note that there are more lies to come. Now, we contacted Twitter and we asked them to take it down. Here's the response. "Thanks for reaching out. We escalated this to our support team for their review, and they have determined that this is not a violation of our policies."

So, Mr. Dorsey, how could a complete lie—it is admitted it is a lie, how does that not affect civic integrity? How could you view that not as being election interference? That could definitely impact my bill to get reelected. How could that not be a violation of voter suppression? Obviously, if people think I am strangling my neighbor's dog, they may not show up at the polls. That would be voter suppression. So why didn't you take that—by the way, that tweet was retweeted like some suddenly 17,000 times and viewed by over and loved commented and appreciated by over 50,000 people. How is that not voter suppression? How is that not election interference? How does that not affect civic integrity?

Mr. DORSEY. We will have to look into our enforcement or non-enforcement in this case for the tweet and we can get back to you with more context.

Senator JOHNSON. So Mr. Zuckerberg, in that same June hearing—real quick, Mr. Dorsey, you referred to that June hearing with Stephan Wolfgang had all kinds of good ideas. That was 16 months ago. Why haven't you entered—why haven't you implemented any of those transparency ideas you thought were pretty good 16 months ago?

Mr. DORSEY. Well, he was talking about algorithm choice, and we have implemented one of them, which is we allow people to turn off the ranking over a timeline. The rest is work and it is going to take some time.

Senator JOHNSON. So I would get to it if I were you. Thank you, Mr. Chairman.

The CHAIRMAN. Senator Johnston, thank you. Let me just make sure I understood the answer. Mr. Dorsey and Mr. Zuckerberg, Mr. Dorsey did I understand you to say that you have no information indicating that the *New York Post* story about Hunter Biden has a Russian source? Did I understand correctly?

Mr. DORSEY. Yes. Not that I am aware of.

The CHAIRMAN. And is that also your answer, Mr. Zuckerberg, that you have no information at all to indicate that Russia was the source of this *New York Post* article?

Mr. ZUCKERBERG. Senator, I would rely on the FBI to make that assessment.

The CHAIRMAN. But you don't have any such information, do you?

Mr. ZUCKERBERG. I do not myself.

The CHAIRMAN. We are just trying to clarify the answer to Senator Johnson's question. Thank you very much for indulging me there. Senator Tester, you are next, sir.

STATEMENT OF HON. JON TESTER,
U.S. SENATOR FROM MONTANA

Senator TESTER. I want to thank you, Mr. Chairman. And I want to thank Sundar and Jack and Mark for being in front of this committee. There is no doubt that there are two major issues with Google and Facebook and Twitter that Congress needs to address. Quite frankly, big tech is the unregulated Wild West that needs to be held accountable. And we do need to hear from all three of you about a range of critical issues that Americans deserve answers, online data privacy, anti-trust, the proliferation of misinformation on your platforms. In a moment, I am going to ask all of you to commit to returning to this committee early next year to have a hearing on these important issues. But the truth is, my Republican colleagues arranged this hearing less than a week from Election Day for one specific reason, to make a last ditch case based on shoddy evidence that these companies are censoring conservative voices. It is a stunt and it is a cheap stunt at that.

It is crystal clear that this hearing is designed to cast doubt on the fairness of the upcoming election and to work with the platforms to allow bad information to stay up as November 3 approaches. It is also crystal clear that the directive to hold this political hearing comes straight from the White House. And it is a sad day when the U.S. Senate, an equal part of an independent branch of Government, allows the Senate's halls to be used for the President's political stance. There is a national election in six days, Mr. Chairman. You have nearly two years to hold this hearing. It is happening six days before the election.

The idea that we should have a somber hearing about putting the reins on big tech six days before the election, quite frankly, doesn't pass the smell test. Today, this hearing is about electoral politics. I know it. You know it, everybody in this room knows it. And I know the American people are smart enough to figure that out. I am going to talk a little more about that in a second. But first, I want to thank the panel once again for being here.

And I will start by asking question about making a more sincere effort to discuss the issues that surround big tech down the road. So the question for the panel and yes is a yes or no answer. Will

you commit to returning to testify again in the new Congress? Let's start with you, Jack.

Mr. DORSEY. Yes, we are always happy—myself or teammates—to talk with the American people.

Senator TESTER. Sundar.

Mr. PICHAI. Senator, yes, we have engaged many times and we are happy to continue that engagement with Congress.

Senator TESTER. How about you, Mark?

Mr. ZUCKERBERG. Senator, yes. I hope we can continue to have this conversation and hopefully not just with the CEOs of the companies, but also with experts who work on these issues every day as is part of their jobs.

Senator TESTER. Absolutely. I think the more information, the better, but not based on politics, based on reality. And I want to thank you for that, because we are in a very unreal time when it comes to politics. Quite frankly, we are in a time when fake news is real and real news is fake. And you guys try to shut down the fake news, whether it comes from Joe Biden's smile or whether it comes from Donald Trump's mouth. And the fact is, if Joe Biden said suddenly crazy and offensive stuff that the President has said, you would get fact checked in the same way. Wouldn't you agree? You can nod your head to that. Wouldn't you agree if Joe Biden said the same stuff that Trump said that you would do the same sort of fact checking on him?

The CHAIRMAN. Shall we take on Mr. Dorsey, Mr. Pichai and Mr. Zuckerberg in that order?

Mr. DORSEY. If we found violations of our policy, we would do the appropriate enforcement action.

Senator TESTER. Thank you.

The CHAIRMAN. Just go ahead then, Mr. Pichai.

Mr. PICHAI. Senator, yes, we would apply our policies without regard to who is strong and be applied neutrally.

Senator TESTER. OK, thank you. Mark.

Mr. ZUCKERBERG. Senator, I agree with what Jack and Sundar said, we would also apply our policies to everyone. And in fact, when Joe Biden tweets or posts and cross posts to Facebook about the election, we put the same label, adding context about voting on his post as we do for other candidates.

Senator TESTER. Thank you for that. In 2016, Russians built a network of bots and fake accounts to spread disinformation. This year, it seems they are seeding your networks with disinformation and relying on Americans, including some folks in Congress, to amplify and distribute it. What tools do you have to fight foreign disinformation on your platforms when it is spread by Americans? Jack?

Mr. DORSEY. We are looking at—you know, our policies are against platform integration, period, no matter where it comes from. So whether that's foreign or domestic, we see patterns of people, organizations that attempt to manipulate platform in the conversation, artificially amplify information.

Senator TESTER. Thanks. Mark.

Mr. ZUCKERBERG. Senator, the efforts are a combination of AI systems that look for anomalous behavior by accounts or networks of accounts, a large human effort where we have 35,000 employees

who work on security and content review, and partnerships that we have made with the other tech companies here, as well as law enforcement and intelligence community and election officials across the world to make sure that we have all the appropriate input signals and can share signals on what we are seeing with the other platforms as well.

Senator TESTER. OK. Sundar?

Mr. PICHAI. Two things to add, Senator, to give it different examples. We partner with over 5,000 civic entities, campaign organizations, at the Federal and state level to protect their campaign's digital assets with advanced protection program and training. And others I would echo, there has been an enormous increase in cooperation between the tech companies. You know, as companies, we are sharing a lot of information and doing more together than ever before.

Senator TESTER. Thank you. I just want to close with one thing. We have learned a lot of information out here today where when you hire somebody you are supposed to ask them their political affiliation. You are supposed to ask them who they have donated to? They are supposed to be a political litmus test. If you hire a Biden person, you are supposed to hire a Trump person. Why not hire a test person? And let's talk about business. We want to regulate business, and if that business is won by a liberal, we are going to regulate them different than if they are run by a conservative outfit. That reminds me a lot of the Supreme Court where you have two sets of polls—one for a Democratic president and one for a Republican. This is baloney, folks. Get off the political garbage and let's have the commerce committee do its job. Thank you very much.

The CHAIRMAN. Thank you, Senator Tester. Senator Scott.

STATEMENT OF HON. RICK SCOTT,
U.S. SENATOR FROM FLORIDA

Senator SCOTT. Thank you, Chairman, for hosting this. I think, first off, if you followed this today, which you will clearly come to a conclusion is Republicans believe that you censor and Democrats think it is pretty good what you are doing. We are blessed to live in the United States. Our democracy where we are granted individual freedoms and liberties under the Constitution. This isn't the case around the world. We can look at what's happening in Communist China right now. General Secretary Xi is committing horrible human rights abuses against China's minority communities and censoring anyone that speaks out about their oppression. The Chinese Communist Party who surveils their citizens and uses state-run media to push through propaganda, control information their citizens consume, and hide their human rights abuses. Twitter and Facebook are banned in communist China, so you can understand why it is concerning to be even discussing issues that big technology companies are interfering with free speech.

The American people entrust your companies with their information. They believe that you will protect their information, allow them to use your platforms to express themselves freely. I don't think any one person has signed up for any of your platforms and expects to be blocked or kicked off because of their political views.

But it is becoming obvious that your companies are unfairly targeting conservatives. That is clearly the perception today. Facebook is actively targeting ads by conservative groups ahead of the election, either removing the ads completely or adding their own disclosure if they claim they didn't pass their fact checks system.

But their fact check is based on imports from known liberal media groups like PolitiFact, which clearly is a liberal media group. Twitter censored Senator Mitch McConnell and put warnings on several of the President's tweets. And until recently, they completely blocked the American people from sharing the *New York Post* story about Hunter Biden's laptop and suspended the *New York Post* account. The *New York Post* is one of the most circulated publications in the United States. This isn't some fringe media outlet filled with conspiracy theories. Yet you allowed murderous dictators around the world to freely use your platform. Let me give you a few examples.

On Twitter, Iran's supreme leader, Ayatollah, tweeted, calling for the elimination of the Zionist regime. He said on May 21, 2020, "the elimination of the Zionist regime does not mean the massacre of the Jewish people. The people of Palestine should hold a referendum. Any political system that they vote for should govern all Palestine. The only remedy until the removal of the Zionist regime is firm, armed resistance." I would like to know first why Twitter let that stay up and why the Ayatollah has not been blocked? In May 2019, Maduro, a murderous dictator, tweeted a photo of him backed by his military for a March after 3 people were killed and 130 injured during protests in his country.

The tweet describes the march as a clear demonstration of the moral strength and integrity of our glorious armed forces, which is always prepared to defend peace and sovereignty. I was saying this glorifies violence, which Twitter has flagged President Trump for, but Twitter let that stand jurisdiction's. Secretary Xi's communist regime stood by the fact that it is committing genocide against the Uyghurs, forcing millions into internment camps because of their religion.

On September 1, the Chinese Government account posted on Twitter, "Xinjiang 'camps' more fake news. What the Chinese Government has done in Xinjiang has created the possibility for the locals to lead better lives. But truth that simply goes against "anti-China narrative" will not report by some biased media." Clear lie. It has been widely reported that this claim by the Chinese Government is false, but Twitter took no action. Your companies are inconsistently applying the road rules with an obvious bias. Your companies are censoring free speech to target the President, the White House Press Secretary, Senator Mitch McConnell, the Susan B. Anthony List, a pro-life group, while giving dictators a free, unfettered platform. It is our responsibility to hold your companies accountable and protect Americans ability to speak freely on their platforms, regardless of their political views or the information they choose to share.

You can't just pick and choose which viewpoints are allowed on your platform and expect to keep the immunity granted by Section 230. So, Mr. Dorsey, you allow dangerous dictators on your platform. Tell me why you flag conservatives in America, like President

Trump or Leader McConnell, for potential misinformation while allowing dictators to spew their propaganda on your platform?

Mr. DORSEY. We have taken actions around leaders around the world and certainly with some dictators as well. We looked at—we look at the tweets, we review them and we figure out if they violated a policy of ours or not.

Senator SCOTT. Mr. Dorsey, can you tell me one you did against Iran, the Ayatollah? Can you tell me about one you have ever done against the Ayatollah or Maduro?

Mr. DORSEY. I think we have done more than one actually, but we can send you that information on those actions. But we, you know, we do have a global leader policy that we believe is important where people can see what these leaders are saying and those tweets remain up, but if they are labeled that they violated our service just to show the integrity of our policy and our enforcement.

Senator SCOTT. When the communists struck China, which we all know, put a million people, Uyghurs in camps, you did nothing about the tweet. And they said that they are just helping them lead a better life. I mean, we mean—you can just—anybody that follows the news knows what is happening to the Uyghurs. I mean it is genocide, what they are doing to the Uyghurs. I have never seen anything you've done on calling that lie.

Mr. DORSEY. We don't have a general policy around misleading information and misinformation. We don't. We rely upon people calling that speech out, calling those reports out in those ideas. And that is part of the conversation is if there is something found to be in contest, then people reply to it. People retweet it and say that this is wrong. This is obviously wrong. You would be able to quote, retweet that today and say that this is utterly wrong and we would benefit from more of those voices calling it out.

Senator SCOTT. So but you block Mitch McConnell and Trump's tweets and you just say, right. Here is what I don't get, is that you guys have set up policies that you don't enforce consistently. And then and then what is the recourse to a user? I talked to a lady this week. She has got her Facebook account just eliminated. There is no recourse. There is nothing she can do about it. So every one of you have these policies that you don't enforce consistently. So what should be the recourse?

Mr. DORSEY. We enforce them consistently and as I said in my opening remarks, we believe it is critical that we have more transparency around our process. We have clear and straightforward and efficient appeals. So the woman that you talked to could actually appeal the decision that we made. And that we focus on algorithms and figure out how to give people more choice.

The CHAIRMAN. Thank you, Senator Scott. Senator Rosen.

STATEMENT OF HON. JACKY ROSEN, U.S. SENATOR FROM NEVADA

Senator ROSEN. Thank you, Mr. Chairman. I appreciate the witnesses for being here today, and I want to focus a little bit, thank you Mr. Dorsey, on algorithms, because my colleagues on the Majority called this hearing in order to argue that you are doing too much to stop the spread of disinformation, conspiracy theories and hate speech on your platforms. I am here to tell you that you are

not doing enough. Your platform's recommendation algorithms drive people who show an interest in conspiracy theories far deeper into hate, and only you have the ability to change this. What I really want to say is that on these platforms and what I would like to tell my colleagues, the important factor to realize is that people or users are the initiators of this content and the algorithms are the potentiators, and particularly the recommendation algorithms, the potentiators of this content. Now, I was doing a little cleaning in my garage like a lot of people during COVID.

I am a former computer programmer. I actually found my old hexadecimal calculator and Radio Shack, my little owner's manual here. So I know a little bit about the power of algorithms and what they can and can't do, having done that myself. And I know that you have the ability to remove bigoted, hateful, and incendiary content that will lead and has led to violence. So I want to be clear, it is really not about what you can or cannot do, it is really about what you will or will not do. So we have adversaries like Russia. They continue to amplify propaganda. Everything from the election to coronavirus. We know what they are doing.

Anti-Semitic conspiracy theories. They do it on your platforms weaponizing division and hate to destroy our democracy and our communities. The U.S. intelligence community warned us earlier this year that Russia is now actively inciting white supremacist violence, which the FBI and the Department of Homeland Security say poses the most lethal threat to America. In recent years, we have seen white supremacy and anti-Semitism on the rise, much of it spreading online. And what enables these bad actors to disseminate their hateful messaging to the American public are the algorithms on your platforms, effectively rewarding efforts by foreign powers to exploit divisions in our country. To be sure, I want to acknowledge the work you are already doing in this space. I am relieved to see that Facebook has really taken that long overdue action in banning Holocaust denial content.

But while you have made some policy changes, what we have seen time and time again is what starts online doesn't end online. Hateful words morph into deadly actions, which are then amplified again and again and it is a vicious cycle. Just yesterday, we commemorated the 2-year anniversary of the Tree of Life shooting in Pittsburgh, the deadliest targeted attack on the Jewish community in American history. The shooter in this case had a long history of posting anti-Semitic content on social media sites. And what started online became very real for the families who will now never again see their loved ones. So there has to be accountability when algorithms actively contribute to radicalization and hate. So Mr. Zuckerberg and Mr. Dorsey, when you implement a policy banning hate or disinformation content, how quickly can you adjust your algorithms to reduce this content?

And perhaps what I want to ask even more importantly, to reduce or remove the recommendation algorithm of hate and disinformation. Perhaps it doesn't continue to spread. We know those recommendation algorithms continue to drive someone more specifically and specifically and specifically. Great when you want to buy a new sweater, it is going to be cold out here. It is winter.

Not so great when you are driving them toward hate. Can you talk to us about that please? Mr. Dorsey, you can go first, please.

Mr. DORSEY. As you know, algorithms, machine learning and deep learning are complex. They are complicated and they require testing and training. So as we learn about how, about their effectiveness, we can shift them and we can iterate them. But it does require experience and it does require a little bit of time, so the most important thing that we need to build into the organization is a fast learning mindset and that agility around updating these algorithms. So we do try to focus on the urgency of our updates on any severity of harm, as you mentioned, specifically which tends to lead to off-line dangerous speech that goes into——

Senator ROSEN. Well, Mr. Zuckerberg, I will ask you to answer that then I have some more questions about how, I guess, the nimbleness of your algorithms. Go ahead.

Mr. ZUCKERBERG. Senator, I think you are focused on exactly the right thing in terms of how many people see the harmful content. And as we talk about putting in place regulation or reforming Section 230 in terms of what we want to hold companies accountable for, I think that what we should be judging the companies on is how many people see harmful content before the companies act on it. And I think being able to act on it quickly and being able to act on content that is potentially going viral or going to be seen by more people before it does get seen by a lot of people is critical. This is what we report in our quarterly transparency reports or what percent of the content that a person sees is harmful in any of the categories of harm that we track. And we try to hold ourselves accountable for it, for basically driving the prevalence of that harmful content down. And I think good content regulation here would create a standard like that across the whole industry.

Senator ROSEN. So I like what you said, your recommendation algorithms need to learn to drive the prevalence of this harmful content down. So I have some other questions and I want to ask those. But I would like to see some of the information about how nimble you are on dropping down that prevalence when you do see it trending, when you do see an uptick, whether it is by bots, by human beings, or whatever that is. We need to drive that prevalence down. And so can you talk a little bit, maybe more specifically then on things you might be doing for anti-Semitism? We know that that is white supremacy, the biggest domestic threat—I'm on the Homeland Security Committee. They have testified to this, the largest threat, of course, to our Nation. And I want to be sure that violence is not celebrated and not amplified on your platform.

The CHAIRMAN. We will have to have a brief answer to that. Senator, to whom are you addressing the question?

Senator ROSEN. Mr. Zuckerberg. I think I am the last one but we have just a few seconds we can ask that.

Mr. ZUCKERBERG. Sure, Senator. Thank you. I mean, this is—there is a lot of nuance here, but in general, for each category of harmful content, whether it is terrorist propaganda or incitement and violence and hate speech, we have to build specific systems and specific AI systems. And one of the benefits of, I think, having transparency and transparency reports into how these companies

are doing is that we have to report on a quarterly basis how effectively we are doing at finding those types of content. So you can hold us accountable for how nimble we are. Hate speech is one of the hardest things to train an AI system to get good at identifying because it is linguistically nuanced. We operate in 150 languages around the world. But what our transparency reports show is that over the last few years, we have gone from proactively identifying and taking down about 20 percent of the hate speech on the service.

To now we are proactively identifying I think it is about 94 percent of the hate speech that we end up taking down and the vast majority of that before people even have to report it to us. So but by having this kind of transparency requirement, which is part of what I am advocating for in the Section 230 reform, I think we will be able to have a broader sense across the industry of how all of the companies are improving in each of these areas.

The CHAIRMAN. Thank you for that answer.

Senator ROSEN. I look forward to working with everyone on this. Thank you, Mr. Chairman.

The CHAIRMAN. As do I, Senator Rosen. Thank you very much. When this hearing convened, I estimated that it would last 3 hours and 42 minutes. It has now been 3 hours and 41 minutes. Four of our members have been unable to join us and that is the only reason my prediction was the least bit accurate. So thank you all, thank you very much. And I thank our witnesses. During my first series, during my first question to the panelists, I have referred to a document that I had entered into the record during our committee meeting, I believe, on October 1, entitled "Social Media Companies Censoring Prominent Conservative Voices." That document has been updated. And without objection, it will be added to the record at this point.

[The information referred to follows:]

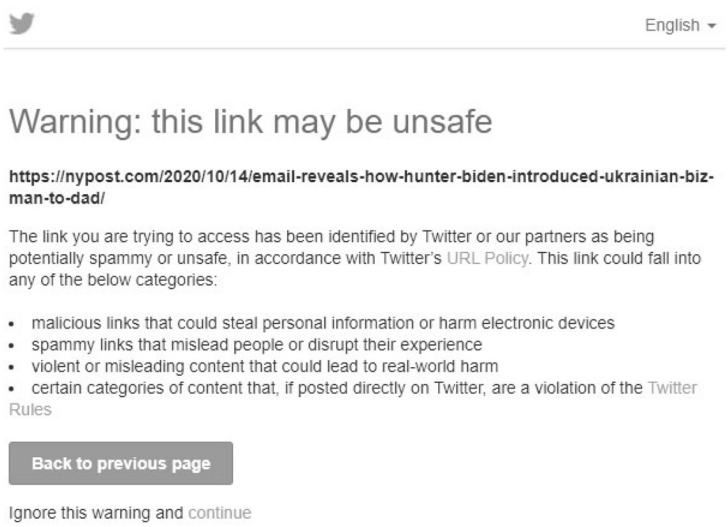
SOCIAL MEDIA COMPANIES CENSORING PROMINENT CONSERVATIVE VOICES

1) Restricted Reach, Deleted Post: Twitter, October 14, 2020—Twitter blocked the distribution of a New York Post article that suggests Hunter Biden introduced Vadym Pozharskyi, an adviser to the board of Burisma, to his father Joe Biden while Joe Biden was Vice President, even though the story was not yet fact-checked.

a) Twitter began by providing some users with a notice that reads "Headlines don't tell the full story. You can read the article on Twitter before Retweeting" when they wanted to retweet it.

b) Twitter then began blocking users from posting any tweets that included a link to the New York Post article. The tweet from the New York Post with the link to the article was deleted from the platform and the4 New York Post's account was suspended. Twitter also blocked users from sending the link to the article via Twitter direct messages.

c) If a user could find a tweet with the link to the New York Post article and clicked on it, they were not taken to the New York Post, but were taken to the following warning page instead:



d) Twitter released the following statement through a representative: "In line with our Hacked Materials Policy, as well as our approach to blocking URLs, we are taking action to block any links to or images of the material in question on Twitter."[1]

e) Twitter's distribution of hacked material policy states: "[W]e don't permit the use of our services to directly distribute content obtained through hacking that contains private information, may put people in physical harm or danger, or contains trade secrets."[2]

f) On October 12, Twitter allowed a fake quote attributed to Susan Collins regarding Judge Amy Coney Barrett to be retweeted more than 6,000 times and received over 17,000 likes. On October 13, the Associated Press conducted a fact check and found the quote was fake.[3] The tweet with the false quote has still not been removed. The false quote reads "At this time I'm not certain that Judge Amy Coney Barrett is the right person to replace Justice Ginsburg. I hope that my colleagues in the Judiciary Committee will be able to alleviate my doubts."[4] The account in question has a history of posting fake quotes from Susan Collins.[5]

g) Twitter suspended several prominent accounts related to President Trump for sharing the New York Post article including White House Press Secretary Kayleigh McEnany's personal account[6] and the official Twitter account of the Trump campaign, @TeamTrump.[7]

h) Reporters were also locked out of their accounts for sharing the link to the article. Politico's Jake Sherman had his account suspended for sharing the link,[8] as well as NewsBusters managing editor Curtis Houck.[9]

i) Twitter Safety released the following statement at 7:44 PM on October 14: "The images contained in the articles include personal and private information—like

---

[1] Karissa Bell, Facebook and Twitter Try to Limit 'NY Post' Story on Joe Biden's Son, Engadget, Oct. 14, 2020, available at *https://www.engadget.com/facebook-twitter-limit-ny-post-story-joe-biden-son-192852336.html*

[2] *https://help.twitter.com/en/rules-and-policies/hacked-materials*

[3] Arijeta Lajka, Fabricated Quote About Supreme Court Nominee Attributed to Maine Senator, Associated Press, Oct. 13, 2020, available at *https://apneas/article/fact-checking-afs:Content:9526587424*

[4] *https://twitter.com/PAULUSV3/status/1315612923000102912*

[5] *https://twitter.com/AndrewSolender/status/1315786510709522437*

[6] *https://twitter.com/TrumpWarRoom/status/1316510056591040513*

[7] *https://twitter.com/mikehahn_/status/1316716049946021888*

[8] *https://twitter.com/JakeSherman/status/1316781581785337857*

[9] Joseph Wulfsohn, Politico's Jake Sherman Says Twitter Suspended Him For Sharing New York Post Report on Hunter Biden, Fox News, Oct. 16, 2020, available at *https://www.foxnews.com/media/politico-jake-sherman-twitter-suspended-hunter-biden*

e-mail addresses and phone numbers—which violate our rules. As noted this morning, we also currently view materials included in the articles as violations of our Hacked Materials Policy." [10]

j) Twitter CEO Jack Dorsey released the following statement at 7:55 PM on October 14: "Our communication around our actions on the @nypost article was not great. And blocking URL sharing via tweet or DM with zero context as to why we're blocking: unacceptable." [11]

k) Twitter updated its Hacked Materials Policy the day after they blocked the New York Post article. Twitter General Counsel Vijaya Gadde announced the following changes: "1. We will no longer remove hacked content unless it is directly shared by hackers or those acting in concert with them 2. We will label Tweets to provide context instead of blocking links from being shared on Twitter" [12] According to Gadde, the changes were made "to address the concerns that there could be many unintended consequences to journalists, whistleblowers and others in ways that are contrary to Twitter's purpose of serving the public conversation." [13]

l) Twitter CEO Jack Dorsey released the following statement regarding the policy changes: "Straight blocking of URLs was wrong, and we updated our policy and enforcement to fix. Our goal is to attempt to add context, and now we have capabilities to do that." [14]

2) Restricted Reach: Facebook, October 14, 2020—Facebook restricted the reach of the New York Post article regarding Hunter Biden even though the article was not yet fact-checked.

a) On October 14, 2020, at 5:00 AM, the New York Post published a story with a newly released e-mail that suggests Hunter Biden introduced Vadym Pozharskyi, an adviser to the board of Burisma, to his father Joe Biden while Joe Biden was Vice President. This allegation contradicts Vice President Biden's public position that he has "never spoken to my son about his overseas business dealings."

b) On October 14, 2020, at 11:10 AM, Andy Stone, Policy Communications Director, Facebook, released the following statement: "While I will intentionally not link to the New York Post, I want be clear that this story is eligible to be fact checked by Facebook's third-party fact checking partners. In the meantime, we are reducing its distribution on our platform."

c) Mr. Stone clarified, at 1:00 PM, that "[t]his is part of our standard process to reduce the spread of misinformation. We temporarily reduce distribution pending fact-checker review." [15] Facebook's policy on temporarily reducing the reach of potential misinformation reads: "[W]e're also working to take faster action to prevent misinformation from going viral, especially given that quality reporting and fact-checking takes time. In many countries, including in the US, if we have signals that a piece of content is false, we temporarily reduce its distribution pending review by a third-party fact-checker." [16] It is not clear what "signals" to Facebook that a piece of content is false.

3) Threats of Harm: Twitter, October 3, 2020—After President Trump was diagnosed with COVID–19, Twitter reiterated that tweets that wish death or bodily harm on any person will be removed and may result in account suspension.[17] After this announcement, Twitter faced criticism that this policy is not forced consistently and Twitter admitted it must improve its enforcement.

a) Twitter's statement in response to criticism: "We hear the voices who feel that we're enforcing some policies inconsistently. We agree we must do better, and we are working together inside to do so." [18]

4) Label: Twitter, September 21, 2020—Tucker Carlson, of Fox's *Tucker Carlson Tonight,* tweeted a news clip of one of his nightly newscasts. Carlson spoke about the destruction and violence in the U.S. and the role that George Soros has played

---

[10] *https://twitter.com/TwitterSafety/status/1316525303930458115*
[11] *https://twitter.com/jack/status/1316528193621327876*
[12] *https://twitter.com/vijaya/status/1316923557268652033*
[13] *https://twitter.com/vijaya/status/1316923550733881344*
[14] *https://twitter.com/jack/status/1317081843443912706*
[15] *https://twitter.com/andymstone/status/1316423671314026496*
[16] *https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/*
[17] *https://twitter.com/TwitterComms/status/1312167835783708672*
[18] *https://twitter.com/TwitterSafety/status/1312498514002243584*

in funding the campaigns of politicians who fail to prosecute those causing the damage.

a) Twitter placed a sensitive label on the video, which required users to click through a filter in order to watch the video.

5) Deleted Post: Facebook, September 18, 2020—Facebook removed an ad run by America First PAC that criticizes Joe Biden's tax policy for raising taxes on all income groups because it lacks context, even though a PolitiFact check finds it is true that Joe Biden's tax policy will result in a tax increase for all income groups.

a) The ad begins with a narrator saying "Sometimes, politicians accidentally tell the truth" and then cuts to a clip of Biden saying "Guess what, if you elect me, your taxes are going to be raised, not cut." The narrator then says "the New York Times says Biden's tax increases are more than double Hillary Clinton's plan. Even the Tax Policy Center admits taxes would increase 'on all income groups."[19]

b) PolitiFact finds it is true that Biden's tax plan would raise taxes for all income groups but rates the ad mostly false because the Biden comment in the ad "came during a South Carolina event, when a member of his crowd said he or she had benefited from the GOP-led tax bill. The full context of the quote shows Biden saying that individual's taxes would be 'raised, not cut, if you benefited from that.'"[20] Facebook removed the ad based on that fact check.

6) Label: Twitter, September 17, 2020—Twitter labeled President Trump's parody tweet of Joe Biden as "manipulated media."

a) At a recent rally, Joe Biden stopped speaking, took out his phone, and played Despacito into the microphone. President Trump shared a video where instead of playing Despacito, Biden plays "F*ck Tha Police."

7) Label, Restricted Reach: Facebook, September 16, 2020—Facebook labeled a Fox News article about Chinese virologist Dr. Li-Meng Yan's claim that China manufactured and intentionally released COVID–19 as containing "false information."

a) The Fox News article summarizes Dr. Yan's interview with Tucker Carlson where she makes her claim and explains that she worked at a World Health Organization (WHO) reference lab at the University of Hong Kong. The article notes that this claim is in conflict with statements made by Dr. Fauci. The article also notes that Dr. Yan helped expose China's attempts to suppress information about how it initially responded to COVID–19.[21]

8) Label, Restricted Reach: Facebook, September 15, 2020—Facebook labeled an ad run by conservative PAC American Principles Project as "missing context and could mislead people" even though PolitiFact says it cannot fact check the ad. Facebook will allow for the ad to be posted organically, but it will not allow for it to be posted as a paid advertisement.

a) The ad, titled "Not Fair" criticizes Joe Biden and Gary Peters for supporting the Equality Act and states "all female athletes want is a fair shot at competition, at a scholarship, at a title, at victory. But what if that shot was taken away by a competitor who claims to be a girl but was born a boy? Senator Gary Peters and Joe Biden support legislation that would destroy girls' sports."[22]

b) The Politifact fact check finds that the Equality Act "would allow transgender students to compete on school sports teams based on their gender identity rather than their sex assigned at birth." It also finds that the "specific criticism is that allowing transgender girls and women to compete on the basis of their gender identity would create an uneven playing field for student athlete and

---

[19] Bill McCarthy, Ad Attacking Joe Biden's Tax Plan Takes His Comments Out of Context, PolitiFact, Aug. 20, 2020, available at *https://www.politifact.com/factchecks/2020/aug/20/america-first/ad-attacking-joe-bidens-tax-plan-takes-his-comment/*

[20] Id.

[21] Sam Dorman, Chinese Virologist: China's Government 'Intentionally' Released COVID–19, Fox News, Sept. 15, 2020, available at *https://www.foxnews.com/media/chinese-virologist-government-intentionally-coronavirus?fbclid=IwAR2q_Jq06e8eN_0oOAjZy8waEu8t_7bckiRg-IUFG9r9abSwIE0ai8KTms4*

[22] Clara Hendrickson, Ad Watch: Conservative PAC Claims Gary Peters Would 'Destroy Girls' Sports', PolitiFact, Sept. 15, 2020, available at *https://www.politifact.com/article/2020/sep/15/ad-watch-peters-supports-ending-discrimination-bas/*

ultimately end girls' and women's sports. That's a prediction we can't fact-check."[23]

c) American Principles Project's Director stated: "Our ad campaign makes a very simple claim: policies supported by Joe Biden, Sen. Gary Peters and other Democrats would destroy girls' sports. There is ample evidence for this claim and more coming in every day. Nothing in the PolitiFact review shows this claim to be false. Yet Facebook has nevertheless decided to declare our ad might 'mislead people' because it is missing context.' Apparently, they believe the ad will only be fair if we also included leftist 'arguments' against us. Do we now need pre-approval from Democrats before we run ads critical of their policies? This is an absurd standard—one which Facebook obviously doesn't hold the other side to."[24]

9) Deleted Post: YouTube, September 14, 2020—YouTube removed a video published by the Hoover Institution of a conversation between Dr. Scott Atlas, a Senior Fellow at the Hoover Institution, and Peter Robinson, a Senior Fellow at the Hoover Institution, because the video "contradicts the WHO or local health authorities medical information about COVID–19"

a) YouTube informed the Wall Street Journal that the video was removed for "falsely stating that a certain age group cannot transmit the virus."[25] In the video, Atlas states children "do not even transmit the disease" but then corrects himself and states that transmission by children is "not impossible, but it's less likely." This is consistent with a review conducted by the American Academy of Pediatrics that states "children are not significant drivers of the COVID–19 pandemic."[26]

b) The video was published in June, before Dr. Atlas left the Hoover Institution to work in the White House. It was removed in September, two days after the New York Times published a story on an open letter from a group of Stanford faculty contesting some of the statements made by Dr. Atlas.[27]

c) Hoover Institution Fellow Lanhee Chen[28] pointed out that, while YouTube has removed the video of Dr. Atlas making a contested claim, it has not removed a video published by the WHO in February that states if you are asymptomatic "you do not have to wear a mask because there is no evidence that they protect people who are not sick."[29]

10) Label: Twitter and Facebook, September 12, 2020—Twitter and Facebook both placed labels on President Trump's posts asking North Carolina voters to go to the polls to see if their mailed ballot had been recorded.

a) President Trump posted: "NORTH CAROLINA: To make sure your Ballot COUNTS, sign & send it in EARLY. When Polls open, go to your Polling Place to see if it was COUNTED. IF NOT, VOTE! Your signed Ballot will not count because your vote has been posted. Don't let them illegally take your vote away from you!"[30]

b) Twitter placed a label requiring users to click through it to view the tweet. The label reads "This Tweet violated the Twitter Rules about civic and election integrity. However, Twitter has determined that it may be in the public's interest for the Tweet to remain accessible."[31]

c) The Facebook label states "voting by mail has a long history of trustworthiness in the U.S. and the same is predicted this year."[32]

d) Twitter refused to place a "civic and election integrity" label on a verifiably false tweet with 25,000 retweets of a picture of locked USPS mailboxes with the caption "A disgrace and immediate threat to American democracy. Shame on them. Shame on the GOP. Where are you Mitch McConnell?" The picture

---

[23] Id.
[24] *https://twitter.com/approject/status/1305903785227714563/photo/1*
[25] Editorial Board, YouTube's Political Censorship, Wall Street Journal, Sept. 14, 2020, available at *https://www.wsj.com/articles/youtubes-political-censorship-11600126230*
[26] Id.
[27] The New York Times, Stanford Doctors Issue a Warning About a Trump Adviser—a Colleague—in an Open Letter, The New York Times, Sept. 10, 2020, available at *https://www.nytimes.com/2020/09/10/world/covid-19-coronavirus.html#link-14e948b0*
[28] *https://twitter.com/lanheechen/status/1305905684785971200*
[29] *https://www.youtube.com/watch?v=Ded_AxFfJoQ&feature=emb_logo*
[30] *https://twitter.com/realDonaldTrump/status/1304769412759724033*
[31] Id.
[32] *https://www.facebook.com/DonaldTrump/posts/north-carolina-to-make-sure-your-ballot-counts-sign-send-it-in-early-when-polls-/10165434016180725/*

of the mailboxes in the tweet is from 2016 and there is a news article from four years ago explaining they were locked to prevent mail theft. The mailboxes are not locked anymore.[33] This tweet was later deleted by the author.

e) Twitter refused to place a "civic and election integrity" label on another verifiably false tweet with over 74,000 retweets and 127,000 likes of a picture of a stack of USPS mailboxes in an industrial lot with the caption "Photo taken in Wisconsin. This is happening right before our eyes. They are sabotaging USPS to sabotage vote by mail. This is massive voter suppression and part of their plan to steal the election." Reuters photojournalist Gary He took the picture included in this tweet and explained that the mailboxes were in the lot of Hartford Finishing Inc. and had been there for years. Hartford Finishing Inc. has a contract with the USPS to refurbish and salvage old mailboxes. A USAToday fact check found the tweet was false.[34] Daniel Goldman, a lawyer who most recently served as majority counsel in the impeachment inquiry and staff counsel to the House Managers, retweeted the tweet with the caption "What possible cost-cutting measure does removing existing mailboxes and mail sorters support? None, of course. Trump admitted that he is using his crony postmaster general to try to steal the election by suppressing votes. This is the most anti-democratic thing he has done."[35] There was no "civic and election integrity" label on his tweet which was retweeted over 8,000 times and received over 17,000 likes.

11) Failure to Act: Facebook, September 9, 2020—Facebook allowed a video of a Mississippi man taking his own life to remain on the platform for three hours. Facebook did not take the video down after an initial review finds that it did not violate any Community Standards.

a) A Mississippi man on a Facebook livestream died of a self-inflicted gunshot wound. Viewers immediately reported the video to Facebook, but Facebook did not take down the video after an initial review. The video remained on the platform for three hours which allowed for nefarious actors to copy it. The video is now circulating on other social media platforms. A close friend of the victim who immediately reported the video believes "if Facebook had intervened 10 minutes after I made the initial report, this is not a situation we're discussing. This is not a situation where a video is spread virally from Facebook to websites, intercut into videos on TikTok. It just doesn't exist."[36]

12) Label: Twitter and Facebook, September 3, 2020—Twitter and Facebook placed labels on President Trump's posts about going to the polling place to ensure a ballot cast in the mail is tabulated.

a) The President posted: "go to your Polling Place to see whether or not your Mail In Vote has been Tabulated (Counted). If it has you will not be able to Vote & the Mail In System worked properly. If it has not been Counted, VOTE (which is a citizen's right to do). If your Mail In Ballot arrives after you Vote, which it should not, that Ballot will not be used or counted in that your vote has already been cast & tabulated. YOU ARE NOW ASSURED THAT YOUR PRECIOUS VOTE HAS BEEN COUNTED, it hasn't been "lost, thrown out, or in any way destroyed". GOD BLESS AMERICA!!!"[37]

b) The Facebook and Twitter labels are the same as the ones placed on his September 12th post which is detailed above.

c) Twitter labeled this unverifiable claim by President Trump but refused to label two viral examples of election misinformation accusing Republicans of sabotaging the USPS to prevent voting by mail despite the fact that both tweets are verifiably false, one even rated false by a USAToday fact check (see above for details).

13) Deleted Post: Facebook, August 28, 2020—The day after a New York Times article noted that the most-shared Facebook post containing the term "Black Lives

---

[33] Tobia Hoonhout, Twitter Says Viral Mailbox Misinformation Does Not Violate Company's Policies, National Review, Aug. 19, 2020 available at *https://www.nationalreview.com/news/twitter-says-viral-mailbox-misinformation-does-not-violate-companys-policies/*

[34] Id.

[35] *https://twitter.com/danielsgoldman/status/1294640736688844803*

[36] C.J. LeMaster, Criticism Lobbed at Social Media Companies After Suicide Video of Mississippi Man Goes Viral, WLBT3, Sept. 9, 2020 available at *https://www.wlbt.com/2020/09/09/criticism-lobbed-social-media-companies-after-suicide-video-mississippi-man-goes-viral/*

[37] *https://twitter.com/realDonaldTrump/status/1301528521752354816*

Matter" in the past six months comes from The Hodgetwins,[38] Facebook notified The Hodgetwins that their "page is at risk of being unpublished because of continued Community Standard violations" and removed three of their videos.

a) The Hodgetwins are black conservative comedians. Their BLM video states that the movement is "a damn lie."[39]

b) One of the removed videos is titled "Liberal has mental breakdown," and another is titled "AOC YOU CRAZY," which contains quotes such as "Man how the hell did she get out of high school," and "She's got a Ph.D, a Ph.D in stupidity."[40]

14) Temporary Ban, Deleted Post: Twitter, July 28, 2020—Twitter suspended the account of Donald Trump Jr. for 12 hours for posting a video of doctors telling Americans they do not have to wear masks and arguing that hydroxychloroquine can treat COVID–19.

a) Twitter deleted the tweet and suspended the account because the tweet was in violation of its COVID–19 misinformation policy.

b) Donald. Trump Jr. responded that "when the Chinese Communist government literally spread disinformation about coronavirus, Twitter literally said that is not a violation of their rules."[41]

15) Temporary Ban: Instagram, July 15, 2020—Instagram twice prevented Mary Morgan Ketchel, daughter of Senator Marsha Blackburn, from advertising the book she wrote with Senator Blackburn titled "Camila Can Vote: Celebrating the Centennial of Women's Right to Vote."

a) Instagram classified advertisements for the book as political ads because of the book's potential to "influence the outcome of an election, or existing proposed legislation." Instagram has stricter disclosure and transparency rules for political ads. Ketchel says the book is solely focused on Tennessee's role in passing women's suffrage laws. Ketchel was asked to register her personal account as a political group. Eventually, Instagram conceded it was not a political ad and published it.[42]

16) Temporary Ban: Instagram July 2, 2020—Instagram temporarily suspended the account of Moral Revolution, a nonprofit Christian group that "speaks about healthy Biblical sexuality" and had 129,000 followers at the time, for posting an anti-porn video.

a) Moral Revolution posted a video on its Instagram alleging that PornHub.com is harmful, is possibly connected to sex trafficking, and should be taken down. Instagram removed the post and then removed the account. The account was eventually reinstated after 350 posts with the hashtag #bringbackmoralrevolution and an account dedicated to restoring the deleted one garnered 19,300 followers.[43]

17) Restricted Reach: Instagram, June 25, 2020—Instagram users were unable to share an Instagram Live video of Christian preacher and music artist Sean Feucht because it contains "harmful or false information."

a) Sean Feucht holds worship services in areas where there recently has been civil unrest. Instagram prevented users from sharing one particular service on their stories because it failed to meet "community standards."[44]

---

[38] Kevin Roose, What if Facebook is the Real 'Silent Majority'? The New York Times, Aug. 27, 2020, available at *https://www.nytimes.com/2020/08/27/technology/what-if-facebook-is-the-real-silent-majority.html*

[39] Id.

[40] Corinne Weaver, Facebook Threatens to Unpublish Black Conservative Creators Hodgetwins, Free Speech America, Aug. 28, 2020, available at *https://censortrack.org/case/hodgetwins*

[41] Joshua Nelson, Trump Jr. Rips Twitter for Restricting His Posts but Allowing China's 'Disinformation' About COVID–19, Fox News, July, 30, 2020, available at *https://www.foxnews.com/media/donald-trump-jr-twitter-china-propaganda*

[42] Tristan Justice, Instagram Blocked Sen. Blackburn And Daughter's Children's Book on Launch Day, The Federalist, July 15, 2020, available at *https://thefederalist.com/2020/07/15/instagram-blocked-sen-blackburn-and-daughters-childrens-book-on-launch-day/*

[43] Heather Moon, Instagram Disables, Reinstates Christian Group That Posted Ant-Porn Video, mrcNewsBusters, July 2, 2020, available at *https://www.newsbusters.org/blogs/techwatch/heather-moon/2020/07/02/instagram-disables-reinstates-christian-group-posted-anti*

[44] Michael Brown, Instagram Brands Christian Worship 'Harmful,' Christian Post, June 26, 2020, available at *https://www.christianpost.com/voices/instagram-brands-christian-worship-harmful.html*

b) Senator Hawley tweeted about this.[45]

18) Label: Twitter, June 23, 2020—Twitter added a notice to President Trump's tweet against establishing an "autonomous zone" in Washington, D.C. because the tweet violated its rules against abusive behavior. You have to click through the notice to view the tweet.

a) President Trump tweeted "There will never be an 'Autonomous Zone' in Washington, D.C., as long as I'm your President. If they try they will be met with serious force!"[46]

b) Twitter responded: "We've placed a public interest notice on this Tweet for violating our policy against abusive behavior, specifically, the presence of a threat of harm against an identifiable group."[47]

c) White House Press Secretary Kyleigh McEnany responded: "Twitter labeled it "abusive behavior" for the President of the United States to say that he will enforce the law. Twitter says it is "abusive" to prevent rioters from forcibly seizing territory to set up a lawless zone in our capital. Recall what happened in Seattle's lawless CHOP zone where multiple people have been shot and one 19-year-old tragically lost his life! We must have LAW AND ORDER!"[48]

19) Deleted Post: YouTube, June 19, 2020—YouTube removed a video published by the Heritage Foundation of a panel discussion at its "Summit on Protecting Children from Sexualization" because it violated YouTube's hate speech policies.

a) In the video, Walter Heyer, a formerly transgender person, expresses regret for his transition and argues that children should not be encouraged to try hormones or surgery. YouTube took issue with Heyer's statement that people are "not born transgender. This is a childhood developmental disorder, that adults are perpetrating on our young people today, and our schools are complicit in this."[49]

b) YouTube removed the video because its "speech policy prohibits videos which assert that someone's sexuality or gender identity is a disease or a mental illness."[50]

20) Deleted Post: Facebook, June 18, 2020—Facebook removes one of President Trump's ads criticizing Antifa for violating Facebook's policy against organized hate because the ad includes an inverted red triangle which Nazi's once used to designate political prisoners. The Communications Director for the Trump campaign said "the red triangle is a common Antifa symbol used in an ad about Antifa" which is why it was included in Trump's ad.[51]

a) The ad reads "Dangerous MOBS of far-left groups are running through our streets and causing absolute mayhem. They are DESTROYING our cities and rioting—it's absolute madness. It's important that EVERY American comes together at a time like this to send a united message that we will not stand for their radical actions any longer. We're calling on YOU to make a public statement and add your name to stand with President Trump against ANTIFA." The ad then has a picture of an inverted red triangle, although the President has other ads with the same text and a different alert image.[52]

b) The inverted red triangle is not listed as a hate symbol by the Anti-Defamation League.

c) The ad was originally flagged on Twitter by a journalist from Fortune.[53]

21) Threat of Demonetization: Google, June 17, 2020—Google threatened to ban The Federalist from its advertising platform because of comments made under

---

[45] *https://twitter.com/HawleyMO/status/1275565792705339392*
[46] *https://twitter.com/realDonaldTrump/status/1275409656488382465*
[47] *https://twitter.com/TwitterSafety/status/1275500569940176897*
[48] *https://twitter.com/PressSec/status/1275546706336116736*
[49] Emily Jashinsky, Exclusive: Man Tried to Share His Regrets About Transgender Life. YouTube Censored It, The Federalist, June 19, 2020 available at *https://thefederalist.com/2020/06/19/exclusive-man-tried-to-share-his-regrets-about-transgender-life-youtube-censored-it/*
[50] Id.
[51] Kayla Gogarty and John Whitehouse, Facebook Finally Removed Trump Campaign Ads with Inverted Red Triangle—an Infamous Nazi Symbol, Media Matters, June 18, 2020, available at *https://www.mediamatters.org/facebook/facebook-let-trump-campaign-run-ads-inverted-red-triangle-infamous-nazi-symbol*
[52] Id.
[53] *https://twitter.com/JohnBuysse/status/1273291676912701441*

a Federalist article titled "The Media Are Lying To You About Everything, Including the Riots."

a) Google's statement: "The Federalist was never demonetized. We worked with them to address issues on their site related to the comments section. Our policies do not allow ads to run against dangerous or derogatory content, which includes comments on sites, and we offer guidance and best practices to publishers on how to comply."[54]

b) The Federalist temporarily deleted its comments section to avoid demonetization.

c) The Federalist notes "Google would be incredibly hard-pressed to moderate or police the millions of incendiary comments posted on YouTube (a website owned by Google). Nor, it should be noted, has Google clamped down on the toxic comment sections of left-wing websites like Daily Kos, Jezebel, or The Young Turks."[55]

22) Demonetization: Google, June 16, 2020—Google banned ZeroHedge, a far-right website, from it advertising platform because of comments made under stories about Black Lives Matter protests.

a) Google stated "We have strict publisher policies that govern the content ads can run on and explicitly prohibit derogatory content that promotes hatred, intolerance, violence or discrimination based on race from monetizing."[56]

23) Label: Twitter, May 29, 2020—Twitter placed a public interest label on President Trump's tweet about the riots in Minnesota. Users have to click through the label to view the tweet.

a) President Trump tweeted: "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. Thank you!"[57]

b) The label reads: "This Tweet violated the Twitter Rules about glorifying violence. However, Twitter has determined that it may be in the public's interest for the Tweet to remain accessible."[58]

c) Ajit Pai, Chairman of the Federal Communications Commission, highlighted four tweets from the Iranian leader, Ayatollah Ali Khamenei, which Twitter did not place a public interest label on for glorifying violence.[59] The tweets at issue include the phrases:

(1) "The Zionist regime is a deadly, cancerous growth and a detriment to this region. It will undoubtedly be uprooted and destroyed."[60]

(2) "The only remedy until the removal of the Zionist regime is firm, armed resistance."[61]

(3) "The struggle to free Palestine is #Jihad in the way of God. Victory in such a struggle has been guaranteed, because the person, even if killed, will receive 'one of the two excellent things.'"[62]

(4) "We will support and assist any nation or any group anywhere who opposes and fights the Zionist regime, and we do not hesitate to say this."[63]

24) Label: Twitter, May 26, 2020—Twitter placed a label on President Trump's tweets related to mail-in ballots for the first time.

a) The President tweeted: "There is NO WAY (ZERO!) that Mail-In Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, bal-

---

[54] *https://twitter.com/Google_Comms/status/1272997427356680195*

[55] Joshua Lawson, Corporate Media Wants to Silence The Federalist Because It Can't Compete, The Federalist, June 18, 2020, available at *https://thefederalist.com/2020/06/18/corporate-media-wants-to-silence-the-federalist-because-it-cant-compete/*

[56] Adele-Momoko Fraser, Google Bans ZeroHedge From Its Ad Platform Over Comments on Protest Article, NBC News, June 16, 2020, available at *https://www.nbcnews.com/tech/tech-news/google-bans-two-websites-its-ad-platform-over-protest-articles-n1231176*

[57] *https://twitter.com/realDonaldTrump/status/1266231100780744704*

[58] Id.

[59] *https://twitter.com/AjitPaiFCC/status/1266368492258816002*

[60] *https://twitter.com/khamenei_ir/status/1263749566744100864*

[61] *https://twitter.com/khamenei_ir/status/1263551872872386562*

[62] *https://twitter.com/khamenei_ir/status/1263742339891298304*

[63] *https://twitter.com/khamenei_ir/status/1263181288338587649*

lots 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. That will be followed up with professionals telling all of these people, many of whom have never even thought of voting before, how, and for whom, to vote. This will be a Rigged Election. No way!" [64]

b) Twitter placed a label on the tweet that reads "get the facts about mail-in ballot." Users do not have to click through a filter to view the tweet. If you click on the label you are taken to a page with Twitter-curated information on election security.[65]

25) Label, Restricted Reach: Facebook, May 19, 2020—Facebook reduced the reach of the PragerU page for "repeated sharing of false news." Facebook flagged a video shared by PragerU titled "Remember This Starving Polar Bear?" as containing false information.

a) The PragerU video contends that the polar bear population is higher than it has been in over 50 years and implores viewers not to "fall for the lies of the climate elites." It is based off of a paper published by Susan Crockford, an Adjunct Professor at the University of Victoria in Canada. After Facebook flagged the post as containing false information, Dr. Crockford defended her claim and noted that while there are conflicting views "this is a classic conflict that happens all the time in science but presents no proof that I'm wrong or that the PragerU video is inherently 'false'." [66]

b) Facebook released this statement: "Third-party fact-checking partners operate independently from Facebook and are certified through the non-partisan International Fact-Checking Network. Publishers appeal to individual fact-checkers directly to dispute ratings." [67]

26) Label, Restricted Reach: Facebook, May 11, 2020—Facebook labeled a LifeNews story as false and reduced its reach after a USAToday fact-check determines it is partly false because the story states that Planned Parenthood is an "abortion business."

a) On February 21, 2019, LifeNews published a story titled "Wisconsin Governor Tony Evers Wants to Force Residents to Fund Planned Parenthood Abortion Business," which states "Governor Tony Evers announced today that he will force state residents to fund the Planned Parenthood abortion business as part of his state budget." [68]

b) More than a year later, it gained steam on Facebook after it was posted in a "Recall Tony Evers" Facebook group. USAToday decided to fact check it after it goes viral and finds the story to be "partly false" because "it's true Gov. Tony Evers tried and failed to restore funding for entities, like Planned Parenthood, that do provide abortion services. But it is false to say residents would be forced to pay for abortions. Even if the measure had passed, under state and Federal law, the money generally couldn't have gone to pay for abortions. Finally, it's an exaggeration to call Planned Parenthood an abortion business, when abortions make up a small portion of the services offered." [69] In response

---

[64] *https://twitter.com/realDonaldTrump/status/1265255835124539392*

[65] *https://twitter.com/i/events/1265330601034256384?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1265255845358645254%7Ctwgr%5Eshare_3&ref_url=https%3A%2F%2Fwww.theguardian.com%2Fus-news%2F2020%2Fmay%2F26%2Ftrump-twitter-fact-check-warning-label*

[66] Susan Crockford, ClimateFeedback Review of PragerU Video Challenges Good News on Polar Bears, Polar Bear Science, May 18, 2020, available at *https://polarbearscience.com/2020/05/18/climate feedback-review-of-prageru-video-challenges-good-news-on-polar-bears/*

[67] Lucas Nolan, Facebook Evicts PragerU From Normal Public Visibility, Claiming 'Repeated Sharing of False News, American Priority, May 20, 2020 available at *https://americanpriority.com/news/facebook-evicts-prageru-from-normal-public-visibility-claiming-repeated-sharing-of-false-news/*

[68] Steven Ertelt, Wisconsin Governor Tony Evers Wants to Force Residents to Fund Planned Parenthood Abortion Business, LifeNews, Feb. 21, 2019, available at *https://www.lifenews.com/2019/02/21/wisconsin-governor-tony-evers-will-force-residents-to-fund-planned-parenthood-abortion-business/*

[69] Haley BeMiller, Fact Check: Planned Parenthood Abortion Funding, Business Claim Goes Too Far, USAToday, May 1, 2020 available at *https://www.usatoday.com/story/news/factcheck/2020/05/01/fact-check-wis-planned-parenthood-abortion-claim-goes-too-far/3057827001/?fbclid=IwAR3dThZ_ZGK5b8kMOr7Dz9c_SMKUitj99e8mTRJi87M_GKdJn2O1uvNiEbg*

to the USAToday article, Facebook labeled the article as false and restricted its reach.

c) Former head of Planned Parenthood Leana Wen said that "First, our core mission is providing, protecting and expanding access to abortion and reproductive health care." [70]

27) Deleted Post: Facebook, March 5, 2020—Facebook removed one of President Trump's ads for violating its policy on misleading census information after initially allowing the ad to stay on the site following criticism from Speaker Nancy Pelosi.

a) The ad reads "We need your help to defeat the Democrats and the FAKE NEWS, who are working around the clock to defeat President Trump in November. President Trump needs you to take the official 2020 Congressional District Census today. We need to hear from you before the most important election in American history. The information we gather from this survey will help us craft strategies in YOUR CONGRESSIONAL DISTRICT. Please respond NOW to help our campaign messaging strategy." The ad then links to a form on President Trump's website which collects information and requests a donation. Facebook allowed the ad to stay up after initially reviewing it. [71]

b) Speaker Nancy Pelosi criticized Facebook for allowing the ad to stay up, stating "I am particularly annoyed today at the actions of Facebook. Facebook has something that is an official document of Donald Trump as saying, 'Fill this out, this is a census form'—it is not. It is an absolute lie, a lie that is consistent with the misrepresentation policy of Facebook. But now they are messing with who we are as Americans." [72] Her characterization of the ad is inconsistent with the text of the ad quoted above.

c) Hours after Speaker Pelosi criticized Facebook, Facebook reversed its decision and removed the ad for "misrepresentation of the dates, locations, times and methods for census participation." Facebook Spokesman Andy stone explained that Facebook reversed its decision because "we conducted a further review." [73] Nothing about the ad changed from the time Facebook initially approved it to when it decided to remove it.

28) Label, Restricted Reach: Facebook, August 30, 2019—Facebook marked two Live Action videos as false for containing the statement "abortion is never medically necessary" and restricted the distribution of the page's posts for repeatedly sharing false information. Live Action is a pro-life group with more than 3 million followers on Facebook.

a) The first video, titled "The Pro-Life Reply to 'Abortion Can be Medically Necessary'" features neonatologist Dr. Kendra Kolb. The second video titled "Abortion is NEVER Medically Necessary" features Live Action's President Lila Rose. Rose says her claim is supported by thousands of OBGYNs and medical experts.[74]

b) Senators Hawley, Cruz, Cramer, and Braun sent a letter to Mark Zuckerberg asking for Facebook to remove the label and any restrictions on distribution.[75] The letter notes that the fact check on this video was conducted by Daniel Grossman, who sits on the board of NARAL Pro-Choice and has testified in support of Planned Parenthood, and Robyn Schickler, who is a Fellow at a pro-abortion advocacy group Physicians for Reproductive Health.

c) Mr. Zuckerberg met with Senator Hawley to discuss the issue. Apparently, Mr. Zuckerberg said there "clearly was bias" in the decision to censor Live Action and told Senator Hawley that bias is "an issue we've struggled with for a long time." [76]

---

[70] Corinne Weaver, Facebook Censors Pro-Life Page for Calling Planned Parenthood an "Abortion Business," LifeNews, May 11, 2020, available at *https://www.lifenews.com/2020/05/11/facebook-censors-lifenews-after-media-falsely-claims-planned-parenthood-not-an-abortion-biz/*

[71] Craig Timberg and Tara Bahrampour, Facebook Takes Down Deceptive Trump Campaign Ads—After First Allowing Them, The Washington Post, March 5, 2020, available at *https://www.washingtonpost.com/ technology/2020/03/05/facebook-removes-trump-ads/*

[72] Id.

[73] Id.

[74] *https://twitter.com/LilaGraceRose/status/1167544725517156352*

[75] Letter to Mark Zuckerberg, from Senators Hawley, Cruz, Cramer, and Braun (Sept. 11, 2019) *https://www.hawley.senate.gov/sites/default/files/2019-09/2019-09-11_Hawley-Cruz-Cramer-Braun-Letter-Facebook-Live-Action.pdf*

[76] *https://twitter.com/HawleyMO/status/1174778110262284288*

d) Facebook removed the labels and restrictions after they had been in place for a month.

29) Temporary Ban: Twitter, August 7, 2019—Twitter temporarily suspended Senator McConnell's "Team Mitch" campaign Twitter account for tweeting a video of protestors gathered outside of Senator McConnell's house shouting violent threats.

a) McConnell was resting at his home and recovering from a broken shoulder. In the video, a protestor outside of the house shouts McConnell "should have broken his little raggedy, wrinkled-(expletive) neck." Another protestor in the video says "Just stab the m——f—— in the heart," in reference to a McConnell voodoo doll. [77]

b) Twitter stated that "the users were temporarily locked out of their accounts for a Tweet that violated our violent threats policy, specifically threats involving physical safety." In a statement to Fox News a Twitter representative said that no account can post calls for violence, even accounts belonging to the targets of those threats.[78]

c) McConnell's campaign manager Kevin Golden released this statement: "This morning, Twitter locked our account for posting the video of real-world, violent threats made against Mitch McConnell. This is the problem with the speech police in America today. The Lexington-Herald can attack Mitch with cartoon tombstones of his opponents. But we can't mock it. Twitter will allow the words 'Massacre Mitch' to trend nationally on their platform. But locks our account for posting actual threats against us. We appealed and Twitter stood by their decision, saying our account will remain locked until we delete the video." [79]

30) Label, Restricted Reach: YouTube, August 6, 2019—Dennis Prager, founder of PragerU wrote an op-ed in the Wall Street Journal describing how the reach of PragerU videos is limited on YouTube.[80]

a) At the time of publication, 56 of the 320 videos PragerU had published on YouTube were age-restricted which limits their reach.

b) Restricted video include: "Israel's Legal Founding" (by Harvard Law professor Alan Dershowitz); "Why America Invaded Iraq" (by Churchill biographer Andrew Roberts); "Why Don't Feminists Fight for Muslim Women?" (by the Somali-American women's-rights activist Ayaan Hirsi Ali); "Are the Police Racist?" (by the Manhattan Institute's Heather Mac Donald); and "Why Is Modern Art So Bad?" (by artist Robert Florczak).

c) The op-ed also cites research by Richard Hanania, a research fellow at the Saltzman Institute of War and Peace Studies at Columbia University which finds: "My results make it difficult to take claims of political neutrality seriously. Of 22 prominent, politically active individuals who are known to have been suspended since 2005 and who expressed a preference in the 2016 U.S. presidential election, 21 supported Donald Trump." [81]

31) Restricted Reach: Instagram, May 21, 2019—Instagram admitted that it inadvertently and incorrectly blocked Let Them Live, a pro-life account with more than 30,000 followers, from using features that let posts go viral.

a) On April 18th, Let Them Live noticed that several of their posts no longer appeared under hashtags which resulted in a large drop in engagement. (If a post is blocked from hashtags then it will not appear when a user clicks on a hashtag and scrolls through all posts with that tag) On April 25th, Let Them Live appealed to Instagram and some of the posts began appearing on hashtags again. On May 11th, all of Let Them Live's posts were blocked from hashtags and were removed from the Explore page. (The Explore page is a page of viral posts from accounts a user does not follow that is curated by algorithm).

---

[77] Gregg Re, Twitter Locks Out McConnell's Campaign for Posting Video of Calls for Violence at His Home, Fox News, August 7, 2019, available at *https://www.foxnews.com/politics/ twitter-locks-out-mcconnell-campaign-for-posting-video-of-calls-for-violence-at-mcconnells-home*

[78] Id.

[79] Id.

[80] Dennis Prager, Don't Let Google Get Away With Censorship, The Wall Street Journal, Aug. 6, 2019, available at *https://www.wsj.com/articles/dont-let-google-get-away-with-censorship-11565132175*

[81] Id.

b) Instagram told the Daily Caller that "we investigated this issue and found that this account was incorrectly filtered from Explore and hashtags. We have removed this restriction and taken steps to ensure this account is not impacted in the future." [82]

32) Temporary Ban, Temporary Deleted Post: Instagram, February 6, 2019—Instagram temporarily disabled Kayleigh McEnany's account and removed her post of Senator Elizabeth Warren's bar registration card because it violated community guidelines on bullying and harassment.

a) At the time, McEnany was not the White House Press Secretary but she was the Republican Party national spokeswoman.

b) Senator Warren said "I never used my family tree to get a break or get ahead" and "I never used it to advance my career." In response, McEnany posted Senator Warren's State Bar of Texas registration card, with Warren's 1986 home address blocked, where Warren listed her race as "American Indian." [83]

c) Senator Warren's registration card was originally obtained and released by the Washington Post and McEnany posted the image of the registration card in the Washington Post article. Instagram apologized and stated "we incorrectly removed this image for including personal information, in this case the home address of someone else, which is not allowed on Instagram. On secondary review, we confirmed that the image included an office address and not a personal home address. The content has now been restored and we apologize for the mistake." [84]

33) Deleted Post, Restricted Reach: Facebook, August 16, 2018—Facebook deleted two PragerU videos and severely limited the reach of seven other videos. Facebook deleted the videos titled "Make Men Masculine Again" and "Where Are the Moderate Muslims?" The seven other videos did not receive more than 10 views despite being shared on the PragerU page that has over 3 million followers.

a) Facebook stated that is mistakenly removed the videos and subsequently restored them which would reverse the reduction in content distribution the PragerU page was experiencing. Facebook also apologized and said it was investigating the matter.[85]

34) Shadow Ban: Twitter, July 26, 2018—Twitter corrected a "glitch" which prevented high-profile conservatives from appearing in the search menu. The phenomenon is referred to as a "shadow ban." RNC Chair Ronna McDaniel, then Representative Mark Meadows, Representative Jim Jordan, Representative Matt Gaetz, and Andrew Surabian, Donald Trump Jr.'s spokesman and former Special Assistant to the President were all affected.

a) Vice News reported on this reduced visibility on July 25, 2018. In the report, Twitter says the reduced visibility is a result of its policies that combat "troll-like behaviors" which is designed such that "people contributing to the healthy conversation will be more visible in conversations and search." The product lead at Twitter said this policy was a result of the behavior of the account and not the content of the account.[86]

b) Twitter CEO Jack Dorsey stated that Twitter does not shadow ban, but that "it suffices to say we have a lot more work to do to earn people's trust on how we work." [87]

35) Restricted Reach: Twitter, October 9, 2017—Twitter blocked Senator Marsha Blackburn from paying to promote a campaign ad titled "Why I'm Running" because it contained an "inflammatory statement" about abortion. In the ad, Senator

---

[82] Chris White, Instagram Explains Why It Effectively Shadow Banned A Pro-Life Group's Content, Daily Caller News Foundation, May 21, 2019, available at *https://dailycaller.com/2019/05/21/instagram-prolife-shadow-ban/*

[83] *https://twitter.com/kayleighmcenany/status/1093191708257538048?lang=en*

[84] Amber Athey, Instagram Apologizes for Removing GOP Spokeswoman's Post About Sen. Warren, The Federalist, Feb. 6, 2020, available at *https://dailycaller.com/2019/02/06/instagram-kayleigh-mcenany-elizabeth-warren/*

[85] *https://hotair.com/archives/john-s-2/2018/08/18/facebook-hey-sorry-making-prageru-videos-disappear/*

[86] Alex Thompson, Twitter Appears to Have Fixed 'Shadow Ban' of Prominent Republicans like the RNC Chair and Trump Jr.'s Spokesman, Vice News, July 25, 2018, available at *https://www.vice.com/en_us/article/43paqq/twitter-is-shadow-banning-prominent-republicans-like-the-rnc-chair-and-trump-jrs-spokesman*

[87] *https://twitter.com/jack/status/1022196722905296896*

Blackburn states that while she was in the House she "fought Planned Parenthood, and we stopped the sale of baby body parts. Thank God." [88]

a) Senator Blackburn served as the Chair of the House Select Investigative Panel on Infant Lives. In that position, Senator Blackburn led the investigation into the illegal fetal-tissue-trafficking industry which was prompted by a series of undercover videos that were released at the time concerning Planned Parenthood.

b) Twitter told the Associated Press that it restricted the paid promotion of the ad because it contained "an inflammatory statement that is likely to evoke a strong negative reaction." Twitter told two employees of Targeted Victory, the digital consulting firm working for Senator Blackburn's campaign that if the line "stopped the sale of baby body parts" was removed then the ad would not be blocked.[89]

c) Twitter reversed the decision after a day and provided this reasoning: "After further review, we have made the decision to allow the content in question from Rep. Blackburn's campaign ad to be promoted on our ads platform. While we initially determined that a small portion of the video used potentially inflammatory language, after reconsidering the ad in the context of the entire message, we believe that there is room to refine our policies around these issues." [90]

36) Deleted Post, Restricted Reach: Facebook, May 10, 2016—Facebook removes its "Trending Topics" after reports that the curators of the feature routinely excluded conservative stories and injected stories that were not actually trending but the curators believed deserved attention.

a) "Trending Topics" was a feature that listed the most talked about news stories of the day in the top right corner of a user's Facebook interface. "news curators" hired by Facebook helped determine which stories were listed. A former "news curator" admitted in an interview with Gizmodo that the workers would routinely remove or prevent stories about "CPAC gathering, Mitt Romney, Rand Paul, and other conservative topics" from being included in the feature. Other suppressed topics include "former IRS official Lois Lerner, who was accused by Republicans of inappropriately scrutinizing conservative groups; Wisconsin Gov. Scott Walker; popular conservative news aggregator the Drudge Report; Chris Kyle, the former Navy SEAL who was murdered in 2013; and former Fox News contributor Steven Crowder." [91]

b) Senator Thune, then Chairman of the Commerce Committee, sent a letter to Mark Zuckerberg asking for more information on the Trending Topics feature.[92]

c) Facebook eventually discontinued the feature.

The CHAIRMAN. And I believe we are now at the point of closing. The hearing record will remain open for two weeks. During this time, Senators are asked to submit any questions for the record. Upon receipt, the witnesses are requested to submit their written answers to the Committee as soon as possible, but by no later than Wednesday, November 25, 2020. I want to thank our witnesses for their cooperation and for bearing with us during a very lengthy hearing. And I want to thank each member of the Committee for their cooperation in the conduct of this hearing. With that, the

---

[88] *https://www.youtube.com/watch?v=wxSPO4V7FYI*

[89] Kevin Robillard, Twitter Pulls Blackburn Senate Ad Deemed 'Inflammatory,' Politico, Oct. 9, 2017, available at *https://www.politico.com/story/2017/10/09/marsha-blackburn-twitter-ad-243607*

[90] Kurt Wagner, Twitter Changed its Mind and Will Let Marsha Blackburn Promote Her 'Inflammatory" Campaign Ad After All, Vox, Oct. 10, 2017, available at *https://www.vox.com/2017/10/10/16455902/twitter-marsha-blackburn-video-ad-reversal-allowed*

[91] Michael Nunez, Former Facebook Workers: We Routinely Suppressed Conservative News, Gizmodo, May 9, 2016, available at *https://gizmodo.com/former-facebook-workers-we-routinely-suppressed-conser-1775461006*

[92] Letter to Mark Zuckerberg, from Senator John Thune, Chairman, Senate Committee on Commerce, Science, and Transportation, (May 10, 2016) *https://www.commerce.senate.gov/services/files/fe5b7b75-8d53-44c3-8a20-6b2c12b0970d*

hearing is concluded and the witnesses are thanked. This hearing is now adjourned.

[Whereupon, at 1:42 p.m., the hearing was adjourned.]

A P P E N D I X

The Leadership Conference
on Civil and Human Rights

1620 L Street, NW
Suite 1100
Washington, DC
20036

202.466.3311 voice
202.466.3435 fax
www.civilrights.org

Officers
Chair
Judith L. Lichtman
National Partnership for
Women & Families
Vice Chairs
Thomas A. Saenz
Mexican American Legal
Defense and Educational Fund
Hilary Shelton
NAACP
Secretary/Treasurer
Lee A. Saunders
American Federation of State,
County & Municipal Employees

Board of Directors
Kevin Allis
National Congress of American Indians
Kimberly Churches
AAUW
Kristen Clarke
Lawyers' Committee for
Civil Rights Under Law
Alphonso B. David
Human Rights Campaign
Rory Gamble
International Union, UAW
Fatima Goss Graves
National Women's Law Center
Jonathan Greenblatt
Anti-Defamation League
Mary Kay Henry
Service Employees International Union
Sherrilyn Ifill
NAACP Legal Defense and
Educational Fund, Inc.
David H. Inoue
Japanese American Citizens League
Benjamin Jealous
People for the American Way
Derrick Johnson
NAACP
Virginia Kase
League of Women Voters of the
United States
Samer E. Khalaf
American-Arab
Anti-Discrimination Committee
Marc Morial
National Urban League
Janet Murguía
UnidosUS
Debra L. Ness
National Partnership for
Women & Families
Christian F. Nunes
National Organization for Women
Rabbi Jonah Pesner
Religious Action Center
Of Reform Judaism
Rebecca Pringle
National Education Association
Lisa Rice
National Fair Housing Alliance
Anthony Romero
American Civil Liberties Union
Maria Town
American Association of
People with Disabilities
Richard L. Trumka
AFL-CIO
Randi Weingarten
American Federation of Teachers
John C. Yang
Asian Americans Advancing Justice |
AAJC

President & CEO
Vanita Gupta

The Leadership Conference on Civil and Human Rights

October 27, 2020

Senator Roger Wicker
Chairman
Committee on Commerce, Science, and Transportation
U.S. Senate
Washington, DC 20510

Senator Maria Cantwell
Ranking Member
Committee on Commerce, Science, and Transportation
U. S. Senate
Washington, DC 20510

Dear Chairman Wicker and Ranking Member Cantwell,

On behalf of The Leadership Conference on Civil and Human Rights ("The Leadership Conference"), a coalition charged by its diverse membership of more than 220 national organizations to promote and protect the rights of all persons in the United States, we thank you for the opportunity to submit our views regarding the need for major tech companies to address threats to civil rights created or facilitated by their platforms and improve civil rights infrastructure, and ask that this statement be entered into the record of the committee hearing entitled "Does Section 230's Sweeping Immunity Enable Big Tech's Bad Behavior?" on October 28, 2020.

The internet has created immense positive value by connecting people, facilitating civil rights advocacy and adding new voices to our culture and public debate. However, it can also enable discriminatory conduct, exacerbate existing disparities, and give new tools to those who want to threaten, harass, intimidate, defame, or violently attack people different from themselves. While The Leadership Conference welcomes scrutiny of the role of social media companies in our democracy, we urge caution regarding potential changes to Section 230 to ensure any proposed changes will not do more harm than good. We encourage the committee to focus on the most important opportunities to ensure these platforms serve all people, which we discuss in more detail below.

Technological progress should promote equity and justice as it enhances safety, economic opportunity and convenience for everyone. On October 21, The Leadership Conference joined dozens of leading civil rights and technology advocacy organizations in releasing updated Civil Rights Principles for the Era of Big Data, in response to the current risks to civil rights – including COVID-19, a surge in hate-based violence, private sector and government surveillance, and disinformation on social media platforms designed to manipulate or suppress voter participation – and with an eye toward how technology can meet its promise and affirmatively promote justice and equity. These principles provide

important guidelines to aid this committee in ensuring that new technologies—including algorithmic decision making, artificial intelligence and machine learning— protect civil rights, prevent unlawful discrimination, and advance equal opportunity.

Congress should use this opportunity to ask platforms what actions they are taking or plan to take to reduce online activities that harm communities of color, religious minorities, and other marginalized communities. For years, we have urged major tech platforms to take responsibility for ensuring that their products and business processes protect civil and human rights and do not result in harm or bias against historically marginalized groups, but they have failed to take sufficient action. And despite years of advocacy urging the companies to rectify the problems, misinformation regarding time, place, manner, and qualifications to vote and content intended to suppress or deter people from voting continue to proliferate. The failure of tech platforms to address these activities harms people of color and members of other marginalized communities. Moreover, despite new policies that ostensibly forbid white supremacy, white supremacists continue to use platforms to incite racist violence on multiple platforms against Asian Americans, African Americans, Jews, Muslims, people with disabilities, and members of the LGBTQ community. Platforms have the tools and the ability to respond effectively to these concerns if they only had the will. Congress should press tech companies on the actions they are taking to improve and enforce their own policies and stop the weaponization of their platforms to suppress the vote, spread hate, and undermine our democracy.

Congress should not be distracted by baseless claims of "anti-conservative" bias and should instead focus on platforms' efforts to respond to online voter suppression and other threats to our democracy. A commitment to civil and human rights is not a 'right' or 'left' issue – it is about right versus wrong. Baseless allegations of so-called anti-conservative bias should not distract tech companies. Research shows that anti-conservative bias is a phantom problem; a number of studies, articles, and reports[i] show that the voices of marginalized communities are more likely to be regarded as "toxic" by content moderators and content moderation artificial intelligence.

Congress should instead focus on some of the more significant challenges facing social media platforms such as safeguarding our elections and the census from manipulation and disinformation, as well as fighting hate and harassment online. We have made a series of recommendations to obviate false, misleading, and harmful content on the companies' platforms that could lead to voter suppression and the spread of hate speech. While Google, Facebook, Twitter, and other social media platforms have made some recent policy changes, their lack of consistent enforcement makes these policies insufficient to prevent the spread of voter suppression. Platforms must also better utilize disinformation tools for voter suppression content as they have done for other issues like COVID-19; and must prevent disinformation in political ads.[ii]

Section 230 must be considered carefully and in context. President Trump's actions to use Section 230 to pressure social media companies is a threat to our civil liberties. The President's Executive Order and the FCC's recent announcement defy both statutory and constitutional principles in the name of protecting the president's own speech online regardless of the consequences to everyone else. And many of the current legislative proposals around Section 230 would do more harm than good. One such example, the EARN IT Act, threatens to not only exacerbate the censorship that many LGBTQ persons face online, but

also threatens the welfare and safety of the sex worker community. Instead of looking at simply changing Section 230 as a means of platform regulation, Congress should clearly define the problem and carefully consider whether Section 230 has a role in causing or exacerbating the problem before turning to making changes to Section 230 as part of the solution.

Congress should press tech companies to conduct independent civil rights audits as well as improve their civil rights infrastructure. Structural changes within the platforms will also help better protect civil rights by ensuring platforms can hold themselves accountable to their commitment to civil rights, diversity, and inclusion. Among the companies appearing at the committee hearing, only Facebook has undertaken a civil rights audit with outside auditors, though civil rights groups have urged all the major platforms to do so. Congress must press the other tech companies to conduct credible independent civil right audits. But Facebook's example demonstrates that without institutional commitment and outside pressure, the impact of an audit will be limited and short-lived.

That is why, in addition to pushing for civil rights audits, Congress must also urge tech companies to adopt structural reforms that comply with federal civil rights law and demonstrate that the companies understand that civil rights are not a partisan issue, but instead are fundamental to protecting the constitutional rights of all people and thus should be part of the organic structure and operations of these companies. This means that tech companies must hire staff with civil rights expertise in senior leadership. The civil rights infrastructure within the companies must be well-resourced and empowered within the company and consulted on the companies' major decisions. New and clarified policies should be subject to vetting and review by internal teams with real civil rights expertise and experience, prior to their implementation. Finally, tech companies should provide a process and format through which civil rights advocates and the public can engage with the companies and monitor their progress.

Congress must also press tech companies to do more to address meaningful diversity and inclusion at their workplaces and the lack of people of color in senior executive, engineering, and technical positions. People of color who are working at the companies often face discrimination and unequal pay, as well as a culture where they are devalued. Tech companies must ensure that this does not happen in their workplaces and must address the inequities that may have already occurred. They also must expand strategies to attract and retain talent in diverse communities to expand access to jobs and opportunities.

Prevention of harm, not damage and after-the-fact repair, must be the goal. This goal cannot be fully accomplished if those with civil rights expertise are not part of decision-making processes. Congress must continue to review and scrutinize tech companies to make sure that they are taking the necessary steps to accomplish this goal.

Congress should consider other meaningful ways to protect civil and human rights. Congress should also focus on other means to protect civil and human rights. For example, invasive data collection and use practices can lead to civil rights violations. Congress should pass comprehensive federal consumer privacy legislation that protects consumers by requiring companies to minimize the data they collect, define permissible and impermissible purposes for collecting, sharing, and using personal data, prohibit discriminatory uses of personal data, and provide for algorithmic transparency and fairness in automated decisions. Congress should ensure federal agencies are focusing on identifying and ending data



processing and algorithmic practices that discriminate on the basis of protected characteristics with respect to access to credit, housing, education, public accommodations and elsewhere.

Thank you for the consideration of our views. If you have any questions about the issues raised in this letter, please contact Leadership Conference Media/Telecommunications Task Force Co-Chairs Cheryl Leanza, United Church of Christ, Office of Communication, Inc., at Cleanza@alhmail.com and Kate Ruane, American Civil Liberties Union, kruane@aclu.org; or Corrine Yu, Leadership Conference Senior Program Director, at yu@civilrights.org.

Sincerely,

Vanita Gupta
President and CEO

LaShawn Warren
Executive Vice President for Government Affairs

---

[i] *See* Casey Newton, *Leaving content moderation to volunteers is empowering racists*, The Verge (June 9, 2020) https://www.theverge.com/interface/2020/6/9/21283442/content-moderation-racism-facebook-reddit-nextdoor-karen; Bertram Lee, *Moderating Race on Platforms*, Public Knowledge (January 29, 2020) https://www.publicknowledge.org/blog/moderating-race-on-platforms/; Thomas Davidson et al., *Racial Bias in Hate Speech and Abusive Language Detection Datasets*, Cornell University arXiv:1905.12516v1 (May 29, 2019) https://arxiv.org/abs/1905.12516v1; Maarten Sap et al., *The Risk of Racial Bias in Hate Speech Detection*, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics 1668–1678 (2019) https://homes.cs.washington.edu/~msap/pdfs/sap2019risk.pdf; Shirin Ghaffary, *The algorithms that detect hate speech online are biased against black people*, Vox: Recode (August 15, 2019) https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter.

[ii] *Leadership Conference Urges Social Media Platforms to Address Online Voter Suppression*, The Leadership Conference on Civil and Human Rights (October 13, 2020) https://civilrights.org/resource/leadership-conference-urges-social-media-platforms-to-address-online-voter-suppression/

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. ROGER WICKER TO JACK DORSEY

Question 1. What does "good faith" in Section 230 mean? Is there any action you could take that could not be justified as done in "good faith"? Do you agree bad faith content moderation is not covered by Section 230? If content is removed pre-textually, or if terms and conditions are applied inconsistently depending on the viewpoint expressed in the content, is that removing content in good faith?

Answer. Section 230 of the Communications Decency Act (CDA) affords two critical types of legal protection to "interactive computer service" providers like Twitter. Under subsection (c)(1), the Communications Decency Act provides that neither providers nor the people who use our service are to "be treated as the publisher or speaker of any information provided by another information content provider." Subsection (c)(2) provides that neither interactive computer service providers nor the people who use our service are to be held liable for either "action[s] voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, . . . excessively violent, harassing, or otherwise objectionable" or "action[s] taken to enable or make available . . . the technical means to restrict access" to any such objectionable material.[1] Irrespective of Section 230s protections, companies have a First Amendment right to assert editorial control over content published on their platform.[2]

We believe it is critical that individuals have trust in the platforms they use and transparency into how content moderation decisions are made. Thus, we support efforts to ensure transparency of content moderation policies and processes. In addition, we support efforts to expand procedural fairness, so customers can feel confident that decisions are being made fairly and equitably.

Question 2. Why wouldn't a platform be able to rely on terms of service to address categories of potentially harmful content outside of the explicit categories in Section 230(c)(2)? Why should platforms get the additional protections of Section 230 for removal of yet undefined categories of speech? Does Section 230s "otherwise objectionable" catchall offer immunity for content moderation decisions motivated by political bias? If the "otherwise objectionable" catchall does not offer such immunity, what limiting principle supports the conclusion that the catchall does not cover politically-biased moderation? If the "otherwise objectionable" catchall does offer such immunity now, how would you rewrite Section 230 to deny immunity for politically-biased content moderation while retaining it for moderation of content that is harmful to children?

Answer. Section 230(c)(2)(A) shields an interactive computer service provider like Twitter from liability for good faith attempts to restrict access to "material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable." The purpose of subsection (c)(2) is to encourage Twitter and our industry peers to moderate content without fear of being sued for their moderation decisions. It has been instrumental in providing platforms the flexibility to make content moderation decisions that safeguard the safety of the public conversation.

As explained in more detail in our written testimony, we do not believe that the solution to concerns raised about content moderation is to eliminate Section 230 liability protections. Instead, we believe the solution should be focused on enhancing transparency, procedural fairness, privacy, and algorithmic choice.

Question 3. Are your terms of service easy to understand and transparent about what is and is not permitted on your platform? What notice and appeals process do you provide users when removing or labeling third-party speech? What redress might a user have for improper content moderation beyond your internal appeals process? In what way do your terms of service ensure against politically-biased content moderation and in what way do your terms of service limit your ability to moderate content on your platform? How would you rewrite your terms of service to pro-

---

[1] The operation of one shield is not conditioned on the availability of the other, and one or both may apply depending on the circumstances. Subsection (c)(1) is broad and unconditional (other than several limited exceptions), and subsection (c)(2) "provides an additional shield from liability." Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1105 (9th Cir. 2009), as amended (Sept. 28, 2009) see also Fyk v. Facebook, Inc., 808 F. App'x 597, 598 (9th Cir. 2020)

[2] See, *e.g.,* La'Tiejira v. Facebook, Inc., 272 F. Supp. 3d 981, 991 (S.D. Tex. 2017) (holding that "Facebook [has a] First Amendment right to decide what to publish and what not to publish on its platform" and collecting cases); Zhang v. Baidu.com, Inc., 10 F. Supp. 3d 433 (S.D.N.Y. 2014) (rejecting claim seeking to hold a website "liable for . . . a conscious decision to design its search-engine algorithms to favor a certain expression on core political subjects . . . [because] to proceed would plainly 'violate the fundamental rule of protection under the First Amendment, that a speaker has the autonomy to choose the content of his own message.'")

tect against politically-biased content moderation? Do you think that removing content inconsistent with your terms of service and public representations is removal of content "in good faith"?

Answer. The Twitter Rules and all incorporated policies, Privacy Policy, and Terms of Service collectively make up the "Twitter User Agreement" that governs an individual's access to and use of Twitter's services. We have the Twitter Rules in place to help ensure everyone feels safe expressing their beliefs and we strive to enforce them with uniform consistency. We are continually working to update, refine, and improve both our enforcement and our policies, informed by in-depth research around trends in online behavior both on and off Twitter and feedback from the people who use Twitter. We believe we have to rely on a straight-forward, principled approach and focus on the long term goal of understanding—not just in terms of the service itself—but in terms of the role we play in society and our wider responsibility to foster and better serve a healthy public conversation.

We have worked to make the Twitter Rules, Terms of Service, and appeals process accessible and transparent. For example, we recently rewrote the Twitter Rules so each Rule can be contained in the length of a Tweet (280 characters) and is straightforward for the people who use our service. In addition, we offer the ability for people who use Twitter to file an appeal if they believe a decision has been made in error. We have also expanded efforts to more clearly communicate additional information about our actions with affected account holders and individuals who reported a Tweet for a potential violation, so they have the information to determine whether they want to take follow-up action like an appeal.

Question 4. Please provide a list of all instances in which a prominent individual promoting liberal or left-wing views has been censored, demonetized, or flagged with extra context by your company. Please provide a list of all instances in which a prominent individual promoting conservative or right-wing views has been censored, demonetized, or flagged with extra context by your company. How many posts by government officials from Iran or China have been censored or flagged by your company? How many posts critical of the Iranian or Communist Chinese government have been flagged or taken down?

Answer. Twitter does not use political viewpoints, perspectives, or party affiliation to make any decisions. In regard to the removal of accounts, our biannual *Twitter Transparency Center* highlights trends in enforcement of our Rules, legal requests, intellectual property-related requests, and e-mail privacy best practices. We provide aggregate numbers of accounts we have actioned across twelve categories of Terms of Service violations in this report, which can be found at transparency.twitter.com. Due to security and privacy concerns, we cannot discuss individual incidents, but we take action on accounts across the world and across the political spectrum.

Question 5. Should algorithms that promote or demote particular viewpoints be protected by Section 230? Why or why not? Do you think the use of an individual company's algorithms to amplify the spread of illicit or harmful materials like online child sexual exploitation should be protected by Section 230?

Answer. We believe that people should have choices about the key algorithms that affect their experience online. We recognize that we can do even more to provide greater algorithmic transparency and fair machine learning. The machine learning teams at Twitter are studying these techniques and developing a roadmap to ensure our present and future algorithmic models uphold a high standard when it comes to transparency and fairness. We believe this is an important step in ensuring fairness in how we operate and we also know that it is critical that we be more transparent about our efforts in this space. Regarding your question on the amplification of harmful materials, Twitter has zero tolerance for any material that features or promotes child sexual exploitation. We strive to proactively detect and remove this abhorrent content and it is not amplified through our algorithms.

Question 6. Should platforms that knowingly facilitate or distribute Federal criminal activity or content be immune from civil liability? Why?/Why not? If your company has actual knowledge of content on your platform that incites violence, and your company fails to remove that content, should Federal law immunize your company from any claims that might otherwise be asserted against your company by victims of such violence? Are there limitations or exceptions to such immunity that you could propose for consideration by the Committee? Should platforms that are willfully blind to Federal criminal activity or content on their platforms be immune from civil liability? Why? Why not?

Answer. The Communications Decency Act currently exempts Federal criminal activity from liability protections. Section 230(e)(1) ("Nothing in this section shall be construed to impair the enforcement of section 223 or 231 of this title, chapter 71

(relating to obscenity) or 110 (relating to sexual exploitation of children) of title 18 or any other Federal criminal statute.").

An individual who uses Twitter may not use our service for any unlawful purposes or in furtherance of illegal activities. By using Twitter, an individual agrees to comply with all applicable laws governing his or her online conduct and content. Additionally, we have Twitter Rules in place that prohibit violence, terrorism/violent extremism, child sexual exploitation, abuse/harassment, hateful conduct, promoting suicide or self-harm, sensitive media (including graphic violence and adult content), and illegal or certain regulated goods or services. More information about each policy can be found in the *Twitter Rules.*

Question 7. Mr. Dorsey, you informed both Senator Cruz and Senator Johnson that you believe Twitter has no ability to influence the election. Do you still stand by this claim? If Twitter has no ability to influence the election, why does it label or remove content on the grounds of election interference? Do voter suppression and misinformation affect elections? Have voter suppression attempts and misinformation appeared on Twitter? How can it then be true that Twitter cannot influence elections?

Answer. We believe that we have a responsibility to protect the integrity of conversations related to elections and other civic events from interference and manipulation. Therefore, we prohibit attempts to use our services to manipulate or disrupt civic processes, including through the distribution of false or misleading information about the procedures or circumstances around participation in a civic process. Combatting attempts to interfere in conversations on Twitter remains a top priority for the company, and we continue to invest heavily in our detection, disruption, and transparency efforts related to state-backed information operations. Twitter defines state-backed information operations as coordinated platform manipulation efforts that can be attributed with a high degree of confidence to state-affiliated actors. Our goal is to remove bad faith actors and to advance public understanding of these critical topics.

Question 8. As you know, many politicians in Washington have expressed concerns around how political speech has been monitored by online platforms like Twitter leading up to the 2020 election. You likely are also aware that the current COVID–19 pandemic has ushered in an unthinkable amount of fraudulent and unsafe activities related to coronavirus "cures" and "treatments" through social media. In your opinion, should Twitter and other online platforms be held responsible for the promulgation of unsafe and fraudulent claims like those related to the COVID–19 pandemic? What specific action(s) has Twitter taken to block, suspend, or report COVID-related fraudulent activity since the pandemic began in March? Has Twitter proactively reported suspected illegal or fraudulent users to the proper authorities upon discovering the sharing of or attempted sale of illegal products like COVID–19 "cures" or other illegal drugs?

Answer. The public conversation occurring on Twitter is critically important during this unprecedented public health emergency. With a critical mass of expert organizations, official government accounts, health professionals, and epidemiologists on our service, our goal is to elevate and amplify authoritative health information as far as possible. To address this global pandemic, on March 16, 2020, we announced *new enforcement guidance,* broadening our definition of harm to address, specifically, content related to COVID–19 that goes directly against guidance from authoritative sources of global and local public health information. We require individuals to remove violative Tweets in a variety of contexts with the goal of preventing offline harm. Additionally, we are currently engaged in an effort launched by the Office of the U.S. Chief Technology Officer under President Trump in which we are coordinating with our industry peers to provide timely, credible information about COVID–19 via our respective platforms. This working group also seeks to address misinformation by sharing emerging trends and best practices.

Response to Written Questions Submitted by Hon. John Thune to Jack Dorsey

Question 1. We have a public policy challenge to connect millions of Americans in rural America to broadband. I know you share in our commitment to connect every American household with broadband not only because it's the right thing to do but because it will add millions of new users to your platforms, which of course, means increase profits. What role should Congress and your companies play in ensuring that we meet all the broadband demands in rural America?

Answer. Bridging the digital divide is of the utmost importance. Twitter is supportive of efforts, including bipartisan proposals to expand funding for broadband

in response to the pandemic, to ensure that all individuals in America can access the Internet. Additionally, on May 15, 2020, Mr. Dorsey *Tweeted* he would personally donate $10 million dollars to #OaklandUndivided to ensure every child in Oakland, CA has access to a laptop and Internet in their homes.

Question 2. The PACT Act would require your platforms to take down content that a court has ruled to be illegal. Do you support a court order-based takedown rule?

Answer. Any court-order takedown rule must be crafted in a way that safeguards free expression and due process. For example, we strongly encourage any court determinations relied upon to be clear in their finality. We do not believe that temporary restraining orders constitute a final legal determination; it would be inappropriate to require permanent removal of content at a preliminary stage of litigation before full consideration of the merits of the case. Additionally, temporary restraining orders raise due process concerns as the proceeding may be ex parte, particularly where the speaker is anonymous. In such cases, the speaker may be unable to participate in the proceeding to advance their interests. We believe strongly that any court order-based takedown rule must protect individuals' due process, speech, and other constitutional rights.

Moreover, any proposed court-order takedown rule should take into account the risk that bad-faith actors submit counterfeit takedown notices to platforms, a phenomenon that has been well-documented, including in *this article* regarding falsified court orders.

Question 3. Section 230 was initially adopted to provide a "shield" for young tech start-ups against the risk of overwhelming legal liability. Since then, however, some tech platforms like yours have grown larger than anyone could have imagined. Often a defense we hear from Section 230 proponents is that reform would hurt current and future start-ups. The PACT Act requires greater reporting from tech platforms on moderation decisions, largely exempts small business. However, your companies are no longer start-ups, but rather some of the most powerful and profitable companies in the world. Do tech giants need "shields" codified by the U.S. government? Have you outgrown your need for Section 230 protections?

Answer. Section 230 is the Internet's most important law for free speech and safety. Weakening Section 230 protections will remove critical speech from the Internet. We must ensure that all voices can be heard, and we continue to make improvements to our service so that everyone feels safe participating in the public conversation—whether they are speaking or simply listening. Eroding the foundation of Section 230 could collapse how we communicate on the Internet, leaving only a small number of giant and well-funded technology companies. We should also be mindful that undermining Section 230 could result in far more removal of online speech and impose severe limitations on our collective ability to address harmful content and protect people online.

Question 4. As discussed during the hearing, please provide for the record a complete list of U.S. newspaper articles that Facebook suppressed or limited the distribution of over the past five years, as Facebook did with the October 14, 2020 New York Post article entitled "Smoking-Gun E-mail Reveals How Hunter Biden Introduced Ukrainian Businessman to VP Dad." For each article listed, please also provide an explanation why the article was suppressed or the distribution was limited.

Answer. We believe that Facebook is in the best position to respond to the question posed.

Question 5. What does "good faith" in Section 230 mean? Is there any action you could take that could not be justified as done in "good faith"? Do you agree bad faith content moderation is not covered by Section 230? If content is removed pre-textually, or if terms and conditions are applied inconsistently depending on the viewpoint expressed in the content, is that removing content in good faith?

Answer. Section 230 of the Communications Decency Act (CDA) affords two critical types of legal protection to "interactive computer service" providers like Twitter. Under subsection (c)(1), the Communications Decency Act provides that neither providers nor the people who use our service are to "be treated as the publisher or speaker of any information provided by another information content provider." Subsection (c)(2) provides that neither interactive computer service providers nor the people who use our service are to be held liable for either "action[s] voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, . . . excessively violent, harassing, or otherwise objectionable" or "action[s] taken to enable or make available . . . the technical means

to restrict access" to any such objectionable material.[3] Irrespective of Section 230s protections, companies have a First Amendment right to assert editorial control over content published on their platform.[4]

We believe it is critical that individuals have trust in the platforms they use and transparency into how content moderation decisions are made. Thus, we support efforts to ensure transparency of content moderation policies and processes. In addition, we support efforts to expand procedural fairness, so customers can feel confident that decisions are being made fairly and equitably.

Question 6. Mr. Pichai noted in the hearing that without the "otherwise objectionable" language of Section 230, the suppression of teenagers eating tide pods, cyberbullying, and other dangerous trends would have been impossible. Could the language of Section 230 be amended to specifically address these concerns, by including the language of "promoting selfharm" or "unlawful" without needing the "otherwise objectionable" language that provides online platforms a blank check to take down any third-party speech with which they disagree?

Answer. Section 230(c)(2)(A) shields an interactive computer service provider like Twitter from liability for good faith attempts to restrict access to "material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable." The purpose of subsection (c)(2) is to encourage Twitter and our industry peers to moderate content without fear of being sued for their moderation decisions. The providers who undertake the moderating must be empowered to exercise the flexibility to determine what content is "objectionable." See, *e.g., Domen v. Vimeo, Inc.,* 433 F.Supp.3d 592, 604 (S.D.N.Y. 2020) that held subsection (c)(2) immunity applied, and refused to question the interactive computer service provider's view regarding objectionable material.

Question 7. What other language would be necessary to address truly harmful material online without needing to rely on the vague term "otherwise objectionable?"

Answer. As outlined in our testimony, we believe that future solutions should be focused on enhancing transparency, procedural fairness, privacy, and algorithmic choice. We caution against changes to existing liability protections that could encourage companies to over-moderate or under-moderate, out of fear of legal liability. Specifically, limiting liability protections for content moderation could counterproductively make it more difficult for companies to take steps to protect safety and combat malicious actors.

Question 8. Why wouldn't a platform be able to rely on terms of service to address categories of potentially harmful content outside of the explicit categories in Section 230(c)(2)? Why should platforms get the additional protections of Section 230 for removal of yet undefined categories of speech?

Answer. The Twitter Rules and all incorporated policies, Privacy Policy, and Terms of Service collectively make up the "Twitter User Agreement" that governs an individual's access to and use of Twitter's services. We are continually working to update, refine, and improve our enforcement and our policies, informed by in-depth research around trends in online behavior and feedback from the people who use Twitter. However, given the evolving nature of online speech, there is no way that any platform can know in advance all of the future speech that will be potentially harmful and include it in their terms of service. A core purpose of Section 230 is to encourage Twitter and our industry peers to moderate content with flexibility to address evolving content issues without fear of being sued for their moderation decisions.

Question 9. Does Section 230s "otherwise objectionable" catchall offer immunity for content moderation decisions motivated by political bias? If the "otherwise objectionable" catchall does not offer such immunity, what limiting principle supports the conclusion that the catchall does not cover politically-biased moderation? If the "otherwise objectionable" catchall does offer such immunity now, how would you rewrite

---

[3] The operation of one shield is not conditioned on the availability of the other, and one or both may apply depending on the circumstances. Subsection (c)(1) is broad and unconditional (other than several limited exceptions), and subsection (c)(2) "provides an additional shield from liability." Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1105 (9th Cir. 2009), as amended (Sept. 28, 2009) see also Fyk v. Facebook, Inc., 808 F. App'x 597, 598 (9th Cir. 2020)

[4] See, *e.g.,* La'Tiejira v. Facebook, Inc., 272 F. Supp. 3d 981, 991 (S.D. Tex. 2017) (holding that "Facebook [has a] First Amendment right to decide what to publish and what not to publish on its platform" and collecting cases); Zhang v. Baidu.com, Inc., 10 F. Supp. 3d 433 (S.D.N.Y. 2014) (rejecting claim seeking to hold a website "liable for . . . a conscious decision to design its search-engine algorithms to favor a certain expression on core political subjects . . . [because] to proceed would plainly 'violate the fundamental rule of protection under the First Amendment, that a speaker has the autonomy to choose the content of his own message.'")

Section 230 to deny immunity for politically-biased content moderation while retaining it for moderation of content that is harmful to children?

Answer. Twitter does not use political viewpoints, perspectives, or party affiliation to make any decisions. In regard to the removal of accounts, our biannual Twitter Transparency Report highlights trends in enforcement of our Rules, legal requests, intellectual property-related requests, and e-mail privacy best practices. We provide aggregate numbers of accounts we have actioned across twelve categories of terms of service violations in this report, which can be found at transparency.twitter.com.

Question 10. Are your terms of service easy to understand and transparent about what is and is not permitted on your platform?

Answer. The Twitter Rules and all incorporated policies, Privacy Policy, and Terms of Service collectively make up the "Twitter User Agreement" that governs an individual's access to and use of Twitter's services. These are easy to understand and are transparent for the individuals who use Twitter.

Question 11. What notice and appeals process do you provide users when removing or labeling third-party speech?

Answer. We may suspend an account if it has been reported to us as violating our Rules. We may suspend it temporarily or, in some cases, permanently. Individuals may be able to unsuspend their own accounts by providing a phone number or confirming an e-mail address. An account may also be temporarily disabled in response to reports of automated or abusive behavior. For example, an individual may be prevented from Tweeting from his or her account for a specific period of time or may be asked to verify certain information before proceeding. If an account was suspended or locked in error, an individual can appeal. First, the individual must log in to the account that is suspended and file an appeal. The individual must describe the nature of the appeal and provide an explanation of why the account is not in violation of the Twitter Rules. Twitter employees will typically engage with the account holder via e-mail to resolve the appeal.

Question 12. What redress might a user have for improper content moderation beyond your internal appeals process?

Answer. We strive to give people an easy, clear way to appeal decisions we make that they think are not right. Mistakes in enforcement—made either by a human or algorithm—are inevitable, and why we strive to make appeals easier. We believe that all companies should provide a straightforward process to appeal decisions. This makes certain people can let us know when we do not get it right, so that we can fix any mistakes and make our processes better in the future.

Question 13. In what way do your terms of service ensure against politically-biased content moderation and in what way do your terms of service limit your ability to moderate content on your platform?

Answer. We ensure that all decisions are made at Twitter without using political viewpoints, party affiliation, or political ideology, whether related to automatically ranking content on our service or how we develop or enforce the Twitter Rules. Our Twitter Rules are not based on ideology or a particular set of beliefs. We believe strongly in being impartial, and we strive to enforce our Twitter Rules fairly.

Question 14. How would you rewrite your terms of service to protect against politically-biased content moderation?

Answer. Twitter does not use political viewpoints, perspectives, or party affiliation to make any decisions, whether related to automatically ranking content on our service or how we develop or enforce our rules. Our rules are not based on ideology or a particular set of beliefs. Instead, the Twitter Rules are based on behavior.

Question 15. Do you think that removing content inconsistent with your terms of service and public representations is removal of content "in good faith"?

Answer. We do not remove content inconsistent with our terms of service and public representations, however, mistakes in enforcement are inevitable and we strive to make appeals easier.

Question 16. As it stands, Section 230 has been interpreted not to grant immunity if a publishing platform "ratifies" illicit activity. Do you agree? How do you think "ratification" should be defined?

Answer. Under our policies, an individual who uses Twitter may not use our service for any unlawful purposes or in furtherance of illegal activities. By using Twitter, an individual agrees to comply with all applicable laws governing his or her online conduct and content. Twitter does not "ratify" illicit activity and this activity is *expressly prohibited.*

Question 17. Do you agree that a platform should not be covered by Section 230 if it adds its own speech to third-party content?

Answer. Section 230 defines "information content provider" as "any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service." 47 U.S.C. §230(f)(3). A court would be required to make a legal judgment based on specific facts to determine whether a platform had a role in "creat[ing] or develop[ing]" the particular content at issue in a specific case.

Question 18. When a platform adds its own speech, does it become an information content provider under Section 230(f)(3)?

Answer. Section 230 defines "information content provider" as "any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service." 47 U.S.C. §230(f)(3). A court would be required to make a legal judgment based on specific facts to determine whether a platform had a role in "creat[ing] or develop[ing]" the particular content at issue in a specific case.

Question 19. Should algorithms that promote or demote particular viewpoints be protected by Section 230? Why or why not?

Answer. We believe that people should have choices about the key algorithms that affect their experience online. We recognize that we can do even more to provide greater algorithmic transparency and fair machine learning. The machine learning teams at Twitter are studying these techniques and developing a roadmap to ensure our present and future algorithmic models uphold a high standard when it comes to transparency and fairness. We believe this is an important step in ensuring fairness in how we operate and we also know that it is critical that we be more transparent about our efforts in this space.

Question 20. Do you think the use of an individual company's algorithms to amplify the spread of illicit or harmful materials like online child sexual exploitation should be protected by Section 230?

Answer. Twitter has zero tolerance for any material that features or promotes child sexual exploitation. We strive to proactively detect and remove this abhorrent content and it is not amplified through our algorithms. When we detect childhood sexual exploitation material, we report it to the National Center for Missing and Exploited Children (NCMEC) or International Center for Missing and Exploited Children (ICMEC). We also are an active member of the Technology Coalition and Thorn Technical Task Force, which are collaborative efforts to tackle this challenge.

Question 21. Should platforms that knowingly facilitate or distribute Federal criminal activity or content be immune from civil liability? Why or why not?

Answer. The Communications Decency Act does not provide liability protections for individuals or platforms that engage in Federal criminal activity. Additionally, an individual who uses Twitter may not use our service for any unlawful purposes or in furtherance of illegal activities. By using Twitter, an individual agrees to comply with all applicable laws governing his or her online conduct and content. Additionally, we have Twitter Rules in place that prohibit violence, terrorism/violent extremism, child sexual exploitation, abuse/harassment, hateful conduct, promoting suicide or self-harm, sensitive media (including graphic violence and adult content), and illegal or certain regulated goods or services. More information about each policy can be found in the *Twitter Rules.*

Question 22. If your company has actual knowledge of content on your platform that incites violence, and your company fails to remove that content, should Federal law immunize your company from any claims that might otherwise be asserted against your company by victims of such violence? Are there limitations or exceptions to such immunity that you could propose for consideration by the Committee?

Answer. Twitter does not permit people on Twitter to promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. We also do not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories.

Question 23. Should platforms that are willfully blind to Federal criminal activity or content on their platforms be immune from civil liability? Why or why not?

Answer. The Communications Decency Act currently exempts Federal criminal activity from liability protections. Section 230(e)(1) ("Nothing in this section shall be construed to impair the enforcement of section 223 or 231 of this title, chapter 71 (relating to obscenity) or 110 (relating to sexual exploitation of children) of title 18 or any other Federal criminal statute.") We do not permit people on Twitter to use our service for any unlawful purposes or in furtherance of illegal activities. By using Twitter, an individual agrees to comply with all applicable laws governing his or her online conduct and content.

Question 24. You informed both Senator Cruz and Senator Johnson that you believe Twitter has no ability to influence the election. Do you still stand by this claim?

Answer. We believe that we have a responsibility to protect the integrity of conversations related to elections and other civic events from interference and manipulation. Therefore, we prohibit attempts to use our services to manipulate or disrupt civic processes, including through the distribution of false or misleading information about the procedures or circumstances around participation in a civic process. Combatting attempts to interfere in conversations on Twitter remains a top priority for the company, and we continue to invest heavily in our detection, disruption, and transparency efforts related to state-backed information operations. Twitter defines state-backed information operations as coordinated platform manipulation efforts that can be attributed with a high degree of confidence to state-affiliated actors. Our goal is to remove bad faith actors and to advance public understanding of these critical topics.

Question 25. If Twitter has no ability to influence the election, why does it label or remove content on the grounds of election interference?

Answer. We believe we have a responsibility to protect the integrity of conversations related to elections and other civic events from interference and manipulation. Twitter's civic integrity policy addresses four categories of misleading behavior and content. First, we label or remove false or misleading information about how to participate in an election or other civic process. Second, we take action on false or misleading information intended to intimidate or dissuade people from participating in an election or other civic process. Third, we do not allow individuals to create fake accounts which misrepresent their affiliation to a candidate, election, official, political party, electoral authority, or government entity. Fourth, we label or remove false or misleading information intended to undermine public confidence in an election or other civic process.

Question 26. Do voter suppression and misinformation affect elections? Have voter suppression attempts and misinformation appeared on Twitter? How can it then be true that Twitter cannot influence elections?

Answer. We have heard from the people who use Twitter that we should not determine the truthfulness of Tweets and we should provide context to help people make up their own minds in cases where the substance of a Tweet is disputed. When we label Tweets, we link to Twitter conversation that shows three things for context: (1) factual statements; (2) counterpoint opinions and perspectives; and (3) ongoing public conversation around the issue. We will only add descriptive text that is reflective of the existing public conversation to let people determine their own viewpoints. Reducing the visibility of Tweets means that we will not amplify the Tweets on a number of surfaces across Twitter.

Response to Written Questions Submitted by Hon. Roy Blunt to Jack Dorsey

Question 1. Twitter is used for a variety of purposes, and this hearing has been called to examine whether it is appropriate for Twitter to censor some speech or speakers, while it enjoys freedom from liability for that speech though Section 230. Music is an essential driver of user engagement on Twitter. According to the Recording Industry Association of America, seven of the top ten most followed accounts are those of recording artists, who have tens of millions of followers. These artists' accounts generate millions of views and subsequently result in new Twitter accounts that are created by fans. Your company greatly benefits from the personal data that is collected from this user base to drive advertising revenue. Despite this, I have heard that Twitter has been slow to respond to copyright infringement on its platform and also refused to negotiate licenses or business agreements with music publishers or record labels to compensate music creators. How many copyright-related takedown notices has Twitter received so far this year? Of those, how many of those have resulted in takedowns, or removal of copyright protected content? What do you believe constitutes a reasonable attempt to remove copyrighted material from your site?

Answer. Twitter responds to copyright complaints submitted under the Digital Millennium Copyright Act, also known as the DMCA. We are required to respond to reports of alleged copyright infringement, including allegations concerning the unauthorized use of a copyrighted video or music. We review takedown notices and respond expeditiously as the law requires. We are as transparent as possible regarding the removal or restriction of access to user-posted content. We report this information to Lumen, and we clearly mark withheld Tweets and media to indicate to

viewers when content has been withheld. We provide detailed metrics regarding *enforcement of the DMCA* in our bi-annual Twitter Transparency Report.

Question 2. Twitter has taken steps to identify and remove both illegal and illicit content across its platform, however, content creators remain outwardly concerned that reasonable steps aren't being taken to protect their work. How can Twitter build on its existing work to better identify copyrighted content, and ensure that crimes like digital piracy are not permitted to occur on your platform?

Answer. If the owner of the copyright has considered fair use, and the owner still wishes to continue with a copyright complaint, the individual may want to first reach out to the user in question to see if the owner can resolve the matter directly with the user. The owner of the copyright can reply to the user's Tweet or send the user a Direct Message and ask for them to remove the copyrighted content without having to contact Twitter. An individual can report alleged copyright infringement by visiting Twitter's Help Center and filing a copyright complaint. An individual logged in to twitter.com can visit the Twitter Help Center directly from a Twitter account by clicking the 'Help' link located in the sidebar.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. JERRY MORAN TO JACK DORSEY

Question 1. How much money does your company spend annually on content moderation in general?

Answer. Putting a dollar amount on our broader efforts on content moderation is a complex request. We have made acquisitions of companies, our staff work on a series of overlapping issues, and we have invested in technology and tools to support our teams reviewing content.

Question 2. How many employees does your company have that are involved with content moderation in general? In addition, how many outside contractors does your company employ for these purposes?

Answer. We have a global workforce of over 5,000 employees, a substantial portion of whom are directly involved in reviewing and moderating content on Twitter.

Question 3. How much money does your company currently spend on defending lawsuits stemming from users' content on your platform?

Answer. Twitter is currently involved in, and may in the future be involved in, legal proceedings, claims, investigations, and government inquiries and investigations arising in the ordinary course of business. These proceedings, which include both individual and class action litigation and administrative proceedings, have included, but are not limited to matters involving content on the platform, intellectual property, privacy, data protection, consumer protection, securities, employment, and contractual rights. Legal fees and other costs associated with such actions are expensed as incurred.

Question 4. Without Section 230s liability shield, would your legal and content moderation costs be higher or lower?

Answer. There are various Executive and Congressional efforts to restrict the scope of the protections from legal liability for content moderation decisions and third-party content posted on online platforms that are currently available to online platforms under Section 230 of the Communications Decency Act, and our current protections from liability for content moderation decisions and third-party content posted on our platform in the United States could decrease or change, potentially resulting in increased liability for content moderation decisions and third-party content posted on our platform and higher litigation costs.

Question 5. Twitter entered a settlement with the FTC in 2011 to resolve charges that the company deceived consumers by putting their privacy at risk for failing to safeguard their personal information. This settlement was to remain in effect for twenty years. Understanding that an investigation is ongoing, are you able to indicate if Twitter's recent security incident from July 15, 2020, that involved high profile accounts to facilitate a cryptocurrency scam violated that 2011 settlement?

Answer. We would respectfully refer you to the U.S. Federal Trade Commission for additional information on this matter.

Question 6. How many liability lawsuits have been filed against your company based on user content over the past year?

Answer. We are currently involved in, and may in the future be involved in, legal proceedings, claims, investigations, and government inquiries and investigations arising in the ordinary course of business.

Question 7. Please describe the general breakdown of categories of liability, such as defamation, involved in the total number of lawsuits over the past year.

Answer. These proceedings, which include both individual and class action litigation and administrative proceedings, have included, but are not limited to matters involving content on the platform, intellectual property, privacy, data protection, consumer protection, securities, employment, and contractual rights.

Question 8. Of the total number of liability lawsuits based on user content, how many of them did your company rely on Section 230 in its defense?

Answer. Twitter is currently involved in, and may in the future be involved in, legal proceedings, claims, investigations, and government inquiries and investigations arising in the ordinary course of business. Any filings we have made in response to specific lawsuits in the United States are publicly available. While we do not have a comprehensive analysis to share at this time, please refer to the Internet Association's recent *study* on Section 230 cases.

Question 9. Of the liability lawsuits based on user content in which your company relies on Section 230 in its defense, what categories of liability in each of these lawsuits is your company subject to?

Answer. These proceedings, which include both individual and class action litigation and administrative proceedings, have included, but are not limited to matters involving content on the platform, intellectual property, privacy, data protection, consumer protection, securities, employment, and contractual rights.

Question 10. In a defamation case based on a user content, please describe the typical procedural steps your company takes to litigate these claims.

Answer. We engage in standard litigation practices in response to legal claims.

Question 11. Of the claims that have been dismissed on Section 230 grounds, what is the average cost of litigation?

Answer. Litigation can involve defense and settlement costs, diversion of management resources, and other costs.

Question 12. I understand the U.S.-Mexico-Canada Agreement (USMCA) contains similar intermediary liability protections that Section 230 established domestically. The recent trade deal with Japan also included similar provisions. If Congress were to alter Section 230, do you expect litigation or free trade agreement compliance issues related to the United States upholding trade agreements that contain those provisions?

Answer. While we cannot speak for the outcome of any actions Congress may wish to take in the future, Twitter advocates for regional and global regulatory alignment on principle. As a platform for the global public conversation, Twitter will support regulations that benefit free expression globally.

Question 13. How does the inclusion of Section 230-like protections in the aforementioned trade deals affect your business operations in the countries party to said trade deals? Do you expect fewer defamation lawsuits and lower legal costs associated with intermediary liability in those countries due to these trade deals?

Answer. As mentioned above, Twitter advocates for regional and global regulatory alignment on principle. Countries and companies that do not share this core value do so at the expense of global free expression, fair competition and global connectivity.

Question 14. In countries that do not have Section 230-like protections, are your companies more vulnerable to litigation or liability as a result?

Answer. We are subject to legislation in Germany that may impose significant fines for failure to comply with certain content removal and disclosure obligations. Other countries, including Singapore, India, Australia, and the United Kingdom, have implemented or are considering similar legislation imposing penalties for failure to remove certain types of content. We could incur significant costs investigating and defending these claims. If we incur material costs or liability as a result of these occurrences, our business, financial condition and operating results would be adversely impacted.

Question 15. How do your content moderation and litigation costs differ in these countries compared to what you might expect if Section 230-like protections were in place?

Answer. Legal risk may be enhanced in jurisdictions outside the United States where our protection from liability for content published on our platform by third parties may be unclear and where we may be less protected under local laws than we are in the United States.

Question 16. As American companies, does Section 230s existence provide you any liability protection overseas in countries that do not have similar protections for tech companies?

Answer. Section 230 is the Internet's most important law for free speech and safety, and we encourage countries to use Section 230 as a model for protecting critical speech online. Moreover, as we see more and more attempts by governments to undermine open conversation around the world, Section 230 sets a powerful democratic standard. This has helped to protect not only free expression, but other fundamental human rights that are interconnected with speech. Thus, changes to Section 230 could have broader ripple effects across the globe.

Question 17. To differing extents, all of your companies rely on automated content moderation tools to flag and remove content on your platforms. What is the difference in effectiveness between automated and human moderation?

Answer. Twitter uses a combination of machine learning and human review to adjudicate reports of violations and make determinations on whether the activity violates our rules. One of the underlying features of our approach is that we take a behavior-first approach. That is to say, we look at how accounts behave before we look at the content they are posting. This is how we were able to scale our efforts globally.

Question 18. What percentage of decisions made by automated content moderation systems are successfully appealed, and how does that compare to human moderation decisions?

Answer. We may suspend an account if it has been reported to us as violating our Rules. We may suspend it temporarily or, in some cases, permanently. Individuals may be able to unsuspend their own accounts by providing a phone number or confirming an e-mail address. An account may also be temporarily disabled in response to reports of automated or abusive behavior. For example, an individual may be prevented from Tweeting from his or her account for a specific period of time or may be asked to verify certain information before proceeding. If an account was suspended or locked in error, an individual can appeal. First, the individual must log in to the account that is suspended and file an appeal. The individual must describe the nature of the appeal and provide an explanation of why the account is not in violation of the Twitter Rules. Twitter employees will engage with the account holder via e-mail to resolve the suspension.

Question 19. Please describe the limitations and benefits specific to automated content moderation and human content moderation.

Answer. Automated content moderation is necessary to engage in content moderation at scale. An additional benefit is that it reduces the burden on individuals to report conduct that may violate the Twitter Rules. More than 50 percent of Tweets we take action on for abuse are now proactively surfaced using technology, rather than relying on reports to Twitter.

Question 20. In your written testimonies, each of you note the importance of tech companies being transparent with their users. Have you already, or do you plan to make public the processes that your automated moderation system undertakes when making decisions about content on your platform?

Answer. An important component of our transparency efforts is the Twitter Transparency Center. This year, we expanded our biannual transparency report site to become a comprehensive Twitter Transparency Center. Our goal with this evolution is make our transparency reporting more easily understood and accessible to the general public. This site includes data visualizations making it easier to compare trends over time and more information for the individuals who use Twitter, academics, researchers, civil society groups and others who study what we do to understand bigger societal issues. The Center includes data on enforcement actions under the Twitter Rules that requires the removal of specific Tweets or to suspend accounts. The Center also includes sections covering information requests, removal requests, copyright notices, trademark notices, e-mail security, platform manipulation, and state-backed information operations. We believe it is now more important than ever to be transparent about our practices and we continue to explore additional ways to increase our transparency.

Question 21. Given the complexity of the algorithms that are now governing a portion of the content across your platforms, how have you or how do you plan to explain the functions of your automated moderation systems in a simple manner that users can easily understand?

Answer. In December 2018, Twitter introduced an icon located at the top of everyone's timelines that allows individuals using Twitter to easily switch to a reverse chronological order ranking of the Tweets from accounts or topics they follow. This improvement gives people more control over the content they see, and it also provides greater transparency into how our algorithms affect what they see. It is a good start. We believe this points to an exciting, market-driven approach where people

can choose what algorithms filter their content so they can have the experience they want.

Question 22. Acknowledging Twitter updated its "Hacked Materials Policy" on October 15, 2020, the company cited this policy the day before when it blocked the distribution of the New York Post article pertaining to Joe Biden and his son Hunter Biden's business dealings. Are you able to describe the substance of the changes in the updates to this policy?

Answer. Following our enforcement actions, we received significant feedback—both positive and negative—on how we enforced the Distribution of Hacked Materials Policy. After reviewing the feedback, we made changes within 24 hours to the policy to address concerns that there could be unintended consequences to journalists, whistleblowers and others in ways that are contrary to Twitter's purpose of serving the public conversation. We also noted publicly that the only enforcement action available under the Distribution of Hacked Materials Policy was removal, which was no longer in alignment with new product capabilities, such as a label, that provide people with additional context.

On October 23, we issued a revised policy on the Distribution of Hacked Materials that states that we will no longer remove hacked content unless it is directly shared by hackers or groups directly associated with a hack. We also laid out our intent to use labels instead of removal of Tweets in other circumstances for violations of our policy to provide more context.

Question 23. While I understand that Twitter's policy of outright blocking URLs is no longer in place, I found it concerning that reports indicated that the top line of the Warning Page that users were directed to in trying to access the article stated that "This link may be unsafe." While the Warning Page goes on to describe that the link was blocked for what could be several categories of reasons, including "violations of the Twitter Rules," I believe it is deceptive for a platform to state that a site a user is trying to visit is "unsafe" when in actuality the platform determines the content of the site to be untrue, or worse, against its own political agenda. What issues could we reasonably expect to arise should companies begin to tie data security and privacy concerns to materials or sites that they claim are untrue or determine to be disadvantageous to the politics of their platform?

Answer. Based on our enforcement decision under our Hacked Materials policy, people on Twitter were blocked from sharing certain links from the @NYPost, publicly or privately, as those specific articles contained the source materials themselves. References to the contents of the materials or discussion about the materials were not restricted under the policy. In order to address the unique facts in the @NYPost case, we determined that we should change our practices to allow for circumstances when actions on a specific account have led to a policy change. Accordingly, we updated the relevant policy, informed @NYPost, and the newspaper's account was restored. While we may have taken longer than some would have wanted to take these actions, we believe that this process and associated review have helped us create strong and more transparent policies.

Question 24. How has COVID–19 impacted your company's content moderation systems? Is there a greater reliance on automated content moderation? Please quantify how content moderation responsibilities have shifted between human and automated systems due to COVID–19.

Answer. In March 2020, we released information to the public relating to our *contingency planning* regarding the increase of our use of machine learning and automation to take a wide range of actions on potentially abusive and manipulative content.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. MIKE LEE TO JACK DORSEY

Question 1. Mr. Dorsey, during the hearing I asked each of the witnesses to provide me with one example of a high-profile person or entity from a liberal ideology that your company has censored and what particular action you took. In response, you stated, "We can give you a more exhaustive list." Later you told me that your company has taken action against "two democratic congresspeople," although when pressed, you were unable to provide their names. Could you provide me with the "exhaustive list" and the names of the "two democratic congresspeople," along with the actions taken against them?

Answer. We would like to clarify that at the time of the hearing we had taken enforcement action on two accounts of Democratic candidates for the U.S. House of Representatives for violations of our civic integrity policy. The account holders were

candidates and do not currently serve in Congress. Mr. Dorsey inadvertently stated the candidates were members of Congress.

Question 2. Mr. Dorsey, you have previously stated, "Let me be clear about one important and foundational fact: Twitter does not use political ideology to make any decisions, whether related to ranking content on our service or how we enforce our rules. . ." The term "misinformation" can be subjective and Twitter has often used it when labeling content that contains political viewpoints especially during this election year. How are you consistent with your promise to not "use political ideology" when you subjectively take down political views that you deem "misinformation"?

Answer. We have heard from the people who use Twitter that we should not determine the truthfulness of Tweets and we should provide context to help people make up their own minds in cases where the substance of a Tweet is disputed. When we label Tweets, we link to Twitter conversation that shows three things for context: (1) factual statements; (2) counterpoint opinions and perspectives; and (3) ongoing public conversation around the issue. We will only add descriptive text that is reflective of the existing public conversation to let people determine their own viewpoints. Reducing the visibility of Tweets means that we will not amplify the Tweets on a number of surfaces across Twitter.

Question 3. Mr. Dorsey, while Facebook "reduced distribution" of the *NY Post,* Twitter wouldn't even let anyone post the article. Twitter cited its "hacked materials" policy for why it chose to block the posting of the *NY Post* article. Did you apply the "hacked materials" policy to the release of President Trump's tax records? Or the Steele Dossier? Did you block the leak of the Edward Snowden documents?

Answer. We issued the Distribution of Hacked Materials Policy in advance of the U.S. 2018 midterm elections to discourage and mitigate harms associated with hacks and unauthorized exposure of private information. Pursuant to these policies, on October 14, 2020, we took action on Tweets related to two articles published by the New York Post that, based on preliminary information, linked to content we determined to be in violation of our policies. Following our enforcement actions, we received significant feedback—both positive and negative—on how we enforced the Distribution of Hacked Materials Policy. After reviewing the feedback, we made changes within 24 hours to the policy to address concerns that there could be unintended consequences to journalists, whistleblowers and others in ways that are contrary to Twitter's purpose of serving the public conversation. We also noted publicly that the only enforcement action available under the Distribution of Hacked Materials Policy was removal, which was no longer in alignment with new product capabilities, such as a label, that provide people with additional context.

On October 23, we issued a revised policy on the Distribution of Hacked Materials that states that we will no longer remove hacked content unless it is directly shared by hackers or groups directly associated with a hack. We also laid out our intent to use labels instead of removal of Tweets in other circumstances for violations of our policy to provide more context.

Question 4. Congress is in the midst of a debate over future reforms to Section 230. This is an important discussion that Congress should have. a. In making decisions to moderate third-party content on your platform, do you rely solely on Section 230? In other words, could you still moderate third-party content without the protections of Section 230? If the provisions of Section 230 were repealed or severely limited, how would your content moderation practices shift?

Answer. Under the First Amendment, Twitter is permitted to exercise affirmative editorial control over content created by the people who use our service.[5] Section 230 is the Internet's most important law for free speech and safety. Weakening Section 230 protections will remove critical speech from the Internet. We must ensure that all voices can be heard, and we continue to make improvements to our service so that everyone feels safe participating in the public conversation—whether they are speaking or simply listening. The protections offered by Section 230 help us achieve this important objective. Eroding the foundation of Section 230 could collapse how we communicate on the Internet, leaving only a small number of giant and well-funded technology companies. We should also be mindful that undermining Section 230 will result in far more removal of online speech and impose severe limi-

[5] See, *e.g.,* La'Tiejira v. Facebook, Inc., 272 F. Supp. 3d 981, 991 (S.D. Tex. 2017) (holding that "Facebook [has a] First Amendment right to decide what to publish and what not to publish on its platform" and collecting cases); Zhang v. Baidu.com, Inc., 10 F. Supp. 3d 433 (S.D.N.Y. 2014) (rejecting claim seeking to hold a website "liable for . . . a conscious decision to design its search-engine algorithms to favor a certain expression on core political subjects. . .[because] to proceed would plainly 'violate the fundamental rule of protection under the First Amendment, that a speaker has the autonomy to choose the content of his own message.'").

tations on our collective ability to address harmful content and protect people online.

Question 5. How many content posts or videos are generated by third-party users per day on Facebook, Twitter, and YouTube? b. How many decisions on average per day does your platform take to moderate content? Are you able to provide data on your takedown numbers over the last year? c. Do you ever make mistakes in a moderation decision? If so, how do you become aware of these mistakes and what actions do you take to correct them? d. What remedies or appeal process do you provide to your users to appeal an action taken against them? On average, how long does the adjudication take until a final action is taken? How quickly do you provide a response to moderation decision appeals from your customers? e. Can you provide approximate numbers, by month or week, for the times you took down, blocked, or tagged material from November 2019 to November 2020?

Answer. An important component of our transparency efforts is the *Twitter Transparency Center.* This year, we expanded our biannual transparency report site to become a comprehensive Twitter Transparency Center. Our goal with this evolution is make our transparency reporting more easily understood and accessible to the general public. This site includes data visualizations making it easier to compare trends over time and more information for the individuals who use Twitter, academics, researchers, civil society groups and others who study what we do to understand bigger societal issues. The Center includes data on enforcement actions under the Twitter Rules that requires the removal of specific Tweets or to suspend accounts. The Center also includes sections covering information requests, removal requests, copyright notices, trademark notices, e-mail security, platform manipulation, and state-backed information operations. We believe it is now more important than ever to be transparent about our practices.

If an account was suspended or locked in error, an individual can appeal. First, the individual must log in to the account that is suspended and file an appeal. The individual must describe the nature of the appeal and provide an explanation of why the account is not in violation of the Twitter Rules. Twitter employees will typically engage with the account holder via e-mail to resolve the appeal.

Question 6. The first major case to decide the application of Section 230 was *Zeran v. AOL* . In *Zeran,* Judge Wilkinson recognized the challenges of conferring "distributor liability" to a website because of the sheer number of postings. That was 1997. If we imposed a form of "distributor liability" on your platforms that would likely mean that your platform would be liable for content if you acquired knowledge of the content. I think there is an argument to be made that you "acquire knowledge" when a user "flags" a post, video, or other form of content. f. How many "user-generated" flags do your companies receive daily? g. Do users ever flag posts solely because they disagree with the content? h. If you were liable for content that was "flagged" by a user, how would that affect content moderation on your platform?

Answer. More than 186 million people last quarter used Twitter each day in dozens of languages and countless cultural contexts. The *Twitter Transparency Center,* described in response to Question 5, can provide a sense of the scale of our efforts to safeguard the conversation and enforce our policies.

Question 7. Section 230 is often used as a legal tool to have lawsuits dismissed in a pre-trial motion. i. How often is your company sued under a theory that you should be responsible for the content posted by a user of your platform? How often do you use Section 230 as a defense in these lawsuits? And roughly how often are those lawsuits thrown out? j. If Section 230 was eliminated and a case seeking to make your platform liable for content posted by a third party went to the discovery phase, roughly how much more expensive would that case be as opposed to its dismissal pre-discovery?

Answer. Twitter is currently involved in, and may in the future be involved in, legal proceedings, claims, investigations, and government inquiries and investigations arising in the ordinary course of business. These proceedings, which include both individual and class action litigation and administrative proceedings, have included, but are not limited to matters involving content on the platform, intellectual property, privacy, data protection, consumer protection, securities, employment, and contractual rights. We are happy to work with your teams to provide feedback on the impact on specific proposals that seek to amend Section 230.

Question 8. Section 230s Good Samaritan provision contains the term "otherwise objectionable." k. How do you define "otherwise objectionable"? l. Is "otherwise objectionable" defined in your terms of service? If so, has its definition ever changed? And if so, can you provide the dates of such changes and the text of each definition? m. In most litigation, a defendant relies on Section 230(c)(1) for editorial decisions. If a company could only rely on 230(c)(2) for a moderation decision (as has been dis-

cussed in Congress), how would that affect your moderation practices? And how would striking "otherwise objectionable" from 230(c)(2) further affect your moderation practices?

Answer. The specific details of a proposal are critical to understanding its potential impact on our business practices. We are happy to work with you further to provide additional feedback on the potential impacts of specific proposals.

Question 9. Are your terms of service a legally binding contract with your users? How many times have you changed your terms of service in the past five years? What recourse do users of your platform have when you allege that they have violated your terms of service?

Answer. The Twitter Rules and all incorporated policies, Privacy Policy, and Terms of Service collectively make up the "Twitter User Agreement" that governs an individual's access to and use of Twitter's services. We have the Twitter Rules in place to help ensure everyone feels safe expressing their beliefs and we strive to enforce them with uniform consistency. We are continually working to update, refine, and improve both our enforcement and our policies, informed by in-depth research around trends in online behavior both on and off Twitter, feedback from the people who use Twitter. We have updated our Terms of Service six times in five years, and all versions are available *here.*

We strive to give people an easy, clear way to appeal decisions we make that they think are not right. Mistakes in enforcement—made either by a human or algorithm—are inevitable, and why we strive to make appeals easier. We look forward to working with the Committee on efforts to enhance procedural fairness, including through a straightforward appeals process.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. RON JOHNSON TO JACK DORSEY

Question 1. During the hearing, in response to both Senator Cruz's line of questioning and mine, you claimed that Twitter does not have the ability to influence nor interfere in the election. a. In hindsight, do you stand by this claim? To reiterate, I am not asking if you have the intent or have actively taken steps to influence/interfere, but rather if Twitter has the ability. b. If you continue to claim that you do not have the ability to influence or interfere in the election, can you explain Twitter's rational for suppressing content that Twitter deems to be Russian misinformation on the basis that it influences the election?

Answer. We believe that we have a responsibility to protect the integrity of conversations related to elections and other civic events from interference and manipulation. Therefore, we prohibit attempts to use our services to manipulate or disrupt civic processes, including through the distribution of false or misleading information about the procedures or circumstances around participation in a civic process. Combatting attempts to interfere in conversations on Twitter remains a top priority for the company, and we continue to invest heavily in our detection, disruption, and transparency efforts related to state-backed information operations. Twitter defines state-backed information operations as coordinated platform manipulation efforts that can be attributed with a high degree of confidence to state-affiliated actors. Our goal is to remove bad faith actors and to advance public understanding of these critical topics.

Question 2. In Senator Rosen's testimony, Mr. Zuckerberg stated that Congress could hold Facebook accountable by monitoring the percentage of users that see harmful content before Facebook acts to take it down. While this is important, it does not address the problem of Facebook biasedly and inconsistently enforcing content moderation policies of political speech. Unfortunately, Twitter has the same biases and inconsistencies. c. In regards to this issue, what role do you think Congress should have in holding Twitter accountable? d. Do you have an example of a mechanism by which Congress can currently hold Twitter accountable on this issue? If there are none, can you please at a minimum acknowledge that there are none?

Answer. We ensure that all decisions are made at Twitter without using political viewpoints, party affiliation, or political ideology, whether related to automatically ranking content on our service or how we develop or enforce the Twitter Rules. Our Twitter Rules are not based on ideology or a particular set of beliefs. We believe strongly in being impartial, and we strive to enforce our Twitter Rules fairly. As detailed in our testimony, we fully support efforts to enhance procedural fairness, so that the public can trust that rules are developed and applied consistently and equitably.

Question 3. When Senator Gardner asked you about misinformation related to holocaust deniers, you said "We have a policy against misinformation in 3 categories: 1) manipulated media, 2) public health, specifically Covid, and 3) civic integrity, election interference, and voter suppression. That is all we have a policy on for misleading information." The puppy tweet example I raised in my testimony falls squarely into the 3rd category. When my staff brought this to your company's attention, we got the following response, "Thanks for reaching out. We escalated this to our Support Team for their review and they have determined that it is not in violation of our Policies." e. Can you please further explain your company's decision? How does this not violate your misinformation policy?

Answer. Twitter strongly condemns antisemitism, and hateful conduct has absolutely no place on the service. At the time of testimony, Mr. Dorsey did not mention that Twitter's Hateful Conduct Policy prohibits a wide range of behavior, including making references to violent events or types of violence where protected categories of people were the primary victims, or attempts to deny or diminish such events. Twitter also has a robust policy against glorification of violence and takes action against content that glorifies or praises historical acts of violence and genocide, including the Holocaust.

Regarding the specific Tweet you referenced, not all false or untrue information about politics or civic processes constitutes manipulation or interference that violates our *Civic Integrity Policy*. Our policy specifically references that "inaccurate statements about an elected or appointed official" are generally not in violation of this policy.

Question 4. During your testimony you said that Twitter should "enable people to choose algorithms created by 3rd parties to rank and filter their own content," in reference to Dr. Stephen Wolfram's research. f. Which of the methods described in his research and testimony have you deployed on your platform? g. What other methods would you like to see put in place? h. What is preventing you from implementing more methods such as these?

Answer. We are inspired by the approach suggested by Dr. Stephen Wolfram, Founder and Chief Executive Officer of Wolfram Research, in his testimony before the Subcommittee on Communications, Technology, Innovation, and the Internet in June 2019. Enabling people to choose algorithms created by third parties to rank and filter their content is an incredibly energizing idea that is in reach. As a first step, in December 2018, Twitter introduced an icon located at the top of everyone's timelines that allows individuals using Twitter to easily switch to a reverse chronological order ranking of the Tweets from accounts or topics they follow. This improvement gives people more control over the content they see, and it also provides greater transparency into how our algorithms affect what they see. We believe this points to an exciting, market-driven approach where people can choose what algorithms filter their content so they can have the experience they want.

Question 5. Do you agree that Twitter competes with local newspapers and broadcasters for local advertising dollars? i. Should Congress allow local news affiliates, such as local newspapers and local broadcast stations, to jointly negotiate with Twitter for fair market compensation for the content they create when it is distributed over your platform?

Answer. Providing news media organizations an effective antitrust exemption to collectively negotiate against other online platforms creates a situation where the media organizations gain leverage in negotiations by banding together. Ultimately, the online companies that can then afford to pay these heightened premiums for content are the large, dominant companies. It is critical that smaller online companies are not effectively shut out of these discussions.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. MARIA CANTWELL TO JACK DORSEY

Foreign Disinformation. Facebook/Instagram, Twitter, and Google/YouTube have each taken concrete steps to improve defensive measures through automated detection and removal of fake accounts at creation; increased internal auditing and detection efforts; and established or enhanced security and integrity teams who can identify leads and analyze potential networks engaging in coordinated inauthentic behavior.

Social media companies have hired a lot of staff and assembled large teams to do this important work and coordinate with the FBI led Foreign Influence Task Force (FITF).

Small companies in the tech sector do not have the same level or expertise or resources, but they face some of the same and growing threats.

Likewise, public awareness and understanding of the threats foreign actors like Russia pose is key to helping fight back against them.

Question 1. What specific steps are you taking to share threat information with smaller social media companies that do not have the same level of resources to detect and stop those threats?

Answer. Information sharing and collaboration are critical to Twitter's success in preventing hostile foreign actors from disrupting meaningful political conversations on the service. We have significantly deepened our partnership with industry peers, establishing formal processes for information sharing with both large and small companies and a regular cadence of discussion about shared threats. In addition to this, as explained in more detail in our response to Question 2, we publicly release data about action we have taken on state-backed information operations, to better enable the public, researchers, and others to better understand these threats.

In addition to these efforts, we have well-established relationships with Federal government agencies, including the Federal Bureau of Investigation Foreign Influence Task Force and the U.S. Department of Homeland Security's Election Security Task Force. For example, on Election Day and the week after, we virtually participated in an operations center run by the Federal Bureau of Investigation, including officials from the Department of Justice, the Department of Homeland Security, the Office of the Director of National Intelligence, and private sector partners.

Question 2. Intel Chairman Schiff has highlighted the need for social media companies to increase transparency about how social media companies have stopped foreign actors disinformation and influence operations. Where are the gaps in public disclosures of this information and what specific actions are you taking to increase transparency about malign foreign threats you have throttled?

Answer. Combatting attempts to interfere in conversations on Twitter remains a top priority for the company, and we continue to invest heavily in our detection, disruption, and transparency efforts related to state-backed information operations. Our goal is to remove bad faith actors and to advance public understanding of these critical topics. We recognize that to better enable the public, lawmakers, and researchers to better understand the threat of state-backed information operations, data is essential. In *October 2018,* we published the first comprehensive archive of Tweets and media associated with known state-backed information operations on Twitter.

We now routinely update this archive to add additional datasets when we disrupt state-backed information operations. This one of a kind resource now spans operations across 15 countries, including more than nine terabytes of media and 200 million Tweets.

It is important to remember that these bad faith actors work across multiple platforms, leveraging a range of services including ad tech, hosting, content distribution, and payment processing. These services can also be used by bad faith actors who create financial incentives to distribute disinformation content or are derived from a state-backed information operation. We support calls for other parts of industry to add their own data to the publicly accessible information Twitter discloses. We also encourage continued disclosures from government partners about activity they observe, including on social media, to enable the wider community to build their own resilience and to expose to the public the tactics used.

Addressing Stop Hate for Profit Recommendations. The Stop Hate for Profit, Change the Terms, and Free Press coalition—all committed to combating racism, violence, and hate online—have called on social media platforms to adopt policies and take decisive actions against toxic and hateful activities.

This includes finding and removing public and private groups focused on white supremacy, promoting violent conspiracies, or other hateful content; submitting to regular, third party, independent audits to share information about misinformation; changing corporate policies and elevating a civil rights to an executive level position.

Question 3. Mr. Dorsey, will you commit to making the removal of racist, violent, and hateful content an executive level priority?

Answer. Ensuring that individuals who use Twitter feel safe using our service is an executive level priority. Currently, we have numerous policies aimed at combating hateful conduct. Our *Hateful Conduct Policy* prohibits promotion of violence against people on the basis of protected categories. This includes wishing or calling for serious harm against a person or group or people, or targeting individuals with the intent to harass with references to violent events or types of violence where protected categories were the primary victims. We recently *expanded* this policy to prohibit language that dehumanizes people on the basis of race, ethnicity, or national origin. In addition, our *Glorification of Violence Policy* prohibits content that glorifies or praises historical acts of violence and genocide, including the Holocaust. And,

our *Violent Organizations Policy* prohibits violent organizations, including violent extremist groups, from using Twitter to promote violence.

Kenosha Wisconsin Violence. On August 25th, a man from Illinois traveled to Kenosha, Wisconsin armed with an assault rifle and fatally shot Joseph Rosenbaum and Anthony Huber, and injured another person, who were protesting the shooting of Jacob Blake, a Black resident, which left him paralyzed.

In the wake of these tragic shootings, we learned that a para-military group called the Kenosha Guard Militia, a group that organized on Facebook, called on followers to "take up arms" and "defend" the city against "evil thugs". This event post had been flagged 455 times by Facebook users, yet Facebook did not take down the group's page until after these lives were already lost.

While the Illinois shooter may not have been a member of the Kenosha Guard Militia, this brings up a very important point—that hate spread on social media platforms can lead to real life violence.

In May of this year, the *Wall Street Journal* reported that Facebook had completed internal research that said its internal algorithms "exploit the human brain's attraction to divisiveness", which could allow Facebook to feed more divisive content to gain user attention and more time on the platform. In response, the *Journal* reported that Facebook buried the research and did little to address it because it ran counter to other Facebook initiatives.

Sowing divisions in this country and further polarizing public discourse is dangerous, and can have deadly consequences.

Question 4. Mr. Dorsey, Twitter also targets information to people based on what your data tells them they want to see, which can lead to people being stuck in an echo chamber that makes them less likely to listen to other viewpoints. What responsibility do you believe you have to stem the divisive discourse in this country?

Answer. Twitter's purpose is to serve the public conversation. People from around the world come together on Twitter in an open and free exchange of ideas. We want to make sure conversations on Twitter are healthy and that people feel safe in expressing their points of view. Our role is not to tell people what is the truth or to dictate their experience on our service. On the contrary, we (1) invest in efforts to provide individuals choices about the algorithms that affect their online experience, and (2) work to ensure our approach to harmful misinformation is aimed at providing people additional context to help inform their views.

In December 2018, Twitter introduced an icon located at the top of everyone's timelines that allows individuals using Twitter to easily switch to a reverse chronological order ranking of the Tweets from accounts or topics they follow. This improvement gives people more control over the content they see, and it also provides greater transparency into how our algorithms affect what they see. We believe this points to an exciting, market-driven approach where people can choose what algorithms filter their content so they can have the experience they want.

In addition, we have conducted research aimed at identifying the needs of people who use Twitter, including how to address potential misinformation on the platform and how to ensure that people have access to credible information to inform their viewpoints. A key piece of feedback we got from the people who use our service is that we should not determine the truthfulness of Tweets; we should provide context to help people make up their own minds in cases where the substance of a Tweet is disputed. Consistent with this feedback, we continue to explore ways to expand our enforcement options beyond a binary choice of leaving content up or taking it down. Instead, we are looking at ways to ensure that people can access additional information to inform their viewpoints.

We have expanded our enforcement options to allow us to label misinformation related to manipulated media, COVID–19, and civic integrity. When we label Tweets, we link to Twitter conversation that shows three things for context: (1) factual statements; (2) counterpoint opinions and perspectives; and (3) ongoing public conversation around the issue. We continue to explore additional improvements to our service that we can make to ensure individuals have access to credible information and diverse viewpoints.

Notwithstanding these efforts, we recognize that Twitter is only part of a larger ecosystem that impacts the online discourse. For example, a *Knight Foundation study* found that "evidence for echo chambers is actually strongest in offline social networks, which can increase exposure to like-minded views and information and amplify partisan messages." The study also found that "several studies have found evidence for offline echo chambers that are as strong or stronger than those documented online."[1]

Russian Election Interference. The U.S. Intelligence community found that foreign actors including Russia tried to interfere in the 2016 election and used social media platforms among other influence operations.

In 2017, the FBI established the Foreign Influence Task Force (FITF), which works closely with state and local partners to share information on threats and actionable leads.

The FBI has also established relationships with social media companies to enable rapid sharing of threat information. Social media companies independently make decisions regarding the content of their platforms.

The U.S. Intelligence Community warned that Russia was using a range of active measures to denigrate former Vice President Joe Biden in the 2020 election. They also warned about Iran and China.

Social media companies remain on the front lines of these threats to our democracy.

Question 5. What steps are you taking to prevent amplification of false voter fraud claims after the 2020 presidential election and for future elections? What challenges do you face trying to prevent foreign actors who seek to influence our elections?

Answer. Twitter has several policies aimed at protecting the conversation around the 2020 U.S. elections and elections around the world. Notably, our civic integrity policy provides for the labeling or removing of false and misleading information related to civic processes. The policy prohibits false or misleading information about how to participate in an election or civic process; content intended to intimidate or dissuade people from participating; misrepresentation about affiliation (for ex., a candidate or political party); content that causes confusion about laws and regulations of a civic process or officials and institutions executing those civic processes; disputed claims that could undermine public confidence in the election including unverified information about election rigging, ballot tampering, vote tallying, or certification of election results; and content that misleads about outcomes (*e.g.,* claiming victory before results are in or inciting unlawful conduct to prevent the procedural or practical implementation of election results).

In 2020, we found that many of the efforts to disrupt the conversation around the election originated domestically. While technology companies play an important role in safeguarding against efforts to undermine the integrity of the conversation regarding civic integrity, they cannot address all facets of complex social issues. Looking forward, we need broader conversations around government institutional preparedness, media literacy, and other efforts that are necessary to build civic resilience.

Question 6. How the U.S. Government improved information sharing about threats from foreign actors seeking to interfere in our elections since 2016? Is information that is shared timely and actionable? What more can be done to improve the cooperation to stop threats from bad actors?

Answer. As explained in the response to Question 2, we have strengthened our relationship with government agencies, including the Federal Bureau of Investigations (FBI), to better share information about foreign efforts to interfere with elections. In recent months we detected limited state backed information operations tied to the conversation regarding the 2020 U.S. election. On September 1, we suspended five Twitter accounts for platform manipulation that we can reliably attribute to Russian state actors based on information we received from the FBI. On September 24, we also permanently suspended two distinct networks of accounts that we can reliably attribute to state-linked entities in Russia.

On September 29, based on intelligence provided by the FBI, we removed 132 accounts that appeared to originate in Iran. The accounts were attempting to disrupt the public conversation during the presidential debate. Additionally, on October 8, we announced we had identified a network of primarily compromised accounts on Twitter operating from Iran. These accounts artificially amplified conversations on politically sensitive topics, including the Black Lives Matter movement, the death of George Floyd, and other issues of racial and social justice in the United States.

Question 7. How are you working with civil society groups like the University of Washington's Center for an Informed Public and Stanford Internet Observatory and Program?

Answer. In order to safeguard the conversation regarding the 2020 U.S. election, we have developed critical partnerships with government entities, news organizations, civil society, and others. These partnerships have been instrumental in informing policies and helping to identify potential threats regarding the integrity of the election conversation occurring on Twitter. In advance of the general election, we partnered with the Election Integrity Partnership, a coalition of research entities focused on supporting real-time information exchange to detect and mitigate the im-

pact of attempts to prevent or deter people from voting or to delegitimize election results. The foundational Partnership consists of four leading institutions: the Stanford Internet Observatory and Program on Democracy and the Internet, Graphika, the Atlantic Council's Digital Forensic Research Lab, and the University of Washington's Center for an Informed Public.

Question 8. How are you raising social media users' awareness about these threats? What more can be done? How do you ensure the actions you take do not cross the line into censorship of legitimate free speech?

Answer. As explained more fully in the response to Question 4, we are working to expand our enforcement options to better enable us to serve the public conversation. As part of that, we are working to expand our enforcement options beyond a binary choice of leaving content up or taking it down, to include options that provide additional context to the people who use Twitter about the content they see on our service. We believe these efforts will provide us more flexibility to address harmful misinformation while safeguarding free expression.

In addition, in certain contexts, we have worked to curb potentially harmful misinformation by facilitating the ability of individuals to access information from official and credible sources. For example, prior to the November election, we showed everyone on Twitter in the U.S. a *series of pre-bunk prompts.* These prompts, which were seen 389 million times, appeared in people's home timelines and in Search, and reminded people that election results were likely to be delayed and that voting by mail is safe and legitimate.

Foreign Disinformation & Russian Election Interference. Since four years ago, our national security agencies and the private sector have made improvements to address foreign cyber and influence efforts that target our electoral process. However, there still needs to be more public transparency about foreign disinformation.

We need to close any gaps to stop any foreign disinformation about the 2020 election and disinformation in future elections. We can not allow the Russians or other foreign actors to try to delegitimize election results or exacerbate political divisions any further.

Question 9. What more could be done to maximize transparency with the public about suspected foreign malign activity?

Answer. As described in our response to Question 2, when we identify inauthentic activity on Twitter that meets our definition of an information operation, and we are able to confidently attribute it to actors associated with a government, we share comprehensive data about this activity publicly. We encourage our industry peers to engage in similar transparency efforts.

Question 10. How could you share more information about foreign disinformation threats among the private sector tech community and among social media platforms and with smaller companies?

Answer. Since the 2016 elections, we have significantly deepened our partnership with industry peers, establishing formal processes for information sharing and a regular cadence of discussion about shared threats. These collaborations are a powerful way to identify and mitigate malicious activity that is not restricted to a single platform or service. We have worked to share information with smaller companies, and continue to encourage our government partners to expand the companies with whom they meet regularly.

Question 11. What should the U.S. Government be doing to promote information sharing on threats and to increase lawful data-sharing about suspected foreign malign activity?

Answer. Information sharing and collaboration are critical to Twitter's success in preventing hostile foreign actors from disrupting meaningful political conversations on the service. As described in the response to Question 2, we have well-established relationships with law enforcement agencies active in this area, and we have mechanisms in place to share classified information. We encourage our government partners to continue to share as much information as possible with us because in certain circumstances only they have access to information critical to joint efforts to stop bad faith actors. We also encourage continued efforts to ensure relevant information is declassified, as appropriate, to allow industry and the public to fully understand potential threats.

Rohingya/Myanmar. In 2018, Facebook was weaponized against to whip up hate against the Muslim minority—the Rohingya. Myanmar held a general election last month. Prior to that election, there were concerns about the integrity of that election.

Question 12. What did you do and how are you continuing to make sure social media is not abused by any foreign or domestic actors to distort the electoral process in Myanmar and other countries?

Answer. The public conversation occurring on Twitter is never more important than during elections, the cornerstone of democracies across the globe. As platform manipulation tactics evolve, we are continuously updating and expanding our rules to better reflect what types of inauthentic activity violate our guidelines. We continue to develop and acquire sophisticated detection tools and systems to combat malicious automation on our service.

Individuals are not permitted to use Twitter in a manner intended to artificially amplify, suppress information, or engage in behavior that manipulates or disrupts other people's experience on the service. We do not allow spam or platform manipulation, such as bulk, aggressive, or deceptive activity that misleads others and disrupts their experience on Twitter. We also prohibit the creation or use of fake accounts.

Impact of S. 4534. As you are aware, Chairman Wicker and two of our Republican colleagues have offered legislation to amend Section 230 to address, among other issues, what they call 'repeated instances of censorship targeting conservative voices."

That legislation would make significant changes to how Section 230 works, including limiting the categories of content that Section 230 immunity would cover and making the legal standard for removal of content more stringent. Critics of the Chairman's bill, S. 4534, suggest that these changes would inhibit companies' ability to remove false or harmful content from their platforms.

Question 13. I would like you to respond yes or no as to whether you believe that bills like the Chairman's would make it more difficult for Twitter to remove the following types of content—

- Bullying?
- Election disinformation?
- Misinformation or disinformation related to COVID–19?
- Foreign interference in U.S. elections?
- Efforts to engage in platform manipulation?
- Hate speech?
- Offensive content directed at vulnerable communities or other dehumanizing content?

Answer. Section 230 has been critical to Twitter's efforts to safeguard the public conversation, including our efforts to combat hateful conduct, harmful misinformation in the civic integrity and COVID–19 contexts, doxxing, platform manipulation, and foreign interference. The law allows us to act responsibly to promote healthy conversations by removing content that contains abuse, foreign interference, or illegal conduct, among other categories of content that violates our rules. Eliminating Section 230s important liability protections could discourage platforms like ours from engaging in content moderation for the purpose of protecting public safety and ensuring that our service is a safe place for all voices.

Combating "Garbage" Content. Santa Clara University Law Professor Eric Goldman, a leading scholar on Section 230, has argued that the Online Freedom and Viewpoint Diversity Act (S. 4534) wants Internet services to act as "passive" receptacles for users' content rather than content curators or screeners of "lawful but awful" third-party content.

He argues that the bill would be counterproductive because we need *less* of what he calls "garbage" content on the Internet, not more. Section 230 lets Internet services figure out the best ways to combat online trolls, and many services have innovated and invested more in improving their content moderation functions over the past few years.

Professor Goldman specifically points out that the bill would make it more difficult for social media companies to remove "junk science/conspiracy theories, like anti-vax content or quack COVID19 cures."

Question 14. Would S. 4534—and similar bills—hurt efforts by Twitter to combat online trolls and to fight what Professor Goldman calls "lawful but awful . . . garbage" content?

Answer. As noted in Question 13, eliminating Section 230s protections would make it more difficult for platforms to address a range of harmful issues, including harmful misinformation. As explained in more detail in our written testimony, we do not believe that the solution to concerns raised about content moderation is to

eliminate Section 230 liability protections. Instead, we believe the solution should be focused on enhancing transparency, procedural fairness, privacy, and algorithmic choice, which can be achieved through additions to Section 230, industry-wide self-regulation best practices, or additional legislative frameworks.

The FCC's Capitulation to Trump's Section 230 Strategy. The Chairman of the Federal Communications Commission, Ajit Pai, announced recently that he would heed President Trump's call to start a rulemaking to "clarify" certain terms in Section 230.

And reports suggest that the President pulled the renomination of a sitting FCC Commissioner due to his concerns about that rulemaking, replacing him with a nominee that helped develop the Administration's petition that is the foundation of this rulemaking. This capitulation to President Trump by a supposedly independent regulatory agency is appalling.

It is particularly troubling that I—and other members of this committee—have been pressing Chairman Pai to push the envelope to interpret the agency's existing statutory authority to, among other things, use the E-Rate program to close the homework gap, which has only gotten more severe as a result of remote learning, and to use the agency's existing authority to close the digital divide on Tribal lands. And we expressed serious concern about Chairman Pai's move to repeal net neutrality, which the FCC majority based upon a highly conservative reading of agency's statutory authority.

In contrast, Chairman Pai is now willing to take an expansive view of the agency's authority when asked to support the President's pressure campaign against social media in an attempt not to fact check or label the President's posts.

Question 15. What are your views on Chairman Pai's announced rulemaking and the FCC's legal analysis of section 230? Would you agree that his approach on this issue is in tension with his repeal of the essential consumer protections afforded by the net neutrality rules?

Answer. Section 230 protects American innovation and freedom of expression. Attempts to erode the foundation of Section 230 could collapse how we communicate on the Internet, leaving only a small number of giant and well-funded technology companies. Twitter strongly supports net neutrality, and we encourage the FCC to instead take steps to reinstate the repealed net neutrality order. In addition, we support the *Save the Internet Act,* which would enshrine net neutrality in law.

Addressing Bad Actors. I have become increasingly concerned with how easy it is for bad actors to use social media platforms to achieve their ends, and how they have been too slow to stop it. For example, a video touting antimalarial drug hydroxychloroquine as a "cure" for COVID was eventually taken down this summer—but not after garnering 17 million views on Facebook.

In May, the watchdog group Tech Transparency Project concluded that white supremacist groups are "thriving" on Facebook, despite assurances that Facebook does not allow such groups on its platform.

These are obviously troubling developments, especially in light of the millions of Americans that rely on social media services. You have to do better.

That said, I am not sure that modifying Section 230 is the solution for these and other very real concerns about your industry's behavior.

Question 16. From your company's perspective, would modifying Section 230 prevent bad actors from engaging in harmful conduct?

Answer. Though we have an important role to play, Twitter is only one part of a larger ecosystem that can address the harmful conduct of bad faith actors. Notwithstanding this, repealing Section 230s protections will likely make it more difficult to combat harm online. For example, Section 230 has been instrumental in allowing Twitter to take action to prevent harm, including through our policies aimed at prohibiting violation organizations, terrorist content, and COVID–19 misinformation.

We continue to improve and expand our policies to better combat harmful conduct. For example, under our *Violent Organizations Policy,* we have removed more than 200 groups from the platform, half of which had links to white supremacy, and permanently suspended 1.7 million unique accounts for violating our policy prohibiting the promotion of terrorism. In addition, to address the global pandemic, on March 16, 2020, we announced *new enforcement guidance,* broadening our definition of harm to address content related to COVID–19 that goes directly against guidance from authoritative sources of global and local public health information. Section 230 enables us to require individuals to remove violative Tweets in a variety of contexts with the goal of preventing offline harm.

Question 17. What do you recommend be done to address the concerns raised by the critics of Section 230?

Answer. We believe the solution to concerns raised by critics regarding our content moderation efforts is not to eliminate Section 230, but rather to build on its foundation. Specifically, we support requiring the publication of moderation processes and practices, a straightforward process to appeal decisions, and best efforts around algorithmic choice, while protecting the privacy of people who use the Internet. We look forward to working with the Committee to achieve these objectives.

Question 18. How do you expect that Twitter would react when faced with increased possibility of litigation over user-submitted content?

Answer. The specifics of any proposal are necessary to assess its potential impact on Twitter. However, we are concerned with proposals that seek to roll back essential Section 230 protections, which have been critical to promoting innovation, ensuring competition, and safeguarding free expression. Specifically, increasing liability for content generated by the people who use the platforms could encourage overmoderation, likely resulting in increased removal of speech. Such an outcome does not align with First Amendment values, nor does it address existing concerns. Instead of eliminating existing liability protections, future solutions should be focused on ensuring transparency, procedural fairness, algorithmic choice, and privacy.

Trump Administration Records. Mr. Dorsey, over the course of nearly four years, President Trump and senior officials in his administration have routinely used Twitter to conduct government business including announcing key policy and personnel decisions on those platforms. In addition, many believe that President Trump and his senior aides have used Twitter to engage in unethical and sometimes illegal conduct.

For example, Special Counsel Mueller cited several of President Trump's tweets as evidence of potentially obstructive conduct, and senior White House aides such as Kellyanne Conway and Ivanka Trump have been cited for violations of the Hatch Act and the misuse of position statute based on their use of Twitter in the conduct of their government jobs. Meanwhile, it appears that on several occasions Twitter has changed or ignored its rules and policies in ways that have allowed administration officials to continue using the platform to violate the rules for government employees and other Twitter users. While government officials are legally obligated to preserve presidential and Federal records created or stored on social media platforms, this administration's actions cast serious doubts on whether they will comply with those obligations, and in many instances, they have already failed to do so. Twitter could play a vital role in ensuring that the historical record of the Trump administration is accessible to the American public, Congress, and other government institutions so that people are "able to see and debate" the "words and actions" of the Trump presidency as well as future presidential administrations.

Question 19. Please describe what steps, if any, Twitter has taken to ensure that Twitter content—including tweets and direct messages—sent or received by Trump administration officials on Twitter accounts used for official government business are collected and preserved by your company.

Answer. Twitter is actively preparing to support the transition of institutional White House Twitter accounts on January 20, 2021. As we did for the U.S. presidential transition in 2017, this process is being done in close consultation with the National Archives and Records Administration (NARA), who has historically played a significant role in appropriately preserving records following administration transitions. Twitter provides technical support to NARA and other government entities to transfer accounts in order to facilitate preservation efforts. We defer to these entities to make critical decisions about the information and accounts to preserve, agency compliance with existing legal obligations, and efforts to make this information available to the public.

In addition, all Twitter account holders can *download a machine-readable archive of information associated with their account.* This *archive dataset* includes an account's profile information, Tweets, Direct Messages, Moments, media (images, videos, and GIFs attached to Tweets, Direct Messages, or Moments), a list of followers, a list of followed accounts, address book, lists an account has created, been a member of or follow, inferred interest and demographic information, and more. We have advised NARA with regard to this self service tool for account holders and how to request archives for each account.

Question 20. Please describe what steps, if any, Twitter has taken to ensure that the National Archives and Records Administration can obtain and preserve all Twitter content—including tweets and direct messages—sent or received by Trump administration officials on Twitter accounts used for official government business.

Answer. Please see the answer to Question 20.

Question 21. Please describe what steps, if any, Twitter has taken to ensure that the White House can preserve all Twitter content—including tweets and direct messages—sent or received by Trump administration officials on Twitter accounts used for official government business.

Answer. Please see the answer to Question 20.

Question 22. Will you commit to ensuring that all Twitter content—including tweets and direct messages—sent or received by Trump administration officials on Twitter accounts used for official government business are collected and preserved by your company?

Answer. Please see the answer to Question 20.

Online Disinformation. I have serious concerns about the unchecked spread of disinformation online. From false political claims to harmful health information, each day the problem seems to get worse and worse. And I do not believe that social media companies—who make billions of dollars from ads based in part on user views of this disinformation—are giving this problem the serious attention that it deserves.

Question 23. Do you agree that Twitter can and should do more to stop the spread of harmful online disinformation?

Answer. As outlined in the responses to Question 4 and Question 8, we have taken significant steps to combat harmful misinformation online as it relates to COVID–19, civic integrity, and synthetic and manipulated media. We continue to improve and expand these policies, and welcome working with the Committee to address these issues more broadly.

Question 24. Can you commit that Twitter will take more aggressive steps to stop the spread of this disinformation? What specific additional actions will you take?

Answer. We continue to expand our policies on misinformation, prioritizing our work on the areas with the greatest likelihood of harm. For example, in December 2020, we updated our policy approach to misleading information about COVID–19, originally issued in March 2020. Beginning December 21, we may require people to remove Tweets which advance harmful false or misleading narratives about COVID–19 vaccinations, including:

- False claims that suggest immunizations and vaccines are used to intentionally cause harm to or control populations, including statements about vaccines that invoke a deliberate conspiracy;
- False claims which have been widely debunked about the adverse impacts or effects of receiving vaccinations; or
- False claims that COVID–19 is not real or not serious, and therefore that vaccinations are unnecessary.

Starting in early 2021, we may label or place a warning on Tweets that advance unsubstantiated rumors, disputed claims, and incomplete or out-of-context information about vaccines. Tweets that are labeled under this expanded guidance may link to authoritative public health information or *the Twitter Rules* to provide people with additional context and authoritative information about COVID–19. We will enforce this policy in close consultation with local, national, and global public health authorities around the world, and will strive to be iterative and transparent in our approach.

Question 25. What are the clickthrough rates on your labelling on disputed or fact-checked content related to civic integrity, either when content is hidden or merely labelled? What metrics do you use to gauge the effectiveness of labelling? Please share typical numerical values of the metrics you describe.

Answer. We continue to work to fully assess the impact of our misinformation policies, including in the civic integrity context. For example, an initial examination of our efforts from October 27 to November 11 around the U.S. 2020 election found that:

- Approximately 300,000 Tweets were labeled under our Civic Integrity Policy for content that was disputed and potentially misleading. These represent 0.2 percent of all U.S. election-related Tweets sent during this time period;
- 456 of those Tweets were also covered by a warning message and had engagement features limited (Tweets could be Quote Tweeted but not Retweeted, replied to, or liked); and
- Approximately 74 percent of the people who viewed those Tweets saw them after we applied a label or warning message.

We continue to assess the impact of our ongoing election efforts, and look forward to sharing this information with the Committee.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. RICHARD BLUMENTHAL TO JACK DORSEY

For the following questions, please provide information about your firm's content moderation decisions related to election misinformation and civic integrity covering the 2020 election period.

Question 1. Please describe what processes were used to make decisions about labeling or taking down organic and paid content related to elections or civic integrity.

Answer. In the lead up to the 2020 elections, we made significant enhancements to our policies to protect the integrity of the election. Most notably, this year, we updated our civic integrity policy to more comprehensively enforce labeling or removing of false and misleading information. The updated policy, which we not only announced publicly but also briefed the Presidential campaigns, civil society, and other stakeholders on, covers the following activities:

- False or misleading information about how to participate in an election or civic process;
- Content intended to intimidate or dissuade people from participating;
- Misrepresentation about affiliation (for ex., a candidate or political party);
- Content that causes confusion about laws and regulations of a civic process, or officials and institutions executing those civic processes;
- Disputes of claims that could undermine public confidence in the election (*e.g.,* unverified information about election rigging, ballot tampering, vote tallying, or certification of election results); and
- Content that misleads about outcomes (*e.g.,* claiming victory before results are in, inciting unlawful conduct to prevent the procedural or practical implementation of election results).

The civic integrity policy augmented and enhanced other important rules aimed at preventing interference with the election. Twitter banned all political advertising in 2019, only allowing some cause-based advertising for non-partisan civic engagement, in line with our belief that the reach of political speech should be earned, not bought. Additionally, we adopted rules prohibiting deceptively shared synthetic or manipulated media, sometimes referred to as "deep fakes," that may lead to serious offline harm; and labeling deceptive or synthetic media to provide additional context. Moreover, we have rules prohibiting platform manipulation, impersonation, hateful conduct, ban evasion, and attributed activity, among other harmful activities. We have also labeled specific government and state-media accounts from UN P–5 nation states, and plan to expand this effort in the near future.

Question 2. How many posts were reported or identified as potentially containing election misinformation or violations of civic integrity policies?

Answer. We continue to analyze and assess our efforts regarding the 2020 U.S. election, and have publicly released initial findings *here.* Our efforts to safeguard the conversation on Twitter regarding the 2020 U.S. election are ongoing. We will continue to work with the Committee to provide comprehensive information to the public regarding the activity we saw on our service.

Question 3. How many posts had enforcement action taken for containing election misinformation or violations of civic integrity policies?

Answer. Our efforts to safeguard the conversation on Twitter regarding the 2020 U.S. election are ongoing and we continue to apply labels, warnings, and additional restrictions to Tweets that included potentially misleading information about the election. During the period from October 27 to November 11, 2020, we labeled approximately 300,000 Tweets under our Civic Integrity Policy for content that was disputed and potentially misleading. These represent 0.2 percent of all U.S. election-related Tweets sent during this time period. Approximately 450 of those Tweets were also covered by a warning message and had engagement features limited, including Tweets could be Quote Tweeted but not Retweeted, replied to or liked. Approximately 74 percent of the people who viewed those Tweets saw them after we applied a label or warning message. We saw an estimated 29 percent decrease in Quote Tweets of these labeled Tweets due in part to a prompt that warned people prior to sharing.

Question 4. Who did your firm consult to draft and implement election misinformation and civic integrity policies?

Answer. During the 2020 U.S. Election, we cultivated close partnerships with election stakeholders that informed our approach, including collaborating with Federal and state government officials, election officials, industry peers, and civil society. We are provided ongoing briefings by the FBI and its Foreign Influence Task

Force regarding specific threats from foreign actors. We also participate in monthly briefings with U.S. government partners and industry peers to discuss the evolving threat landscape. We have established routine communications channels with the FBI, the Department of Justice, U.S. Department of Homeland Security, and the Office of the Director of National Intelligence. We have strengthened our relationships with the elections community, including state elections officials at the state and local levels non-governmental organizations, and advocacy groups. Twitter also works in close collaboration with our industry peers and we share and receive threat information on an ongoing basis. In addition, we have engaged key civil rights and civil liberties groups on elections related issues.

Question 5. Who made final decisions about labeling or taking down a post related to election misinformation or civic integrity? Who did that person or those persons consult?

Answer. Twitter uses a combination of machine learning and human review to adjudicate reports of violations and make determinations on whether the activity violates our rules. In addition, Twitter has created an internal cross-functional analytical team whose mission is to monitor site and service integrity. Drawing on expertise across the company, this team can respond immediately to escalations of inauthentic, malicious automated or human-coordinated activity on the service. To supplement its own analyses, Twitter's analytical team also receives and responds to reports from across the company and from external third parties. The results from all of the team's analyses are shared with key stakeholders at Twitter and provide the basis for policy changes, product initiatives, and the removal of accounts.

Question 6. Does a different or specialized process exist for content from Presidential candidates, and if so, how does that process for review differ from the normal review?

Answer. Our mission is to provide a forum that enables people to be informed and to engage their leaders directly. Everything we do starts with an understanding of our purpose and of the service we provide: a place where people can participate in public conversation and get informed about the world around them.

We assess reported Tweets from world leaders against the Twitter Rules, which are designed to ensure people can participate in the public conversation freely and safely. We focus on the language of reported Tweets and do not attempt to determine all potential interpretations of the content or its intent. In cases involving a world leader, we will err on the side of leaving the content up if there is a clear public interest in doing so. We place the violative content behind a warning notice.

Question 7. Based on enforcement actions taken, is there a discernible difference in engagement between a labeled post and unlabeled posts? Please provide any supporting information.

Answer. Yes, our analysis of the 300,000 Tweets that had a label or warning applied between October 27 and November 11, 2020 indicates that approximately 74 percent of the people who viewed those Tweets saw them after we applied a label or warning message. We also confronted potentially misleading information by showing everyone on Twitter in the United States a *series of pre-bunk prompts.* These prompts, which were seen 389 million times, appeared in people's home timelines and in Search, and reminded people that the announcement of election results were likely to be delayed, and that voting by mail is safe and legitimate. Finally, we encouraged people to add their own commentary when amplifying content by prompting Quote Tweets instead of Retweets. This change introduced some friction, and gave people an extra moment to consider why and what they were adding to the conversation. Since making this change, we observed a 23 percent decrease in Retweets and a 26 percent increase in Quote Tweets, but on a net basis the overall number of Retweets and Quote Tweets combined decreased by 20 percent. In short, this change slowed the spread of misleading information by virtue of an overall reduction in the amount of sharing on the service.

Question 8. What was the average time to add a misinformation label to a post?

Answer. Our efforts to safeguard the conversation on Twitter regarding the 2020 U.S. election are ongoing. We will continue to work with the Committee to provide comprehensive information to the public regarding the enforcement actions we took on our service.

For the following questions, please provide information about your firm's content moderation decisions related to hate speech, election interference, civic integrity, medical misinformation, or other harmful misinformation over the previous year.

Question 9. How many pieces of content were reported by users to the platform related to hate speech, election interference, civic integrity, and medical misinformation, broken down by category?

Answer. An important component of our transparency efforts is the *Twitter Transparency Center*. This year, we expanded our biannual transparency report site to become a comprehensive Twitter Transparency Center. Our goal with this evolution is make our transparency reporting more easily understood and accessible to the general public. This site includes data visualizations making it easier to compare trends over time and more information for the individuals who use Twitter, academics, researchers, civil society groups and others who study what we do to understand bigger societal issues. The Center includes data on enforcement actions under the Twitter Rules that requires the removal of specific Tweets or to suspend accounts. The Center also includes sections covering information requests, removal requests, copyright notices, trademark notices, e-mail security, platform manipulation, and state-backed information operations. We believe it is now more important than ever to be transparent about our practices.

Question 10. How many pieces of content were automatically identified or identified by employees related to hate speech, election interference, civic integrity, and medical misinformation, broken down by category?

Answer. Any individual, including employees, can report behavior that violates our terms of service directly from a Tweet, profile, or Direct Message. An individual navigates to the offending Tweet, account, or message and selects an icon that reports that it is harmful. Multiple Tweets can be included in the same report, helping us gain better context while investigating the issues to resolve them faster. For some types of report Twitter also prompts the individual to provide more information concerning the issue that is being reported. More than 50 percent of Tweets we take action on for abuse are now proactively surfaced using technology, rather than relying on reports to Twitter.

Question 11. Of the content reported or flagged for review, how many pieces of content were reviewed by humans?

Answer. Twitter uses a combination of machine learning and human review to adjudicate reports of violations and make determinations on whether the activity violates our rules. One of the underlying features of our approach is that we take a behavior-first approach. That is to say, we look at how accounts behave before we look at the content they are posting. This is how we were able to scale our efforts globally.

Question 12. How many pieces of content were subject to enforcement action? Please provide a break down for each type of enforcement action taken for each category.

Answer. Please see the response to Question 9.

Question 13. For content subject to enforcement action due to violation of hate speech rules, please identify how many pieces of content targeted each type of protected category (such as race or gender) covered by your rules. Do you track this information?

Answer. Twitter does not permit people on Twitter to promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. We also do not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories. As published in our *Twitter Transparency Center,* we took action on over 970,000 accounts for violations of our Hateful Conduct Policy between July and December 2019. Although we do not currently report statistics on what protected class may be targeted by hateful conduct. We continue to explore improvements to tracking the reporting of more detailed information regarding the enforcement of our Rules.

Response to Written Questions Submitted by Hon. Edward Markey to Jack Dorsey

Question 1. Mr. Dorsey, in 2018, you testified before the House Energy & Commerce Committee and committed to conducting a public civil rights audit at Twitter. Please provide a detailed status report on that audit. Please also describe in detail the steps Twitter will take to comply with the recommendations of that audit in a transparent manner.

Answer. We agree that third-party feedback and metrics can be valuable resources to inform our work. Our focus is not only assessment but building a framework both internally and externally to make substantive change over time. To that end, currently, several national organizations that represent the interests and defense of civil rights, serve in advisory roles on our Trust and Safety Council. In addi-

tion, we have also established a global, cross-functional group to inform and evaluate our work related to civil rights.

Question 2. Mr. Dorsey, children and teens are a uniquely vulnerable population online, and a comprehensive Federal privacy law should provide them with heightened data privacy protections. Do you agree that Congress should prohibit online behavioral advertising, or "targeted marketing" as defined in S.748, directed at children under the age of 13?

Answer. MoPub, a division of Twitter and Twitter International Company, provides advertising technology services that allow publishers of mobile applications to show ads in their apps, and for advertisers to reach audiences that may be interested in their products and services. MoPub's policies clearly prohibit the use of MoPub's services in violation of the Children's Online Privacy Protection Act, or COPPA. MoPub Publisher Partners—companies that develop mobile applications (or "apps" as they are more commonly known) and integrate the MoPub Services to show in-app advertising—are explicitly required to comply with COPPA in the collection and use of "Personal Information" from children under 13 years old. We believe strongly in protecting children on Twitter.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. GARY PETERS TO JACK DORSEY

Question 1. Twitter does not alert users if they have seen or interacted with content that content moderators have deemed harmful disinformation or extremist. Has Twitter examined or considered a policy of notifying users when content they have viewed has been removed? Why has such a policy not been implemented?

Answer. Our policies regarding terrorism, violent organizations, and hateful conduct are strictly enforced, as are all our policies. We take additional steps to safeguard the public conversation from manipulation. We proactively enforce policies and use technology to halt the spread of content propagated through manipulative tactics, such as automation or attempting to deliberately game trending topics. For example, we typically challenge 8 to 10 million accounts per week for these behaviors, requesting additional details, like e-mail addresses and phone numbers in order to authenticate the account.

Question 2. Twitter's community standards often draw the line at specific threats of violence for the removal of content, rather than conspiracy theories that may set the predicate for radicalization and future action. When it comes to conspiracy theories and misinformation, Twitter often chooses not to remove content, but rather to reduce the spread and to attach warnings. What testing or other analysis has Twitter done that shows your work to reduce the spread of disinformation and misinformation is effective?

Answer. As it relates specifically to violation of our Civic Integrity Policy related to the 2020 U.S. Election, we continue to apply labels, warnings, and additional restrictions to Tweets that included potentially misleading information about the election. During the period from October 27, 2020 to November 11, 2020, we labeled approximately 300,000 Tweets under our Civic Integrity Policy for content that was disputed and potentially misleading. These represent 0.2 percent of all U.S. election-related Tweets sent during this time period. Approximately 450 of those Tweets were also covered by a warning message and had engagement features limited, including Tweets could be Quote Tweeted but not Retweeted, replied to or liked. Approximately 74 percent of the people who viewed those Tweets saw them after we applied a label or warning message. We saw an estimated 29 percent decrease in Quote Tweets of these labeled Tweets due in part to a prompt that warned people prior to sharing. Since taking action on coordinated harmful activity, we have reduced impressions on this content by 50 percent.

Question 3. It is clear that the existence of conspiracy theories, disinformation campaigns, and misinformation has led to violence, even if not specifically planned on your platform. Recently, Twitter has taken action against the QAnon conspiracy for this reason. Why did QAnon reach that threshold now, and how will Twitter address other conspiracies? Is there a set number of violent incidents that must occur before Twitter considers a group unfit for the platform?

Answer. The *Twitter Rules* exist to ensure that people can participate in the public conversation freely and safely. In some cases, we identify groups, movements, or campaigns that are engaged in coordinated activity resulting in harm on and off of Twitter. We evaluate these groups, movements, or campaigns against an analytical framework, with specific on-Twitter consequences if we determine that they are harmful. Coordinated harmful activity is an actor-level framework, meaning we as-

sess groups, movements, and campaigns and then take enforcement action on any accounts which we identify as associated with those entities. In order to take action under this framework, we must find both evidence that individuals associated with a group, movement, or campaign are engaged in some form of coordination and that the results of that coordination cause harm to others. We respectfully direct you to *our policy* for greater detail on our enforcement approach.

Question 4. While I appreciate that Twitter continues to evolve and learn about threats of violence on the platform, would you agree that as groups evolve and change their tactics you will always be one step behind extremist groups that seek to use social media to recruit and plan violent acts? How do you address this problem?

Answer. The challenges we face as a society are complex, varied, and constantly evolving. These challenges are reflected and often magnified by technology. The push and pull factors influencing individuals vary widely and there is no one solution to prevent an individual turning to violence. This is a long-term problem requiring a long-term response, not just the removal of content.

While we strictly enforce our policies, removing all discussion of particular viewpoints, no matter how uncomfortable our customers may find them, does not eliminate the ideology underpinning them. Quite often, it moves these views into darker corners of the Internet where they cannot be challenged and held to account. As our peer companies improve in their efforts, this content continues to migrate to less-governed platforms and services. We are committed to learning and improving, but every part of the online ecosystem has a part to play.

Question 5. When prioritizing which content to evaluate, Twitter does not always consider the amount of time that content is on the platform but rather the spread. While this may make sense for disinformation, where the threat lies in misleading the population, when dealing with content to inspire violence, who sees the content can be more important than how many. As we have seen time and again, lone actors inspired to violence can cause significant harm. How do you address this issue?

Answer. The trend we are observing year-over-year is a steady decrease in terrorist organizations attempting to use our service. This is due to zero-tolerance policy enforcement that has allowed us to take swift action on ban evaders and other identified forms of behavior used by terrorist entities and their affiliates. In the majority of cases, we take action at the account creation stage—before the account even Tweets. We are reassured by the progress we have made, including recognition by independent experts. For example, Dublin City University Professor Maura Conway found in a detailed study that "ISIS's previously strong and vibrant Twitter community is now . . . virtually non-existent."

Question 6. What has Twitter done to work with the National Archives and Records Administration to help preserve the API and the content of Federal records on your platform in a way consistent with Federal records management processes?

Answer. As we have in prior years, we continue to engage with the National Archives and Records Administration on an ongoing basis to support their efforts during transition-related periods.

Question 7. How will Twitter address maintaining records in a way consistent with Federal records management processes from the "personal" social media accounts of the president, his advisors, and other members of the administration?

Answer. Twitter defers to the National Archives and Records Administration (NARA) and its Management Guide Series to provide Federal agencies with guidance on the management of records and other types of documentary materials accumulated by Federal agencies and officials. However, we work with NARA and other Federal entities to ensure they have the tools to preserve records, in cases where such preservation may be required or in the public interest.

Response to Written Questions Submitted by Hon. Kyrsten Sinema to Jack Dorsey

COVID–19 Misinformation. The United States remains in the midst of a global pandemic. More than 227,000 Americans have died of COVID–19, including nearly 6,000 in my home state of Arizona. COVID has impacted the health, employment, and education of Arizonans, from large cities to tribal lands like the Navajo Nation. And at the time of this hearing, the country is facing another significant surge in cases.

The persistent spread of COVID–19 misinformation on social media remains a significant concern to health officials. Digital platforms allow for inflammatory, dangerous, and inaccurate information—or outright lies—to spread rapidly. Sometimes

it seems that misinformation about the virus spreads as rapidly as the virus itself. This misinformation can endanger the lives and livelihoods of Arizonans.

Social distancing, hand washing, testing, contact tracing, and mask wearing should not be partisan issues, nor should they be the subject of online misinformation.

Question 1. What has Twitter done to limit the spread of dangerous misinformation related to COVID–19 and what more can it do?

Answer. The public conversation occurring on Twitter is critically important during this unprecedented public health emergency. With a critical mass of expert organizations, official government accounts, health professionals, and epidemiologists on our service, our goal is to elevate and amplify authoritative health information as far as possible. To address this global pandemic, on March 16, 2020, we announced *new enforcement guidance,* broadening our definition of harm to address, specifically, content related to COVID–19 that goes directly against guidance from authoritative sources of global and local public health information. We require individuals to remove violative Tweets in a variety of contexts with the goal of preventing offline harm. Additionally, we are currently engaged in an effort launched by the Office of the U.S. Chief Technology Officer under President Trump in which we are coordinating with our industry peers to provide timely, credible information about COVID–19 via our respective platforms. This working group also seeks to address misinformation by sharing emerging trends and best practices.

Spreading Accurate Information. Arizonans need accurate, scientifically based information to help get through this pandemic. Many Arizonans get their news from sources such as Twitter. As a result, your companies can play a role in helping people receive accurate information that is relevant to their communities and can aid them in their decisions that keep their families healthy and safe.

For example, earlier this month, the CDC issued a report illustrating that COVID–19 cases fell dramatically in Arizona after prevention and control measures were put into place. I shared this information on social media, and this is the type of information we should emphasize to help save lives.

Question 2. What more can Twitter do to better amplify accurate, scientifically-based health information to ensure that Arizonans understand how best to protect themselves from the pandemic?

Answer. In January 2020, we launched a dedicated search prompt feature to ensure that when individuals come to the service for information about COVID–19, they are met with credible, authoritative content at the top of their search experience. We have been continuously monitoring the conversation on the service to make sure keywords—including common misspellings—also generate the search prompt.

In the United States, people who search for key terms on Twitter are directed to the dedicated website on coronavirus and COVID–19 administered by the Centers for Disease Control and Prevention (CDC). In each country where we have launched the initiative, we have partnered with the national public health agency or the World Health Organization (@WHO) directly.

The proactive search prompt is in place with official local partnerships in the United States and nearly 70 markets around the world. We have also ensured the Events feature on Twitter contains credible information about COVID–19 and is available at the top of the home timeline for everyone in the U.S. and a number of other countries.

Scientific Evidence-based COVID Information. Our best sources of information related to the pandemic are doctors, researchers, and scientists. We should be relying on their expertise to help stop the spread of the virus and help our country recover from its devastating impacts.

Question 3. Who determines whether content on Twitter is scientifically supported and evidence based?

Answer. Twitter is enhancing our internal and external efforts to build partnerships, protect the public conversation, help people find authoritative health information, and contribute pro bono advertising support to ensure people are getting the right message, from the right source.

We have open lines of communication with relevant multinational stakeholders, including the CDC, the World Health Organization, numerous global government and public health organizations, and officials around the world, to ensure they can troubleshoot account issues, get their experts verified, and seek strategic counsel as they use the power of Twitter to mitigate harm.

COVID Scams Arizonans and Americans have been inundated with fraudulent offers and scams, using social media to spread inaccurate information and perpetrate

criminal scams. I've been using my own social media to help warn Arizonans about common scams related to economic assistance, false coronavirus "cures", and where they can report scams to Federal and state authorities.

Question 4. What has Twitter done to limit the spread of scams and report criminal activity and what more can be done to protect seniors, veterans, and others who have been targeted by fraudsters?

Answer. In September 2019, we updated our policies to clarify our prohibitions against scam tactics. We want Twitter to be a place where people can make human connections and find reliable information. For this reason, bad-faith actors may not use Twitter's services to deceive others into sending money or personal financial information via scam tactics, phishing, or otherwise fraudulent or deceptive methods.

Using scam tactics on Twitter to obtain money or private financial information is prohibited under this policy. Individuals are not allowed to create accounts, post Tweets, or send Direct Messages that solicit engagement in such fraudulent schemes. Our policies outline deceptive tactics that are prohibited. These include relationship/trust-building scams, money-flipping schemes, fraudulent discounts, and phishing scams.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. JACKY ROSEN TO JACK DORSEY

Question 1. Adversaries like Russia continue to amplify propaganda—on everything from the election to the coronavirus to anti-Semitic conspiracy theories—and they do it on your platform, weaponizing division and hate to destroy our democracy and our communities. The U.S. intelligence community warned us earlier this year that Russia is now actively inciting white supremacist violence, which the FBI and Department of Homeland Security say poses the most lethal threat to America. In recent years, we have seen white supremacy and anti-Semitism on the rise, much of it spreading online. What enables these bad actors to disseminate their hateful messaging to the American public are the algorithms on your platform, effectively rewarding efforts by foreign powers to exploit divisions in our country.

Question 1a. Are you seeing foreign manipulation or amplification of white supremacist and anti-Semitic content, and if so, how are your algorithms stopping this? Are your algorithms dynamic and nimble enough to combat even better and more personalized targeting that can be harder to identify?

Answer. Twitter strongly condemns antisemitism, and hateful conduct has absolutely no place on the service. Our Hateful Conduct Policy prohibits a wide range of behavior, including making references to violent events or types of violence where protected categories of people were the primary victims, or attempts to deny or diminish such events. Twitter also has a robust policy against glorification of violence and takes action against content that glorifies or praises historical acts of violence and genocide, including the Holocaust.

Combatting attempts to interfere in conversations on Twitter remains a top priority for the company, and we continue to invest heavily in our detection, disruption, and transparency efforts related to state-backed information operations. Twitter defines state-backed information operations as coordinated platform manipulation efforts that can be attributed with a high degree of confidence to state-affiliated actors. Our goal is to remove bad faith actors and to advance public understanding of these critical topics. Whenever we identify inauthentic activity on Twitter that meets our definition of an information operation, and which we are able to confidently attribute to actors associated with a government, we publicly share comprehensive data about this activity. To date, it is the only public archive of its kind. Using our archive, thousands of researchers have conducted their own investigations and shared their insights and independent analyses with the world.

Question 1b. Have you increased or modified your efforts to quell Russian disinformation in the wake of recently revealed efforts by Russia and Iran to weaponize stolen voter data to exploit divisions in our nation? How have you or will you adjust your algorithms to reduce the influence of such content—knowing that these countries' newly obtained data will allow for even better targeting, making their deception harder to identify?

Answer. Twitter believes that we have a responsibility to protect the integrity of conversations related to elections and other civic events from interference and manipulation. Therefore, we prohibit attempts to use our services to manipulate or disrupt civic processes, including through the distribution of false or misleading information about the procedures or circumstances around participation in a civic process. Combatting attempts to interfere in conversations on Twitter remains a top pri-

ority for the company, and we continue to invest heavily in our detection, disruption, and transparency efforts related to state-backed information operations.

In recent months we detected limited state backed information operations tied to the conversation regarding the 2020 U.S. election. On September 1, we suspended five Twitter accounts for platform manipulation that we can reliably attribute to Russian state actors based on information we received from the Federal Bureau of Investigations (FBI). We also permanently suspended two distinct networks of accounts that we can reliably attribute to state-linked entities in Russia on September 24.

On September 29, based on intelligence provided by the FBI, we removed 132 accounts that appeared to originate in Iran. The accounts were attempting to disrupt the public conversation during the presidential debate. Additionally, on October 8, we announced we had identified a network of primarily compromised accounts on Twitter operating from Iran. These accounts artificially amplified conversations on politically sensitive topics, including the Black Lives Matter movement, the death of George Floyd, and other issues of racial and social justice in the United States.

Question 1c. Are you consulting outside groups to validate moderator guidelines on hate speech, including what constitutes anti-Semitic content? Are you collecting data on hate speech content? If so, what are you doing with that data to combat hate speech on your platforms?

Answer. In addition to applying our iterative and research-driven approach to the expansion of the Twitter Rules, we've reviewed and incorporated public feedback to ensure we consider a wide range of perspectives. With each update to our policy prohibiting hateful conduct, we have sought to expand our understanding of cultural nuances and ensure we are able to enforce our rules consistently. We have benefited from *feedback* from various communities and cultures who use Twitter around the globe. We realize that we don't have all the answers, so in addition to public feedback, we work in partnership with our *Trust & Safety Council* as well as other organizations around the world with deep subject matter expertise in this area. Additionally, we convened a global working group of *third-party experts* to help us think about how we could appropriately address dehumanizing speech around the complex categories of race, ethnicity, and national origin. These experts helped us better understand the challenges we would face.

Question 2. Recently, there have been high profile cybersecurity breaches involving private companies, government agencies, and even school districts—including in my home state of Nevada. A few months ago, a hacker subjected Clark County School District—Nevada's largest school district and our country's fifth largest, serving more than 320,000 students—to a ransomware attack. In the tech industry, there was a notable breach of Twitter in July, when hackers were able to access an internal IT administrator tool used to manage accounts. Dozens of verified accounts with high follower counts—including those of President Obama, Bill Gates, and Jeff Bezos—were used to send out a tweet promoting a Bitcoin scam. What we learned from this breach is stunning . . . the perpetrators were inside the Twitter network in one form or another.

Question 2a. How often do your staff attend cybersecurity training? Do you hire outside cybersecurity firms to look at your systems, offering a fresh look and catching overlooked flaws?

Answer. Every new hire receives security and privacy training during onboarding. That training is augmented on an ongoing basis through mandatory and optional training. For example, each year employees must complete a mandatory compliance training that refreshes security and privacy requirements. In addition, employees in applicable teams receive training on secure coding and appropriate product launch practices. Mandatory training sessions also include phishing training and in July 2020, we convened company-wide meetings to provide guidance on the security vulnerabilities posed by phishing attacks.

Question 3. The COVID–19 pandemic has shined a light on our Nation's digital divide and on the technological inequalities facing millions of American students, including those in Nevada. Lack of access to broadband disproportionately affects low-income communities, rural populations, and tribal nations—all of which are present in my state. In addition to broadband access, many students still do not have regular access to a computer or other connected device, making online learning incredibly difficult, and sometimes impossible.

Twitter stepped up during the pandemic to help close the digital divide, including by offering many educational resources to help teachers and parents during the pandemic.

Question 3a. As classes continue to meet online, or in a hybrid model, what more can Twitter do to help students and teachers?

Answer. Educators can use Twitter to connect with their students, and you can use Twitter to teach about digital citizenship, freedom of expression, and respect. The conversation occurring on Twitter on #stuvoice and other hashtags are great ways to follow students and hear their voices. This is another helpful step in digital literacy and citizenship education because students see others speaking up, they will feel encouraged to raise their voices as well. One of the first lessons of digital literacy is understanding that everyone is a speaker, and each of us brings our own values and perspectives to a conversation. With support from the United Nations Educational, Scientific and Cultural Organization, we released a resource for educators on *Teaching and Learning with Twitter*. We continue to look for opportunities to partner with education stakeholders during this critical time.

Question 3b. How does Twitter plan to remain engaged in K–12 education after we get through the pandemic? In particular, what role can you play in closing not only the urban/rural divide, but also the racial divide in access to technologies and the Internet?

Answer. Bridging the digital divide is of the utmost importance. Twitter is supportive of efforts by all stakeholders to ensure that broadband access reaches every American. On May 15, 2020, Mr. Dorsey *Tweeted* he would personally donate $10 million dollars to #OaklandUndivided to ensure every child in Oakland, CA has access to a laptop and Internet in their homes.

Question 4. One of my top priorities in Congress is supporting the STEM workforce and breaking down barriers to entering and succeeding in STEM fields. This includes ensuring we have a diverse STEM workforce that includes people of color and women. In the past several years, tech companies have begun releasing diversity reports and promising to do better at hiring Black and Latino workers, including women. In overall employment, your companies are doing much better today in building a diverse workforce. However, in 2020, just 5.1 percent of Twitter's tech employees were Black, and only 4.2 percent were Latino.

I know that tech companies in Nevada understand that by increasing the number of women and people of color in tech careers, we diversify the qualified labor pool that the U.S. relies on for innovation. This will help us maintain our global competitiveness and expand our economy, and I hope your companies redouble your efforts to this effect.

Question 4a. Can you discuss the full set of 2020 data on women and the people of color who work at your companies, and would you please discuss what you are doing to increase these numbers in 2021?

Answer. Twitter is on a journey to be the world's most diverse and inclusive company—it is key to serving the public conversation. Our path starts with having a workforce that looks like the amazing people around the world who use our service everyday. We have made steady progress, but our work doesn't end *UntilWe AllBelong*. We provide detailed information in our *report on our Inclusion and Diversity* efforts.

Question 4b. What are you doing more broadly to support STEM education programs and initiatives for women and people of color, including young girls of color?

Answer. Twitter's NeighborNest is a community space dedicated to creating new opportunities through technology for San Francisco's most underserved, particularly homeless families who are over-represented by women and people of color. We do this by partnering with nonprofit organizations, Twitter employees, and community residents. Our programs focus on advancing Internet safety and education, equality and workforce development, and capacity building for NGOs. From career mentorship and STEAM education to Twitter training and hosted event space, we work to empower the community while cultivating empathy and equity. Through its University Recruiting team, Twitter participates in events and programs across the country that support aspiring technologists of color, such as the National Society of Black Engineers Annual Conference, as well as actively recruiting from Historically Black Colleges and Universities and higher education institutions serving LatinX students. In addition, Twitter is a corporate sponsor of the leading national organization dedicated to training young girls to enter careers in software engineering and computer science, Girls Who Code.

Finally, recognizing that great talent can come from non-traditional educational backgrounds, Twitter launched the Engineering Apprenticeship. Participants engage in an one-year program with full-time employment benefits. Apprentices are provided with hands-on experience while being paired with dedicated coaches and mentors to propel them to a successful career in engineering. Upon completion of the program, Engineering Apprentices graduate and are embedded in a Twitter engineering team.

Question 5. To continue being the most innovative country in the world, we need to maintain a workforce that can innovate. By 2026, the Department of Labor projects there will be 3.5 million computing-related jobs, yet our current education pipeline will only fill 19 percent of those openings. While other countries have prioritized STEM education as a national security issue, collaborating with non-profits and industry, the United States has mostly pursued an approach that does not meaningfully include such partnerships. The results of such a strategy are clear. A recent study found that less than half of K–12 students are getting any cyber related education, despite a growing demand for cyber professionals, both in national security fields and in the private sector.

Question 5a. What role can Twitter play in helping the United States boost its competitiveness in STEM fields, so that our economy can better compete with others around the globe?

Answer. Our efforts for corporate philanthropic investment can serve as a model. Under our Twitter for Good program, we have heavily invested in workforce development and youth service programs in San Francisco to prepare these populations for STEM fields. We work with the San Francisco United School District Department of Technology to assist them in their STEM curriculum. Additionally, we recently expanded this support globally through our partnership with JA Worldwide. We also fund specific programs like Techtonica, Hack the Hood, and dev/mission as a way to expand STEM opportunities to underrepresented communities.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. ROGER WICKER TO SUNDAR PICHAI

Question 1. Section 230 of the Communications Decency Act provides platforms like Twitter immunity from civil liability when dangerous—even illegal—third-party content is published on the platform. During the COVID–19 pandemic, search terms like "COVID cure" and "get hydroxychloroquine" have been popular in the United States as unsuspecting Americans grapple with the public health emergency. The search results often direct individuals to unregulated illegal "pharmacies" on the internet. In fact, last year, the National Association of Boards of Pharmacy found that of nearly 12,000 surveyed pharmacy websites, 90 percent were illegal. However, a recent survey found that 7 in 10 Americans erroneously believe that if an online pharmacy website appears high up in a search engine search, it is likely to be legitimate.

What specific action(s) has Google taken to ensure users are directed to safe websites to find reliable and correct information about the COVID–19 pandemic?

How has Google's search algorithm specifically taken illegal online pharmacies into consideration to ensure unsuspecting consumers are not directed to a website selling dangerous or illegal products?

Answer. Because the answers to these questions are related, we have grouped together our response to all subparts of Question No. 1.

We were committed to combating health misinformation before the coronavirus crisis, and our commitment has not wavered. In fact, a number of the policies and product features we are currently using as pat of our response to COVID–19 were already in place before the crisis began. For example, our ranking systems on Google Search and YouTube have been designed to elevate authoritative information in response to health-related searches for years, and we have significantly enhanced our efforts to combat COVID–19 misinformation. Since the outbreak of COVID–19, teams across Google have launched 200 new products, features, and initiatives, and we are contributing over $1 billion in resources to help our users, clients, and partners through this unprecedented time. These efforts include our Homepage "Do the Five" promotion, launch of a COVID–19-focused site (*https://www.google.com/intl/en_us/covid19/*), and amplifying authoritative voices through ad grants. There have been over 400 billion impressions on our information panels for coronavirus related videos and searches since the pandemic began, and since February, we've removed over 270 million coronavirus related ads across all Google advertising platforms, and 600,000 coronavirus videos, globally. We have invested heavily to make sure that we surface authoritative content in our search results, which significantly reduces the spread of misinformation, and we will continue to do so after this unprecedented public health crisis. And our work to update our YouTube recommendation systems to decrease the spread of misinformation, including (but not limited to) health-related misinformation, was announced in January 2019. For more information, please see *https://blog.youtube/news-and-events/continuing-our-work-to-improve.*

Our advertising policies prohibit unauthorized pharmacies and we restrict the promotion of online pharmacies (*https://support/adspolicy/answer/176031*). We do

not allow advertisers to offer drugs without a prescription or target a location where the advertiser is not licensed to sell. We also require online pharmacies to be accredited by either the National Association of Boards of Pharmacy (NABP) or the LegitScript Healthcare Merchant Certification and Monitoring Program and to apply to advertise with Google before they can promote their services. Users on YouTube also are not allowed to include links to an online pharmacy that does not require prescriptions.

In addition, irrespective of any claim of legality, our advertising policies prevent a long list of content from being promoted. This includes a non-exhaustive list of prohibited pharmaceuticals and supplements, products that have been subject to any government or regulatory action or warning, and products with names that are confusingly similar to an unapproved pharmaceutical or supplement or controlled substance. Our policies prevent Lapps, ads or merchant products that promote or sell unapproved substances as well, irrespective of any claims of legality, such as apps that facilitate the sale of marijuana.

In 2018, we also began delisting online rogue pharmacies identified in Food and Drug Administration (FDA) warning letters submitted directly to us by the FDA. Google and the FDA worked upon building and honing procedures to efficiently process these letters. Upon receipt of these letters from the FDA, we review and take delisting action for active sites associated with the rogue pharmacies identified in the warning letters.

We are also founding members of the Center for Safe Internet Pharmacies (CSIP), a nonprofit organization of diverse Internet service providers and technology companies that together with a growing group of U.S. states, supportive nonprofits, and other organizations, offers online pharmacy verification powered by LegitScript, and works closely with concerned organizations, such as the Partnership for Drug Free Kids, to educate consumers and others on vital addiction support and other related resources. For more information, please see *https://safemedsonline.org/resources/opioid-addiction-resources/*.

We have and will continue to identify new or changing illicit drugs or pharmaceutical products and to prohibit them from being surfaced on our products, and to address attempts to evade or circumvent our efforts in this area. We will continue to reassess our policies and procedures to increase the protection of our users and the public, and work to improve our policies.

Question 2. Mr. Pichai, you noted in the hearing that without the "otherwise objectionable" language of Section 230, the suppression of teenagers eating tide pods, cyber-bullying, and other dangerous trends would have been impossible.

Couldn't the language of Section 230 be amended to specifically address these concerns, including by the language of "promoting self harm" or unlawful" without needing the "otherwise objectionable" language that provides online platforms a blank check to take down any third-party speech with which they disagree?

What other language would be necessary to address truly harmful material online without needing to rely on the vague term "otherwise objectionable"?

Answer. Because the answers to these questions are related, we have grouped together our response to all subparts of Question No. 2.

Threats to our platforms and our users are constantly evolving. We certainly agree that we need to be able to limit content that "promot[es] self harm," is "unlawful," or is otherwise truly harmful material. But we have concerns about unintended consequences in removing "otherwise objectionable" material, as the nature of the harmful content we see is always changing. If we were to have specific exceptions, we would lose the ability to act in real time on troubling and dangerous content that we are seeing for the first time. Striking "otherwise objectionable" also could put removals of spam, malware, fraud, scams, misinformation, manipulated media, and hate speech at risk. Our ability to remove such content is particularly important now, when there has been a flood of daily malware, phishing e-mails, and spam messages related to COVID–19.

We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 3. Why wouldn't a platform be able to rely on terms of service to address categories of potentially harmful content outside of the explicit categories in Section 230(c)(2)? Why should platforms get the additional protections of Section 230 for removal of yet undefined categories of speech?

Answer. Section 230 is what permits us to curate content to protect users—and changes could jeopardize removals of terrorist content, spam, malware, scams, misinformation, manipulated media, and hate speech. Given the ever-evolving threats to our platforms and users, and that the nature of the content we see is always changing, it would be ineffective and impractical to attempt to address every pos-

sible harm in advance in our terms of service, and we could lose the ability to act in real time on troubling and harmful content that we are seeing for the first time. It is important that we and other platforms do not have to second guess our ability to act quickly to remove violative content. We are strong proponents of free speech, but have always had rules of the road and are never going to be "neutral" about issues like child abuse, terrorism, and harassment.

Google also remains committed to transparency in our business practices, including our content moderation efforts. In fact, we were the first to launch a Transparency Report (*https://transparencyreport.google.com/*) and have continued to expand and enhance our transparency efforts across numerous products and services over time. We do recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Does Section 230s "otherwise objectionable" catchall offer immunity for content moderation decisions motivated by political bias?

If the "otherwise objectionable" catchall does *not* offer such immunity, what limiting principle supports the conclusion that the catchall does not cover politically-biased moderation?

Answer. Because the answers to these questions are related, we have grouped together our response to these subparts of Question No. 3.

Our products are built for everyone, and we design them with extraordinary care to be a trustworthy source of information without regard to politics or political viewpoint. Our users overwhelmingly trust us to deliver the most helpful and reliable information available on the web. Distorting results or moderating content for political purposes would be antithetical to our mission and contrary to our business interests—it's simply not how we operate.

Consistent with our mission, Google enforces its content moderation policies consistently and impartially, without regard to political viewpoint. Section 230 has enabled us to respond quickly to ever-evolving threats to our platforms and users. For example, when the Christchurch videos happened, we saw a highly distressing type of content on our platforms—something that the "otherwise objectionable" standard allowed us to quickly address. It was important that we and other platforms did not have to second guess our ability to act quickly to remove that content. We also have robust policies and procedures in place to prevent content moderation decisions motivated by improper bias.

We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

If the "otherwise objectionable" catchall *does* offer such immunity now, how would you rewrite Section 230 to deny immunity for politically-biased content moderation while retaining it for moderation of content that is harmful to children?

Answer. Section 230 is one of the foundational laws that has enabled America's technology leadership and success in the Internet sector—allowing freedom of expression to flourish online. Google facilitates the speech of a wide range of people and organizations from across the political spectrum, giving them a voice and new ways to reach their audiences. We have always stood for protecting free expression online, and have enforced our content moderation policies consistently and impartially, and we will continue to do so.

In addition, millions of small and large platforms and websites across the Internet rely on Section 230 to keep users safe by addressing harmful content and to promote free expression. Section 230 is what permits us to curate content to protect users—and changes to Section 230 could jeopardize removals of terrorist content, spam, malware, scams, misinformation, manipulated media, hate speech, and content harmful to children. We are committed to working with Congress to see if there is a more flexible approach that would give overall guidance to platforms to receive complaints, implement appropriate processes, and report out—without overprescribing the precise manner and timelines by which they do so, or causing any unintended consequences.

As to content that is harmful to children, we are committed to protecting children on our platform. We have invested heavily in technologies and efforts to protect children like our Content Safety API and CSAI Match tools (*https://www.youtube.com/csai-match/*). And in 2019 alone, we filed more than 449,000 reports to the National Center for Missing & Exploited Children (NCMEC) Cyber Tipline. We are also a leading member of the Technology Coalition, where child safety experts across the industry build capacity and help companies working to increase their capacity to detect Child Sexual Abuse Material (CSAM) (*https://www.technologycoalition.org/*). In June, the Tech Coalition announced a multi-million dollar Research and Innovation Fund and Project Protect—a cross-industry initiative to combat CSAM through investment, research, and information sharing. For more information, please see

https://www.technologycoalition.org/2020/05/28/a-plan-to-combat-online-child-sexual-abuse.

We're committed to ensuring that our products are safe for children and families online, innovating and investing in measures to combat CSAM, and continuing to work with you to improve the ability to proactively detect, remove, and report this disturbing content. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 4. Are your terms of service easy to understand and transparent about what is and is not permitted on your platform?

Answer. The policies that govern use of our products and services work best when users are aware of the rules and understand how we enforce them. That is why we work to make this information clear and easily available to all. We are always working to provide greater transparency around our products and our business practices, including by making our Google terms of service (*https://policies.google.com/terms*) publicly available and plainly worded.

Our terms of service reflect the way our business works, the laws that apply to our company, and certain things we believe to be true. Among other things, we use examples from how users interact with and use our services to make our terms of service easy to understand. Google also has developed comprehensive help centers, websites outlining our policies, and blog posts that detail the specific provisions of our policies, as well as updates to these policies. In fact, Google was the first to launch a Transparency Report (*https://transparencyreport.google.com/*), we have expanded and enhanced our transparency efforts across numerous products and services over time, and we will continue to do so.

What notice and appeals process do you provide users when removing or labeling third-party speech?

What redress might a user have for improper content moderation beyond your internal appeals process?

Answer. Because the answers to these questions are related, we have grouped together our response to these subparts of Question No. 4.

Our mission at Google is to organize the world's information and make it universally accessible and useful. Core to this mission is a focus on the relevance and quality of the information we present to users. While the breadth of information available online makes it impossible to give each piece of content an equal amount of attention, human review, and deliberation, we certainly enforce our policies in an impartial manner without regard to politics or political viewpoint.

We want to make it easy for good-faith actors to understand and abide by our rules, while making it difficult for bad actors to flout them. If users believe their Google Accounts have been suspended or terminated in error, we seek to provide the opportunity for users to appeal decisions and provide clarification when reasonably possible. To help ensure consistent and fair application of our rules and policies, such decisions are then evaluated by a different member of our Trust and Safety team. Users can learn more about their rights relating to our terms of service at *https://policies.google.com/terms.*

In addition to our general terms of service, we also publish service-specific policies detailing the appeals process, including information on Search Reconsideration Requests (*https://support.google.com/webmasters/answer/35843*), Ads disapprovals and suspensions (*https://support.google.com/adspolicy/topic/1308266*), publisher Policy Center violations (*https://support.google.com/adsense/answer/7003627*), and YouTube Community Guidelines violations (*https://support.google.com/youtube/answer/185111*). We are transparent about our decisions and discuss them further in places like our How Search Works page (*https://www.google.com/search/researchers/mission/open-web/*), Google Transparency Report (*https://transparencyreport.google.com/*), and YouTube Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/removals*).

In what way do your terms of service ensure against politically-biased content moderation and in what way do your terms of service limit your ability to moderate content on your platform?

How would you rewrite your terms of service to protect against politically-biased content Moderation?

Answer. Because the answers to these questions are related, we have grouped together our response to these subparts of Question No. 4.

Our products are built for everyone, and we design them with extraordinary care to be a trustworthy source of information without regard to political viewpoint. Our users overwhelmingly trust us to deliver the most helpful and reliable information available. Distorting results or moderating content for political purposes or based

on ideology would be antithetical to our mission and contrary to our business interests.

Google's publicly available terms of service (*https://policies.google.com/terms*) provide that we reserve the right to take down any content that we reasonably believe breaches our terms of service, violates applicable law, or could harm our users, third patsies, or Google—we enforce these terms and our other policies in a impartial and consistent manner without regard to politics or political viewpoint.

We also have safeguards in place to ensure that we enforce our policies in a way that is free from political bias. In addition to technical controls and machine learning detection systems, we have robust systems to ensure that employees' personal biases do not impact our products and that our policies are enforced in a politically neutral way. These include policies that prohibit employees from engaging in unethical behavior, including altering or falsifying Google's systems to achieve some personal goal or benefit. In addition, Google reviewers, including Search raters, go through regular training and training refreshes. These reviewers are regularly tested and graded for consistency with Google's policies. Our Trust and Safety team also conducts reviews for compliance in accordance with our own policies. Finally, we employ review teams across the globe to ensure we have a diverse set of reviewers who are reviewing publisher sites and apps. We are proud of our processes and are committed to ensuring we are fair and unbiased in enforcing our policies.

Do you think that removing content inconsistent with your terms of service and public representations is removal of content "in good faith"?

Answer. We design and build our products for everyone, and enforce our policies in a good faith, impartial way. We endeavor to remove content only when it is inconsistent with our policies, with no regard to ideology or political viewpoint. As explained above, when we take action or make decisions to enforce our policies, we make it clear to users that we have taken action on their content and provide them the opportunity to appeal that decision and provide any clarification.

Question 5. Please provide a list of all instances in which a prominent individual promoting liberal or left-wing views has been censored, demonetized, or flagged with extra context by your company.

Please provide a list of all instances in which a prominent individual promoting conservative or right-wing views has been censored, demonetized, or flagged with extra context by your company.

How many posts by government officials from Iran or China have been censored or flagged by your company?

How many posts critical of the Iranian or Communist Chinese government have been flagged or taken down?

Answer. Because the answers to these questions are related, we have grouped together our response to all subparts of Question No. 5.

Our products are built for everyone, and we design them with extraordinary care to be a trustworthy source of information without regard to politics or political viewpoint. Billions of people use our products to find information, and we help our users, of every background and belief, find the high-quality information they need to better understand the topics they care about. That is why we are committed to transparency in our business practices, including our content moderation efforts. Our Terms of Service are public and can be found at: *https://policies.google.com/terms.* And while they provide that we reserve the right to take down any content that we reasonably believe breaches our terms of service, violates applicable law, or could harm our users, third parties, or Google—we enforce those terms and our other policies in an impartial manner, without regard to politics or political viewpoint and, wherever possible, we make it clear to creators that we have taken action on their content and provide them the opportunity to appeal that decision and provide clarification.

To give some sense of the scale of our efforts, in 2019 we blocked and removed 2.7 billion bad ads—that's more than 5,000 bad ads per minute. We also suspended nearly 1 million advertiser accounts for policy violations. On the publisher side, we terminated over 1.2 million accounts and removed ads from over 21 million web pages that are pat of our publisher network for violating our policies. Our efforts to protect users from bad ads are described in more detail at *https://blog.google/products/ads/stopping-bad-ads-to-protect-users,* and our efforts to enforce our YouTube Community Guidelines are described in more detail at *https://transparencyreport.google.com/youtube-policy/removals.*

For example, in the lead up to the 2020 election, we refused to publish ads that violated our policies from both the Biden and Trump campaigns. For additional information, please see the Political Advertising section of our Transparency Report (*https://transparencyreport.google.com/political-ads/region/US*). We've also re-

ceived complaints about Search rankings from across the political spectrum, from the World Socialist Web Site to Breitbart.

At the same time, creators on both the left and right have had unprecedented success on our platforms. For example, YouTube has been a tremendous tool for conservative outlets such as PragerU seeking to expand the reach of their message. That channel has over 2 million subscribers and over 1 billion video views. Similarly, candidates from both parties heavily utilized YouTube as a way to reach voters this election cycle, including running ads and publishing content from a wide variety of political commentators.

As to government requests to remove content, including requests from the Chinese and Iranian governments, we provide information about these requests in our transparency report: *https://transparencyreport.google.com/government-removals/by-country*. Additionally, we've provided ongoing updates about our efforts to combat coordinated influence operations, including operations linked to China and Iran, and in May 2020 we announced the launch of a new bulletin to provide more frequent, quarterly updates about our findings: *https://blog.google/threat-analysis-group/updates-about-government-backed-hacking-and-disinformation/*.

We are proud that we have a wide variety of views expressed on our platforms and are committed to ensuring we continue to enforce our policies in a fair and impartial manner.

Question 6. Mr. Pichai, you responded to Senator Gardner that companies should not be given sweeping immunity for "Company-Created Content".

As it stands, Section 230 has been interpreted not to grant immunity if a publishing platform "ratifies" illicit activity. Do you agree? How do you think "ratification" should be defined?

Do you agree that a platform should not be covered by Section 230 if it adds its own speech to third-party content?

When a platform adds its own speech, does it become an information content provider under Section 230(f)(3)?

Answer. Because the answers to these questions are related, we have grouped together our response to all subparts of Question No. 6.

Google is a technology company that facilitates the speech of a wide range of people and organizations from across the political spectrum, giving them a voice and new ways to reach their audiences. We provide a platform for creators, advertisers, academics, politicians, scientists, religious groups, and myriad others. Section 230 was passed recognizing the unique nature of platforms that host user-generated content and allows us to protect our users in the face of ever-evolving content and threats. Section 230 safeguards open access to information and free expression online. Instead of overblocking speech, the law supports platforms' ability to responsibly manage content.

In some cases, we may also act as an information content provider—for instance, when we publish a logos on blog.google. In such instances, we would be treated as the information content provider with respect to that specific content, but that does not and should not affect how our services overall are treated under the law.

We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 7. Should algorithms that promote or demote particular viewpoints be protected by Section 230? Why or why not?

Answer. Our products are built for everyone, and we design them with extraordinary care to be a trustworthy source of information without regard to political viewpoint. Our users overwhelmingly trust us to deliver the most helpful and reliable information available. Distorting results or moderating content for political purposes would be antithetical to our mission and contrary to our business interests.

Our services organize, rank, and recommend content in a wide variety of ways to help meet people's needs and interests. Indeed, this is the essence of most online services today.

Curtailing Section 230 based on the use of ranking algorithms would thus undermine the many benefits of the statute today.

We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Do you think the use of an individual company's algorithms to amplify the spread of illicit or harmful materials like online child sexual exploitation should be protected by Section 230?

Answer. We've always been proponents of free speech, but have always had rules of the road and are never going to be "neutral" about issues like child abuse, terrorism, and harassment. We are very focused on the protection of children on our platforms. Section 230 is what permits us to curate content to protect users, and

changes to Section 230 could jeopardize removals of terrorist content, spam, malware, scams, misinformation, manipulated media, and hate speech.

At Google, we have invested heavily in technologies and efforts to protect children on our platform, like our Content Safety API and CSAI Match tools (*https://www.youtube.com/csai-match/*). We already proactively look for and report illegal child sexual abuse to NCMEC—filing more than 449,000 reports to the NCMEC Cyber Tipline in 2019 alone. We are also a leading member of the Technology Coalition, where child safety experts across the industry build capacity and help companies working to increase their capacity to detect Child Sexual Abuse Material (CSAM) (*https://www.technologycoalition.org/*). In June, the Tech Coalition announced a multi-million dollar Research and Innovation Fund and Project Protect—a cross-industry initiative to combat CSAM through investment, research, and information sharing. For more information, please see *https://www.technologycoalition.org/2020/05/28/a-plan-to-combat-online-child-sexual-abuse.*

This is a very important issue and we're committed to continue working with Congress on it.

Question 8. Should platforms that knowingly facilitate or distribute Federal criminal activity or content be immune from civil liability? Why?/Why not?

If your company has actual knowledge of content on your platform that incites violence, and your company fails to remove that content, should Federal law immunize your company from any claims that might otherwise be asserted against your company by victims of such violence? Are there limitations or exceptions to such immunity that you could propose for consideration by the Committee?

Should platforms that are willfully blind to Federal criminal activity or content on their platforms be immune from civil liability? Why? Why not?

Answer. Because the answers to these questions are related, we have grouped together our response to all subparts of Question No. 8.

Section 230 helps Internet companies address harmful content, including user comments, and while we've always been proponents of free speech, we've also always had rules of the road and are never going to be "neutral" about harmful content. Millions of small and large platforms and websites across the Internet rely on Section 230 to both keep users safe and promote free expression. Google also has worked closely with law enforcement and organizations such as NCMEC, Thorn, and Polaris for years. Under existing law, Section 230s protections for online platforms already exempt all Federal criminal law. We have concerns that changes to Section 230 would negatively impact our ability to remove harmful content of all types and would make our services less useful and safe. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 9. Ranking has been described as Google's "Holy Grail." During the 2020 election (January 1, 2020 to date), how was ranking used for searches related to candidates and election-related issues to control the content seen by its users?

During the 2020 election (January 1, 2020 to date), did Google lower the search visibility, suppress, or de-rank in any way any search results for any candidates, or election-related issues? If so, how and when was this done, and why?

During the 2020 election (January 1, 2020 to date), did Google lower the search visibility, suppress, or de-rank in any way search results for any news outlets, including Breitbart News, the Daily Caller, or the Federalist? If so, how and when was this done, and why?

Answer. Because the answers to these questions are related, we have grouped together our response to all subparts of Question No. 9.

Our business model is dependent on being a useful and trustworthy source of information for everyone and we design Search and other products with extraordinary care to serve our mission without regard to politics or political viewpoint. Our Search algorithm ranks pages to provide the most useful and relevant information by matching search terms against available web pages and looking at factors like the number of times the words appear and freshness of the page. Political viewpoint is not a relevant factor to our Search results or ranking and the 2020 election cycle was no exception. We also seek to ensure that our Search results are providing the most authoritative and responsive results by using external quality raters from across the vast majority of U.S. states. In addition, we have robust systems in place to ensure that our policies are enforced in a politically impartial way across all of our products and services, including Search.

For the 2020 election, we worked with The Associated Press (AP)—a trusted source of information on election results—to provide authoritative elections results on Google. When people come to Search looking for information on election results,

we provide a dedicated feature with data provided by the AP to provide real-time information on Google for Federal and state level races.

Finally, we do not manually intervene in any particular Search result. We remove content from our organic Search results only in very limited circumstances, such as a court order, valid DMCA takedown notice, or violation of our webspam policies. Search ranking considers many factors, but political ideology is not factored into the process. When we make changes to Search, they apply broadly, after extensive testing and controls, rigorous evaluation, and use of detailed metrics. Our publicly available Search Quality Rater Guidelines (*https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf*) provide transparency for ratings, users, and webmasters about how Search works. And they make it clear that ratings should never be based on personal opinions, preferences, religious beliefs, or political views.

Question 10. In our increasingly digital world, consumers are demanding more access to video digital data products and services. The need for these mediums has only increased due to the pandemic with more Americans relying on video players as pat of their online school and work. Competition in this area is key to ensure that the best technically available products are available to consumers at competitive prices, however, products such as ISO's MPEG High Efficiency Video Code have had its access restricted by certain browsers, including Chrome. Is Google actively blocking competing video players in favor of Chrome-specific products? What steps is Google taking to ensure that products and services that directly compete with its own product offerings are easily available and accessible to all consumers when using Chrome?

Answer. Google Chrome is focused on creating the best possible experience for web browsing. To this end, Google has focused on including technologies in Chrome that facilitate the development and delivery of media content on internet-connected devices and that improve the media experience for all users across the browser ecosystem, including OEMs and content providers. These include media technologies developed by third parties outside Google.

The success of Chrome depends on providing users with a fast, secure, and performant browsing experience for websites and services across the web whether or not they are from Google or other providers; understanding that is why a number of users choose Chrome from among the many browser options available. We continually evaluate the need to include technologies in Chrome based on feedback from our entire ecosystem of users, OEMs, and content providers, and we add technologies where we identify shared needs across the ecosystem.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. JOHN THUNE TO SUNDAR PICHAI

Question 1. We have a public policy challenge to connect millions of Americans in rural America to broadband. I know you share in our commitment to connect every American household with broadband not only because it's the right thing to do but because it will add millions of new users to your platforms, which of course, means increase profits. What role should Congress and your companies play in ensuring that we meet all the broadband demands in rural America?

Answer. Broadband technology, specifically high-speed Internet and the devices and tools it enables, is essential to unlock key services, especially during this time of crisis. For example, access to broadband makes telemedicine possible, so patients can easily confer online with their doctors, and this access also enables the tools needed for distance learning and teleworking. But we have seen that many Americans, in both rural and urban areas, are left out of this picture because they do not have access to affordable broadband technology.

Google's commitment to playing its pat to bridge the digital divide means we have invested in expanding access to broadband for those who need it most. For example, we are proud to have pioneered a program called Rolling Study Halls (*https://edu.google.com/why-google/our-commitment/rolling-study-halls*) to equip buses with WiFi to provide Internet access for students in rural communities. In addition, Google Fiber's Community Connections (*https://fiber.google.com/community/#!/*) program offers organizations such as libraries, community centers, and nonprofits free Internet access.

To keep up with the rising demand for bandwidth, the FCC has worked with industry leaders like Google to create the CBRS rules (*https://www.cbrsalliance.org/resource/what-is-cbrs/*) for shared spectrum as a new model for adding capacity at a low cost. By aligning on industry standards, Google is helping the CBRS ecosystem bring better wireless Internet to more people in more places. As the founda-

tion for Google's suite of products and services for CBRS (*https://www.google.com/get/spectrumdatabase/#cbrs*), Google's Spectrum Access System (SAS) controls fundamental access to CBRS. Google's SAS is purpose-built to support dense networks across operators and to scale on-demand—from a small in-building network to the largest nationwide deployment. For more information on how Google is bringing affordable Internet and choice to consumers, please see *https://www.google.com/get/spectrumdatabase/sas/*.

Looking forward, we also believe our mapping and network planning tools could be useful for improving the reach and coverage of next generation broadband networks. While we don't have specific datasets that can solve broadband mapping issues, we're happy to come together with the government and industry to try to solve this problem. We also recognize that any effort would have to be sensitive to privacy best practices and consumer expectations, and we are committed to helping find a solution.

While private sector involvement is very important to making broadband more accessible, government policy is fundamental. Now more than ever, there is a need for coordinated action by Federal and state governments to support investment in world-class digital infrastructure to help close the digital divide and deliver ubiquitous, affordable Internet access to rural, urban, tribal, and other underserved communities. Legislation that increases investment in broadband infrastructure, broadband adoption and digital literacy programs, and the availability of commercial spectrum will pay enormous dividends for those in rural and urban areas today who are on the wrong side of the digital divide. We would be pleased to discuss further existing proposals before Congress, as well as any other ways we can assist with this effort.

Question 2. The PACT Act would require your platforms to take down content that a court has ruled to be illegal. Do you support a court order-based takedown rule?

Answer. With respect to the PACT Act, we are supportive of steps that promote transparency of content policies, provide notices of removals, and encourage responsible practices, while still ensuring a vibrant and responsible online environment. In fact, we were the first to launch a Transparency Report (*https://transparencyreport.google.com/*) and have continued to expand and enhance our transparency efforts across numerous products and services over time. We do think it is important that any legislative approach provides flexibility—allowing platforms to build content moderation systems fit for the type of service they offer and the processes they use. We have concerns that certain changes would negatively impact our ability to remove harmful content of all types and would make our services less useful and safe. We view Section 230 as a foundational American law that has enabled the U.S. to lead the Internet globally, supporting millions of jobs and billions of dollars of economic activity—so we want to be very cautious and thoughtful about potential changes. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 3. Google's Trust and Safety team has the challenging task of filtering out things like violent extremism from your products, but Google largely sets its own standards for what should be blocked under the rationale of protecting users, and many are skeptical about where the line is drawn once you get past incitements to violence. For example, an under-reported but nevertheless newsworthy story might be mislabeled a conspiracy theory by those seeking to suppress it.

Would you agree that no one's "safety" is protected—and the "trust" of many users is actually jeopardized—when Google uses such policies to restrict the exposure of conservative content online? Can you assure the committee that such suppression is not occurring?

Answer. Yes. We are pleased to assure you that we apply our policies objectively and consistently in an impartial manner without regard to politics or political viewpoint. We build products that are for everyone, and we design and enforce our policies in a fair and impartial manner.

For each product and service we offer, we tailor our policies to distinguish between providing access to a diversity of voices and limiting harmful content and behaviors—whether those are our policies against hate speech or material that is excessively violent, unlawful, deceptive, or obscene (*e.g.,* Advertising Policies, *https://support.google.com/adspolicy/answer/6015406;* Publisher Policies, *https://support.google.com/adsense/answer/9335564;* and YouTube Community Guidelines, *https://support.google.com/youtube/answer/9288567*). We also have safeguards in place to ensure that we enforce these policies in a consistent way without bias as to the ideological viewpoint of the content.

Question 4. Section 230 was initially adopted to provide a "shield" for young tech stat-ups against the risk of overwhelming legal liability. Since then, however, some tech platforms like yours have grown larger than anyone could have imagined. Often a defense we hear from Section 230 proponents is that reform would hut current and future stat-ups. The PACT Act requires greater reporting from tech platforms on moderation decisions, largely exempts small business. However, your companies are no longer stat-ups, but rather some of the most powerful and profitable companies in the world.

Do tech giants need "shields" codified by the U.S. government? Have you outgrown your need for Section 230 protections?

Answer. It is no accident that the greatest Internet companies in the world were created in the United States. Section 230 has worked remarkably well, and we believe a cautious and thoughtful approach to potential changes is appropriate. Our platforms empower a wide range of people and organizations from across the political spectrum, giving them a voice and new ways to reach their audiences. Section 230 has enabled that, and millions of small and large platforms and websites across the Internet rely on Section 230 to both keep users safe and promote free expression. Changes to 230 would disproportionately impact up-and-coming platforms without the resources to try and police every comment or defend every litigation, which could deter the next Google, Twitter, or Facebook, as the liability for third-party content would be too great. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 5. Justice Thomas recently observed that "[p]aring back the sweeping immunity courts have read into § 230 would not necessarily render defendants liable for online misconduct. It simply would give plaintiffs a chance to raise their claims in the first place. Plaintiffs still must prove the merits of their cases, and some claims will undoubtedly fail." Do you agree with him? Why shouldn't lawsuits alleging that a tech platform has violated a law by exercising editorial discretion be evaluated on the merits rather than being dismissed because a defendant invokes Section 230 as a broad shield from liability?

Answer. By putting potentially every decision around content moderation up to judicial review, we do have concerns that this type of change would negatively impact our ability to remove harmful content of all types and would make our services less useful and safe. Millions of small and large platforms and websites across the Internet rely on Section 230 to both keep users safe and promote free expression. We believe that Section 230 strikes the appropriate balance that facilitates making more content and diverse points of view available than ever before in history, all while ensuring Internet companies can keep their platforms safe and secure for our users. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 6. What does "good faith" in Section 230 mean? Is there any action you could take that could not be justified as done in "good faith"? Do you agree bad faith content moderation is not covered by Section 230? If content is removed pre-textually, or if terms and conditions are applied inconsistently depending on the viewpoint expressed in the content, is that removing content in good faith?

Answer. Google agrees that exercising good faith is important, and we also believe we already engage in good faith content moderation. Among the many policies and publicly available materials we make available, we have published terms of service (*https://policies.google.com/terms*) and endeavor to remove content only when removal is consistent with our policies. When we make decisions to enforce our policies, we make it clear to creators that we have taken action on their content and provide them the opportunity to appeal that decision and provide clarification.

Our products are built for everyone, and we design them with extraordinary care to be a trustworthy source of information without regard to politics or political viewpoint. Our users overwhelmingly trust us to deliver the most helpful and reliable information available on the web. Distorting results or moderating content for political purposes would be antithetical to our mission and contrary to our business interests—it's simply not how we operate.

We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 7. You noted in the hearing that without the "otherwise objectionable" language of Section 230, the suppression of teenagers eating tide pods, cyber-bullying, and other dangerous trends would have been impossible. Could the language of Section 230 be amended to specifically address these concerns, by including the language of "promoting self-harm" or "unlawful" without needing the "otherwise ob-

jectionable" language that provides online platforms a blank check to take down any third-party speech with which they disagree?

Question 8. What other language would be necessary to address truly harmful material online without needing to rely on the vague term "otherwise objectionable?"

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 7 and 8.

Threats to our platforms and our users are constantly evolving. We certainly agree that we need to be able to limit content that "promot[es] self harm," is "unlawful," or is otherwise truly harmful material. But we have concerns about unintended consequences in removing "otherwise objectionable" material, as the nature of the harmful content we see is always changing. If we were to have specific exceptions, we would lose the ability to act in real time on troubling and dangerous content that we are seeing for the first time. Striking "otherwise objectionable" also could put removals of spam, malware, fraud, scams, misinformation, manipulated media, and hate speech at risk. Our ability to remove such content is particularly important now, when there has been a flood of daily malware, phishing e-mails, and spam messages related to COVID–19.

We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 9. Why wouldn't a platform be able to rely on the terms of service to address categories of potentially harmful content outside of the explicit categories in Section 230(c)(2)? Why should platforms get the additional protections of Section 230 for removal of yet undefined categories of speech?

Answer. Section 230 is what permits us to curate content to protect users—and changes could jeopardize removals of terrorist content, spam, malware, scams, misinformation, manipulated media, and hate speech. Given the ever-evolving threats to our platforms and users, and that the nature of the content we see is always changing, it would be ineffective and impractical to attempt to address every possible harm in advance in our terms of service, and we could lose the ability to act in real time on troubling and harmful content that we are seeing for the first time. It is important that we and other platforms do not have to second guess our ability to act quickly to remove violative content. We are strong proponents of free speech, but have always had rules of the road and are never going to be "neutral" about issues like child abuse, terrorism, and harassment.

Google remains committed to transparency in our business practices, including our content moderation efforts. In fact, we were the first to launch a Transparency Report (*https://transparencyreport.google.com/*) and have continued to expand and enhance our transparency efforts across numerous products and services over time. We recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 10. Does Section 230s "otherwise objectionable" catchall offer immunity for content moderation decisions motivated by political bias?

If the "otherwise objectionable" catchall does not offer such immunity, what limiting principle supports the conclusion that the catchall does not cover politically-biased moderation?

Answer. Because the answers to these questions are related, we have grouped together our response to these subparts of Question No. 10.

Our products are built for everyone, and we design them with extraordinary care to be a impartially source of information without regard to politics or political viewpoint. Our users overwhelmingly trust us to deliver the most helpful and reliable information available on the web. Distorting results or moderating content for political purposes would be antithetical to our mission and contrary to our business interests—it's simply not how we operate.

Consistent with our mission, Google enforces its content moderation policies consistently and impartially, without regard to political viewpoint. Section 230 has enabled us to respond quickly to ever-evolving threats to our platforms and users. For example, when the Christchurch videos happened, we saw a highly distressing type of content on our platforms—something that the "otherwise objectionable" standard allowed us to quickly address. It was important that we and other platforms did not have to second guess our ability to act quickly to remove that content. We also have robust policies and procedures in place to prevent content moderation decisions motivated by improper bias.

We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

If the "otherwise objectionable" catchall does offer such immunity now, how would you rewrite Section 230 to deny immunity for politically-biased content moderation while retaining it for moderation of content that is harmful to children?

Answer. Section 230 is one of the foundational laws that has enabled America's technology leadership and success in the Internet sector—allowing freedom of expression to flourish online. Google facilitates the speech of a wide range of people and organizations from across the political spectrum, giving them a voice and new ways to reach their audiences. We have always stood for protecting free expression online, and have enforced our content moderation policies consistently and impartially, and we will continue to do so.

In addition, millions of small and large platforms and websites across the Internet rely on Section 230 to keep users safe by addressing harmful content and to promote free expression. Section 230 is what permits us to curate content to protect users—and changes to Section 230 could jeopardize removals of terrorist content, spam, malware, scams, misinformation, manipulated media, hate speech, and content harmful to children. We are committed to working with Congress to see if there is a more flexible approach that would give overall guidance to platforms to receive complaints, implement appropriate processes, and report out—without overprescribing the precise manner and timelines by which they do so, or causing any unintended consequences.

As to content that is harmful to children, we are committed to protecting children on our platform. We have invested heavily in technologies and efforts to protect children like our Content Safety API and CSAI Match tools (*https://www.youtube.com/csai-match/*). And in 2019 alone, we filed more than 449,000 reports to the NCMEC Cyber Tipline. We are also a leading member of the Technology Coalition, where child safety experts across the industry build capacity and help companies working to increase their capacity to detect Child Sexual Abuse Material (CSAM), *https://www.technologycoalition.org/*. In June, the Tech Coalition announced a multi-million dollar Research and Innovation Fund and Project Protect–a cross-industry initiative to combat CSAM through investment, research, and information sharing. For more information, please see *https://www.technologycoalition.org/2020/05/28/a-plan-to-combat-online-child-sexual-abuse.*

We're committed to ensuring that our products are safe for children and families online, innovating and investing in measures to combat CSAM, and continuing to work with you to improve the ability to proactively detect, remove, and report this disturbing content. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 11. Are your terms of service easy to understand and transparent about what is and is not permitted on your platform?

Answer. The policies that govern use of our products and services work best when users are aware of the rules and understand how we enforce them. That is why we work to make this information clear and easily available to all. We are always working to provide greater transparency around our products and our business practices, including by making our Google terms of service (*https://policies.google.com/terms*) publicly available and plainly worded.

Our terms of service reflect the way our business works, the laws that apply to our company, and certain things we believe to be true. Among other things, we use examples from how users interact with and use our services to make our terms of service easy to understand. Google has also developed comprehensive help centers, websites outlining our policies, and blog posts that detail the specific provisions of our policies, as well as updates to these policies. In fact, Google was the first to launch a Transparency Report (*https://transparencyreport.google.com/*), we have expanded and enhanced our transparency efforts across numerous products and services over time, and we will continue to do so.

Question 12. What notice and appeals process do you provide users when removing or labeling third-party speech?

Question 13. What redress might a user have for improper content moderation beyond your internal appeals process?

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 12 and 13.

Our mission at Google is to organize the world's information and make it universally accessible and useful. Core to this mission is a focus on the relevance and quality of the information we present to users. While the breadth of information available online makes it impossible to give each piece of content an equal amount of attention, human review, and deliberation, we certainly enforce our policies in an impartial manner without regard to politics or political viewpoint.

We want to make it easy for good-faith actors to understand and abide by our rules, while making it difficult for bad actors to flout them. If users believe their Google Accounts have been suspended or terminated in error, we seek to provide

the opportunity for users to appeal decisions and provide clarification when reasonably possible. To help ensure consistent and fair application of our rules and policies, such decisions are then evaluated by a different member of our Trust and Safety team. Users can learn more about their rights relating to our terms of service at *https://policies.google.com/terms*.

In addition to our general terms of service, we also publish service-specific policies detailing the appeals process, including information on Search Reconsideration Requests (*https://support.google.com/webmasters/answer/35843*), Ads disapprovals and suspensions (*https://support.google.com/adspolicy/topic/1308266*), publisher Policy Center violations (*https://support.google.com/adsense/answer/7003627*), and YouTube Community Guidelines violations (*https://support.google.com/youtube/answer/185111*). We are transparent about our decisions and discuss them further in places like our How Search Works page (*https://www.google.com/search/hawser/mission/open-web/*), Google Transparency Report (*https://transparencyreport.google.com/*), and YouTube Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/removals*).

Question 14. In what way do your terms of service ensure against politically-biased content moderation and in what way do your terms of service limit your ability to moderate content on your platform?

Question 15. How would you rewrite your terms of service to protect against politically-biased content moderation?

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 14 and 15.

Our products are built for everyone, and we design them with extraordinary care to be a impartially source of information without regard to political viewpoint. Our users overwhelmingly trust us to deliver the most helpful and reliable information available. Distorting results or moderating content for political purposes or based on ideology would be antithetical to our mission and contrary to our business interests.

Google's publicly available terms of service (*https://policies.google.com/terms*) provide that we reserve the right to take down any content that we reasonably believe breaches our terms of service, violates applicable law, or could harm our users, third patsies, or Google—we enforce these terms and our other policies in a impartial and consistent manner without regard to politics or political viewpoint.

We also have safeguards in place to ensure that we enforce our policies in a way that is free from political bias. In addition to technical controls and machine learning detection systems, we have robust systems to ensure that employees' personal biases do not impact our products and that our policies are enforced in a politically neutral way. These include policies that prohibit employees from engaging in unethical behavior, including altering or falsifying Google's systems to achieve some personal goal or benefit. In addition, Google reviewers, including Search raters, go through regular training and training refreshes. These reviewers are regularly tested and graded for consistency with Google's policies. Our Trust and Safety team also conducts reviews for compliance in accordance with our own policies. Finally, we employ review teams across the globe to ensure we have a diverse set of reviewers who are reviewing publisher sites and apps. We are proud of our processes and are committed to ensuring we are fair and unbiased in enforcing our policies. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 16. Do you think that removing content inconsistent with your terms of service and public representations is removal of content "in good faith"?

Answer. We design and build our products for everyone, and enforce our policies in a good faith, impartial way. We endeavor to remove content only when it is inconsistent with our policies, with no regard to ideology or political viewpoint. As explained above, when we take action or make decisions to enforce our policies, we make it clear to users that we have taken action on their content and provide them the opportunity to appeal that decision and provide any clarification.

Question 17. As it stands, Section 230 has been interpreted not to grant immunity if a publishing platform "ratifies" illicit activity. Do you agree? How do you think "ratification" should be defined?

Question 18. Do you agree that a platform should not be covered by Section 230 if it adds its own speech to third-party content?

Question 19. When a platform adds its own speech, does it become an information content provider under Section 230(f)(3)?

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 17–19.

Google is a technology company that facilitates the speech of a wide range of people and organizations from across the political spectrum, giving them a voice and new ways to reach their audiences. We provide a platform for creators, advertisers, academics, politicians, scientists, religious groups, and myriad others. Section 230 was passed recognizing the unique nature of platforms that host user-generated content and allows us to protect our users in the face of ever-evolving content and threats. Section 230 safeguards open access to information and free expression online. Instead of overblocking speech, the law supports platforms' ability to responsibly manage content.

In some cases, we may also act as an information content provider—for instance, when we publish a blogpost on blog.google. In such instances, we would be treated as the information content provider with respect to that specific content, but that does not and should not affect how our services overall are treated under the law.

We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 20. Should algorithms that promote or demote particular viewpoints be protected by Section 230? Why or why not?

Answer. Our products are built for everyone, and we design them with extraordinary care to be a trustworthy source of information without regard to political viewpoint. Our users overwhelmingly trust us to deliver the most helpful and reliable information available. Distorting results or moderating content for political purposes would be antithetical to our mission and contrary to our business interests.

Our services organize, rank, and recommend content in a wide variety of ways to help meet people's needs and interests. Indeed, this is the essence of most online services today. Curtailing Section 230 based on the use of ranking algorithms would thus undermine the many benefits of the statute today. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 21. Do you think the use of an individual company's algorithms to amplify the spread of illicit or harmful materials like online child sexual exploitation should be protected by Section 230?

Answer. We've always been proponents of free speech, but have always had rules of the road and are never going to be "neutral" about issues like child abuse, terrorism, and harassment. We are very focused on the protection of children on our platforms. Section 230 is what permits us to curate content to protect users, and changes to Section 230 could jeopardize removals of terrorist content, spam, malware, scams, misinformation, manipulated media, and hate speech.

At Google, we have invested heavily in technologies and efforts to protect children on our platform, like our Content Safety API and CSAI Match tools (*https://www.youtube.com/csai-match/*). We already proactively look for and report illegal child sexual abuse to NCMEC—filing more than 449,000 reports to the NCMEC Cyber Tipline in 2019 alone. We are also a leading member of the Technology Coalition, where child safety experts across the industry build capacity and help companies working to increase their capacity to detect Child Sexual Abuse Material (CSAM) (*https://www.technologycoalition.org/*). In June, the Tech Coalition announced a multi-million dollar Research and Innovation Fund and Project Protect—a cross-industry initiative to combat CSAM through investment, research, and information sharing. For more information, please see *https://www.technologycoalition.org/2020/05/28/a-plan-to-combat-online-child-sexual-abuse.*

This is a very important issue and we're committed to continue working with Congress on it.

Question 22. Should platforms that knowingly facilitate or distribute Federal criminal activity or content be immune from civil liability? Why or why not?

Question 23. If your company has actual knowledge of content on your platform that incites violence, and your company fails to remove that content, should Federal law immunize your company from any claims that might otherwise be asserted against your company by victims of such violence? Are there limitations or exceptions to such immunity that you could propose for consideration by the Committee?

Question 24. Should platforms that are willfully blind to Federal criminal activity or content on their platforms be immune from civil liability? Why or why not?

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 22–24.

Section 230 helps Internet companies address harmful content, including user comments, and while we've always been proponents of free speech, we've also always had rules of the road and are never going to be "neutral" about harmful content. Millions of small and large platforms and websites across the Internet rely on Section 230 to both keep users safe and promote free expression. Google also has

worked closely with law enforcement and organizations such as NCMEC, Thorn, and Polaris for years. Under existing law, Section 230s protections for online platforms already exempt all Federal criminal law. We have concerns that changes to Section 230 would negatively impact our ability to remove harmful content of all types and would make our services less useful and safe. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 25. Ranking has been described as Google's Holy Grail. During the 2020 election (January 1, 2020 to date), how was ranking used for searches related to candidates and election-related issues to control the content seen by its users?

Question 26. During the 2020 election (January 1, 2020 to date), did Google lower the search visibility, suppress, or de-rank in any way any search results for any candidates, or election-related issues? If so, how and when was this done, and for what reason?

Question 27. During the 2020 election (January 1, 2020 to date), did Google lower the search visibility, suppress, or de-rank in any way search results for any news outlets, including Breitbart News, the Daily Caller, or the Federalist? If so, how and when was this done, and why?

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 25–27.

Our business model is dependent on being a useful and trustworthy source of information for everyone and we design Search and other products with extraordinary care to serve our mission without regard to politics or political viewpoint. Our Search algorithm ranks pages to provide the most useful and relevant information by matching search terms against available web pages and looking at factors like the number of times the words appear and freshness of the page. Political viewpoint is not a relevant factor to our Search results or ranking and the 2020 election cycle was no exception. We also seek to ensure that our Search results are providing the most authoritative and responsive results by using external quality raters from across the vast majority of U.S. states. In addition, we have robust systems in place to ensure that our policies are enforced in a politically impartial way across all of our products and services, including Search.

For the 2020 election, we worked with The Associated Press (AP)—a trusted source of information on election results—to provide authoritative elections results on Google. When people come to Search looking for information on election results, we provide a dedicated feature with data provided by the AP to provide real-time information on Google for Federal and state level races.

Finally, we do not manually intervene in any particular Search result. We remove content from our organic Search results only in very limited circumstances, such as a court order, valid DMCA takedown notice, or violation of our webspam policies. Search ranking considers many factors, but political ideology is not factored into the process. When we make changes to Search, they apply broadly, after extensive testing and controls, rigorous evaluation, and use of detailed metrics. Our publicly available Search Quality Rater Guidelines (*https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf*) provide transparency for ratings, users, and webmasters about how Search works. And they make it clear that ratings should never be based on personal opinions, preferences, religious beliefs, or political views.

Response to Written Questions Submitted by Hon. Jerry Moran to Sundar Pichai

Question 1. How much money does your company spend annually on content moderation in general?

Answer. Our mission is to organize the world's information and make it universally useful and accessible. As such, it is difficult for us to separate our content moderation efforts and investments from our overall efforts and investments. However, we estimate that we spent at least $1 billion over the past year on content moderation systems and processes. We continue to invest aggressively in this area.

Question 2. How many employees does your company have that are involved with content moderation in general? In addition, how many outside contractors does your company employ for these purposes?

Answer. We enforce our content policies at scale and take tens of millions of actions every day against content that does not abide by the policies for one or more of our products. To enforce our policies at scale, we use a combination of reviewers and AI moderation systems.

In the last year, more than 20,000 people have worked in a variety of roles to help enforce our policies and moderate content. Content moderation at Google and YouTube is primarily managed by Trust and Safety teams across the company. These teams are made up of engineers, content reviewers, and others who work across Google to address content that violates any of our policies. These teams also work with our legal and public policy teams, and oversee the vendors we hire to help us scale our content moderation efforts, as well as provide the native language expertise and the 24-hour coverage required of a global platform. Moderating content at scale is an immense challenge, but we see this as one of our core responsibilities and we are focused on continuously working towards removing content that violates our policies before it is widely viewed.

Question 3. How much money does your company currently spend on defending lawsuits stemming from users' content on your platform?

Answer. Google is subject to and defends against numerous lawsuits and legal claims each year, including content-related claims, ranging from complex Federal litigation to local small claims count claims. Alphabet Inc.'s annual 10–K filing (*https://abc.xyz/investor/*) includes information on the types of material, public legal matters we defend. As detailed in our 10–K filings, we are subject to claims, suits, and government investigations involving content generated by our users and/or based on the nature and content of information available on or via our services, as well as other issues such as competition, intellectual property, data privacy and security, consumer protection, tax, labor and employment, commercial disputes, goods and services offered by advertisers or publishers using our platforms, and other legal theories.

Question 4. Without Section 230s liability shield, would your legal and content moderation costs be higher or lower?

Answer. Without Section 230, we certainly could face an increased risk of liability and litigation costs for decisions around removal of content from our platforms. For example, YouTube might face legal claims for removing videos we determine could harm or mislead users in violation of our policies. Or we might be sued for trying to protect our users from spam and malware on Gmail and Search.

By putting potentially every decision around content moderation up to judicial review, we have concerns that this type of change would negatively impact our ability to remove harmful content of all types and would make our services less useful and safe. Millions of small and large platforms and websites across the Internet rely on Section 230 to both keep users safe and promote free expression. We believe that Section 230 strikes the appropriate balance that facilitates making more content and diverse points of view available than ever before in history, all while ensuring Internet companies can keep their platforms safe and secure for our users. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 5. How many liability lawsuits have been filed against your company based on user content over the past year?

Question 6. Please describe the general breakdown of categories of liability, such as defamation, involved in the total number of lawsuits over the past year.

Question 7. Of the total number of liability lawsuits based on user content, how many of them did your company rely on Section 230 in its defense?

Question 8. Of the liability lawsuits based on user content in which your company relies on Section 230 in its defense, what categories of liability in each of these lawsuits is your company subject to?

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 5–8.

As noted in our response to Question No. 3 above, Google is subject to and defends against numerous lawsuits and legal claims each year, including content-related claims, ranging from complex Federal litigation to local small claims court claims. Alphabet Inc.'s annual 10–K filing (*https://abc.xyz/investor/*) includes information on the types of material, public legal matters we defend.

Question 9. In a defamation case based on a user content, please describe the typical procedural steps your company takes to litigate these claims.

Answer. While Google is unable to provide privileged or other information that could impact its legal strategy in current or future matters, Google defends against claims to the full extent permitted by law. We have strong policies across our products to protect our users, including our content moderation policies. We take allegations of this nature seriously and take appropriate action.

Question 10. Of the claims that have been dismissed on Section 230 grounds, what is the average cost of litigation?

Answer. As noted in our response to Question Nos. 3 and 5–8 above, Google is subject to and defends against numerous lawsuits and legal claims each year, including content-related claims, ranging from complex Federal litigation to local small claims count claims. Alphabet Inc.'s annual 10–K filing (*https://abc.xyz/investor/*) includes information on the types of material, public legal matters we defend.

Question 11. I understand the U.S.-Mexico-Canada Agreement (MUSCAT) contains similar intermediary liability protections that Section 230 established domestically. The recent trade deal with Japan also included similar provisions.

If Congress were to alter Section 230, do you expect litigation or free trade agreement compliance issues related to the United States upholding trade agreements that contain those provisions?

Answer. Section 230 is one of the foundational laws that has enabled America's technology leadership and success in the Internet sector—allowing freedom of expression to flourish online. And it has worked remarkably well in the United States. Some other countries are increasingly looking to regulate, restrict, and censor content in a way that harms U.S. exporters and U.S. creators. Including pro-speech and pro-innovation rules in a trade agreement helps the U.S. push back on those regimes and defend an open Internet globally. The online liability provisions of USMCA (Article 19.17) and the U.S.-Japan Trade Agreement are aligned with Section 230 and ensure that service providers are not held liable for third-party content published on their platforms.

Because Section 230 is not the only U.S. law included in trade agreements (*e.g.,* copyright protections are often included), the litigation and compliance risks associated with upholding trade agreements containing Section 230-like protections that Congress may change are generally similar to the risks associated with upholding trade agreements containing other U.S. laws that Congress may change.

Question 12. How does the inclusion of Section 230-like protections in the aforementioned trade deals affect your business operations in the countries party to said trade deals? Do you expect fewer defamation lawsuits and lower legal costs associated with intermediary liability in those countries due to these trade deals?

Answer. The importance of Section 230 to the U.S. economy has grown since Section 230 was first introduced in the 1990s. It has generated a robust Internet ecosystem where commerce, innovation, and free expression all thrive—while at the same time enabling providers to develop content detection mechanisms and take aggressive steps to fight online abuse.

Intermediary safe harbors in trade deals are critical to digital trade and contribute to the success of the U.S. economy. A recent report found that over the next decade, Section 230 will contribute 4.25 million additional jobs and $440 billion in growth to the economy. (NetChoice and the Copia Institute, "Don't Shoot The Message Board: How Intermediary Liability Harms Online Investment and Innovation" (June 25, 2019), *https://copia.is/library/dont-shoot-the-message-board/*). Section 230 is also a key contributor to the U.S.'s $172 billion digital trade surplus and helps large and small firms run a global business. (Internet Association, "A Look At American Digital Exports" (January 23, 2019), *https://internetassociation.org/publications/a-look-at-american-digital-exports/*.)

It is important for businesses to be able to moderate content and to prevent censorship from other, more oppressive regimes abroad. Including pro-speech and pro-innovation rules in trade agreements helps us avoid the costs and harm (including lawsuits and legal costs) associated with overbroad intermediary liability.

Question 13. In countries that do not have Section 230-like protections, are your companies more vulnerable to litigation or liability as a result?

Question 14. How do your content moderation and litigation costs differ in these countries compared to what you might expect if Section 230-like protections were in place?

Question 15. As American companies, does Section 230s existence provide you any liability protection overseas in countries that do not have similar protections for tech companies?

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 13–15.

It is no accident that the greatest Internet companies in the world were created in the United States. As noted in other responses related to Section 230, it is one of the foundational laws that has enabled America's technology leadership and success in the Internet sector—allowing freedom of expression to flourish online.

While liability limitations under Section 230 exist in U.S. courts, in countries without Section 230-like protections, we could face an increased risk of liability and litigation costs for decisions around removal of content from our platforms. Threats

to our platforms and our users are ever-evolving, and the nature of the content we see is always changing. Section 230 enables Google and other platforms to act quickly to remove violative content and avoid the increased risk of liability and litigation costs associated with such intermediary liability.

Question 16. To differing extents, all of your companies rely on automated content moderation tools to flag and remove content on your platforms.

What is the difference in effectiveness between automated and human moderation?

Question 17. What percentage of decisions made by automated content moderation systems are successfully appealed, and how does that compare to human moderation decisions?

Question 18. Please describe the limitations and benefits specific to automated content moderation and human content moderation.

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 16–18.

To enforce our policies at scale, we rely on a mix of automated and manual efforts to spot problematic content in violation of our policies. In addition to flags by individual users, sophisticated automated technology helps us detect problematic content at scale. Our automated systems are carefully trained to quickly identify and take action against spam and violative content. This includes flagging potentially problematic content for reviewers, whose judgment is needed for the many decisions that require a more nuanced determination.

Automated flagging also allows us to identify and act more quickly and accurately to remove violative content, lessening both the burden on human reviewers and the time it takes to remove such content. Our machine learning systems are faster and more effective than ever before and are helping our human review teams remove content with speed and volume that could not be achieved with people alone. For example, in the third quarter of 2020, more than 7.8 million videos were removed from YouTube for violating our community guidelines. Ninety-four percent of these videos were first flagged by machines rather than humans. Of those detected by machines, over 45 percent never received a single view, and just over 80 percent received fewer than 10 views. In the same period, YouTube removed more than 1.1 billion comments, 99 percent of which were detected automatically. For more information, please see our YouTube Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/removals*).

As we work continuously to improve and enhance information quality and our content moderation practices, we rely heavily on machines and technology, but reviewers also play a critical role. New forms of abuse and threats are constantly emerging that require human ingenuity to assess and develop appropriate plans for action. Our reviewers perform billions of reviews every year, working to make fair and consistent enforcement decisions in enforcing our policies and helping to build training data for machine learning models.

For example, at YouTube, reviews of Community Guideline flags and other notices are conducted by our technological systems in conjunction with Google reviewers. We have a robust quality review framework in place to make sure our global staff are consistently making the appropriate decisions on reported content, and receive regular feedback on their performance. We also operate dedicated threat intelligence and monitoring teams (*e.g.,* Google's Threat Analysis Group, *https://blog.google/threat-analysis-group*), which provide insights and intelligence to our policy development and enforcement teams so they can stay ahead of bad actors.

If users believe their Google Accounts have been suspended or terminated in error, they can appeal the decision. Users can learn more about their rights relating to our terms of service at *https://policies.google.com/terms.*

We also regularly release reports that detail how we enforce our policies and review content. For example, our YouTube Community Guidelines Enforcement Transparency Report includes information on the number of appealed videos that have been reinstated (*https://transparencyreport.google.com/youtube-policy/appeals*). In addition to our general terms of service, we publish comprehensive guides detailing the appeals process to address take-down concerns, including information on Search Reconsideration Requests (*https://support.google.com/webmasters/answer/35843*), Ads disapprovals and suspensions (*https://support.google.com/adspolicy/topic/1308266*), publisher Policy Center violations (*https://support.google.com/adsense/answer/7003627*), and YouTube Community Guidelines violations (*https://support.google.com/youtube/answer/185111*). We are transparent about our decisions and discuss them further in places like our How Search Works page (*https://www.google.com/search/howsearchworks/mission/open-web/*) and our Transparency Report (*https://transparencyreport.google.com/*).

Over the past two decades, we have worked hard to maintain a safe community, have invested heavily in our enforcement program that relies on both people and technology, and we will continue to do so.

Question 19. In your written testimonies, each of you note the importance of tech companies being transparent with their users.

Have you already, or do you plan to make public the processes that your automated moderation system undertakes when making decisions about content on your platform?

Question 20. Given the complexity of the algorithms that are now governing a potion of the content across your platforms, how have you or how do you plan to explain the functions of your automated moderation systems in a simple manner that users can easily understand?

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 19 and 20.

Google is committed to transparency in our business practices, including our content moderation efforts. Our policies work best when users are aware of the rules and understand how we enforce them. That is why we work to make this information clear and easily available to all.

As to our automated and manual content moderation systems, we publish extensive information explaining how they work, including information relating to automated review for invalid Ads activity (*https://www.google.com/ads/adtrafficquality/how-we-prevent-it/*), YouTube Community Guidelines Enforcement (*https://support.google.com/transparencyreport/answer/9209072*), and YouTube ads review systems (*https://support.google.com/youtube/answer/9269751*). We also publish comprehensive guides regarding our content moderation policies in general, including our Publisher Policies (*https://support.google.com/adsense/answer/9335564*), our Publisher Center for Google News (*https://support.google.com/news/publishercenter/*), permissible content for Ads (*https://support.google.com/adspolicy/answer/6008942*), YouTube's Community Guidelines (*https://www.youtube.com/about/policies/#community-guidelines*), our Webmaster guidelines for Search (*https://support.google.com/webmasters/answer/35769*), Google Play Policies regarding restricted content (*https://play.google.com/about/restricted-content/*), and our Terms of Service relating to content (*https://policies.google.com/terms#toc-content*). We endeavor to make all of our policies publicly available and easy to find.

Question 21. How has COVID–19 impacted your company's content moderation systems?

Question 22. Is there a greater reliance on automated content moderation?

Please quantify how content moderation responsibilities have shifted between human and automated systems due to COVID–19.

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 21 and 22.

We are proud of our efforts during this unprecedented public health crisis. Since the outbreak of COVID–19, teams across Google have launched 200 new products, features, and initiatives and are contributing over $1 billion in resources to help our users, clients, and patners through this unprecedented time. That includes our Homepage "Do the Five" promotion, launch of the COVID–19-focused site (*https://www.google.com/intl/en_us/covid19/*), and amplifying authoritative voices through ad grants. There have been over 400 billion impressions on our information panels for coronavirus related videos and searches, and since February we've removed over 270 million coronavirus related ads across all Google advertising platforms, and 600,000 coronavirus videos, globally. We have invested heavily to make sure that we surface authoritative content in our search results, which significantly reduces the spread of misinformation, and we will continue to do so after the coronavirus crisis.

As to our use of automated tools for content moderation during the pandemic, in the face of temporary reductions in our extended workforce, we reallocated employees to prioritize addressing egregious content and supported their doing this work onsite, taking extra health and safety precautions, and providing private transportation. These content moderators ensured we still had capacity to take action on high priority workflows and flags for egregious content, including flags from our Trusted Flagger program and governments. Where feasible, we relied more heavily on automated systems to reduce the need for people to come into the office. Given the resulting risk of false positives (*e.g.,* more legitimate content being automatically but incorrectly removed), we also worked to ensure content creators could appeal and would not wrongly receive strikes against their accounts.

Question 23. Last year, the Department of Justice's Antitrust Division held a workshop that brought together academics and executives from leading companies,

including buyers and sellers of advertising inventory. The discussion explored the practical considerations that industry participants face and the competitive impact of technological developments such as digital and targeted advertising in media markets, including dynamics between local broadcast and online platforms for advertisement expenditures.

Separately, the FCC has attempted to update its local broadcast ownership rules following its 2018 quadrennial review, including permitting the ownership of two TV stations in local markets. However, this recent attempt by the FCC to modernize the local media marketplace has been halted by the Third Circuit's decision to reject the FCC's update of broadcast ownership restrictions.

For purposes of understanding your companies' general views on the local media marketplace, do your companies compete with local broadcast stations for digital advertising revenue?

Answer. Yes. Google is just one player in a crowded advertising market, competing against a wide array of business, from digital advertising businesses to local broadcast stations. We compete for ad dollars with lots of different formats, from websites to apps to billboards to radio to TV. In this robustly competitive market, advertisers decide where to focus their advertising spend, whether it be on Google properties or non-Google properties like local broadcast stations, and may do so based on a variety of factors unique to an advertiser and its advertising goals.

Do you think Federal regulations determining acceptable business transactions in local media marketplaces should be updated to account for this evolving and increasing competition for digital advertising purchases?

Answer. Our grounding principle is that we should do what's right for users. While no system is perfect, the U.S. approach has encouraged strong competition that has delivered world-leading innovation for consumers. We are committed to continuing to work with Congress on evolving that framework to advance the interests of consumers, but should not lose sight of the significant technological competition we face around the world.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. MIKE LEE TO SUNDAR PICHAI

Question 1. Congress is in the midst of a debate over future reforms to Section 230. This is an important discussion that Congress should have.

a. In making decisions to moderate third-party content on your platform, do you rely solely on Section 230? In other words, could you still moderate third-party content without the protections of Section 230?

b. If the provisions of Section 230 were repealed or severely limited, how would your content moderation practices shift?

Answer. Because the answers to these questions are related, we have grouped together our response to all subparts of Question No. 1.

Section 230 safeguards open access to information and free expression online. Instead of overblocking speech, the law supports platforms' ability to responsibly manage content. In this way, Section 230 is one of the foundational laws that has enabled America's technology leadership and success in the Internet sector. Millions of small and large platforms and websites across the Internet rely on Section 230 to keep users safe by addressing harmful content and to promote free expression. While we could moderate third-party content in the absence of Section 230, we have concerns that changes to Section 230 could jeopardize removals of, among other things: terrorist content, spam, malware, scams, misinformation, manipulated media and hate speech, and other objectionable content.

The ability to remove harmful, but not necessarily illegal, content has been particularly important during COVID–19. In just one week earlier this year, we saw 18 million malware and phishing e-mails related to coronavirus and more than 240 million COVID-related spam messages. Since February, we've removed over 600,000 YouTube videos with dangerous or misleading coronavirus information and over 270 million COVID-related ads.

Furthermore, before companies advertise, they want some assurance that publishers' sites are appropriate for their ads. That's where our longstanding content policies come in—they are business-driven policies to ensure ads do not appear alongside offensive content. Please see our User-Generated Content Overview, *https://support.google.com/adsense/answer/1355699.* For example, a company marketing baby clothes wouldn't want its paid ads to appear alongside violent or mature content. Our content policies cover the entire site where ads are displayed, including user-generated comments sections.

We believe that Section 230 strikes the appropriate balance that facilitates making available more content and diverse points of view than ever before in history, while ensuring Internet companies can keep their platforms safe and secure for our users. We also understand that these are important issues and remain committed to working with Congress on them.

Question 2. How many content posts or videos are generated by third-party users per day on Facebook, Twitter, and YouTube?

Answer. YouTube has more than 500 hours of video uploaded every minute, and every day, people watch over a billion hours of video and generate billions of views.

c. How many decisions on average per day does your platform take to moderate content? Are you able to provide data on your takedown numbers over the last year?

Answer. In addition to our automated systems, our reviewers perform billions of reviews every year, working to make appropriate content policy enforcement decisions and helping build training data for machine learning models. We are transparent about our policies and enforcement decisions and discuss them in places like our Transparency Report (*https://transparencyreport.Google/*) and YouTube Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/removals*). For example, in the third quarter of 2020 alone, more than 7.8 million videos were removed from YouTube for violating our community guidelines (*https://transparencyreport.google.com/youtube-policy/removals*).

d. Do you ever make mistakes in a moderation decision? If so, how do you become aware of these mistakes and what actions do you take to correct them?

e. What remedies or appeal process do you provide to your users to appeal an action taken against them? On average, how long does the adjudication take until a final action is taken? How quickly do you provide a response to moderation decision appeals from your customers?

Answer. Because the answers to these questions are related, we have grouped together our response to these subparts of Question No. 2.

We sometimes make mistakes in our decisions to enforce our policies, which may result in the unwarranted removal of content from our services. To address that risk, wherever possible, we make it clear to creators that we have taken action on their content and—as noted in our responses to Chairman Wicker's Question No. 4 and Senator Moran's Question Nos. 16–18—provide them with the opportunity to appeal that decision and give us any clarifications they feel are relevant.

We want to make it easy for good-faith actors to understand and abide by our rules, while making it challenging for bad actors to flout them. That is why we seek to make room for good-faith errors as we enforce our rules, and provide the opportunity for users to appeal decisions and provide clarification—decisions that are then evaluated by a different member of our Trust and Safety team. Our policies detailing the appeals process to address take-down concerns include information on Search Reconsideration Requests (*https://support.google.com/webmasters/answer/35843*), Ads disapprovals and suspensions (*https://support.google.com/adspolicy/topic/1308266*), publisher Policy Center violations (*https://support.google.com/adsense/answer/7003627*), and YouTube Community Guidelines violations (*https://support.google.com/youtube/answer/185111*).

Many of these policies also detail how long it takes for us to review appeals and provide a response. For example, our publisher Policy Center site (*https://support.google.com/adsense/answer/7003627*) explains that we typically review a site within one week of the publisher's request for review, but sometimes it can take longer. Similarly, our Search Console Help site (*https://support.google.com/webmasters/answer/9044175*) explains that Search reconsideration reviews can take anywhere from several days up to a week or two.

We are transparent about our policy enforcement decisions and discuss them further in places like our How Search Works page (*https://www.google.com/search/howsearchworks/mission/open-web/*), Google Transparency Report (*https://transparencyreport.google.com/*), and YouTube Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/removals*).

f. Can you provide approximate numbers, by month or week, for the times you took down, blocked, or tagged material from November 2019 to November 2020?

Answer. We enforce our content policies at scale and take tens of millions of actions every day against content that does not abide by the "rules of the road" for one or more of our products. For example, in the third quarter of 2020 alone, we removed over 1.1 billion comments, 7.8 million videos, and 1.8 million channels on YouTube for violating our Community Guidelines. In 2019, we blocked and removed 2.7 billion ads, terminated over 1.2 million accounts, and removed ads from over 21

million web pages that are pat of our publisher network for violating our policies. Google is transparent about our content moderation decisions, and we include information on these takedowns in places like our Google Transparency Report (*https://transparencyreport.google.com/*), YouTube Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/removals*), and other updates, such as our quarterly Threat Analysis Group ("TAG") Bulletins (*e.g., https://blog.google/threat-analysis-group/tag-bulletin-q2-2020/*) and annual report on Ads take downs (*e.g., https://www.blog.google/products/ads/stopping-bad-ads-to-protect-users/*).

Question 3. The first major case to decide the application of Section 230 was *Zeran v. AOL*.[1] In *Zeran,* Judge Wilkinson recognized the challenges of conferring "distributor liability" to a website because of the sheer number of postings. That was 1997. If we imposed a form of "distributor liability" on your platforms that would likely mean that your platform would be liable for content if you acquired knowledge of the content. I think there is an argument to be made that you "acquire knowledge" when a user "flags" a post, video, or other form of content.

g. How many "user-generated" flags do your companies receive daily?

Answer. Google is committed to transparency in our business practices, including our content moderation efforts. We publish information on removal requests and user flags in our Google Transparency Report (*https://transparencyreport.google.com/*) and YouTube Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/removals*). For example, in the third quarter of 2020 alone, we removed over 480,000 videos on YouTube flagged by users and our Trusted Flaggers for violating YouTube Community Guidelines (*https://transparencyreport.google.com/youtube-policy/removals*).

h. Do users ever flag posts solely because they disagree with the content?

Answer. Our policies provide that users should only flag content that violates Google policies. That said, users often flag content that does not violate our policies. When this occurs, we do not take action on the flagged content. Many of our policies and blog posts also instruct users not to flag content simply because they disagree with it (*e.g.,* please see our Maps policy on flagging content, *https://support.google.com/contributionpolicy/answer/7445749*).

i. If you were liable for content that was "flagged" by a user, how would that affect content moderation on your platform?

Answer. Section 230 allows us to have content rules that we enforce in an effort to ensure that our platform is safe for our users. Without Section 230, platforms could be sued for decisions around removal of content from their platforms. As a result, search engines, video sharing platforms, political bogs, startups, and review sites of all kinds would either not be able to filter content at all (resulting in more offensive online content, including adult content, spam, security threats, etc.) or would over-filter content (possibly including important cases of political speech)—in either scenario, harming consumers and businesses that rely on and use these services every day.

We also note that we've always had robust policies, but finding all violative content is an immense challenge. We take tens of millions of actions every day against content that does not abide by the "rules of the road" for one or more of our products. For example, in the third quarter of 2020 alone, we removed over 480,000 videos on YouTube flagged by users and our Trusted Flaggers for violating YouTube Community Guidelines. More information about our efforts is available in our Transparency Report at *https://transparencyreport.google.com.*

Question 4. Section 230 is often used as a legal tool to have lawsuits dismissed in a pre-trial motion.

j. How often is your company sued under a theory that you should be responsible for the content posted by a user of your platform? How often do you use Section 230 as a defense in these lawsuits? And roughly how often are those lawsuits thrown out?

Answer. As noted in our response to Senator Moran's Question Nos. 5–8, Google is subject to and defends against numerous lawsuits and legal claims each year, including content-related claims, ranging from complex Federal litigation to local small claims court claims. Alphabet Inc.'s annual 10–K filing (*https://abc.xyz/investor/*) includes information on the types of material, public legal matters we defend.

k. If Section 230 was eliminated and a case seeking to make your platform liable for content posted by a third party went to the discovery phase, roughly how much more expensive would that case be as opposed to its dismissal pre-discovery?

---

[1] Kenneth M. Zeran v. America Online, Inc. 129 F. 3d 327 (4th Cir. 1997)

Answer. While we do not have an exact figure or estimate, without Section 230, we certainly could face an increased risk of liability and litigation costs—including the routinely high costs of discovery—for decisions around removal of content from our platforms.

Question 5. Section 230s Good Samaritan provision contains the term "otherwise objectionable."

l. How do you define "otherwise objectionable"?

m. Is "otherwise objectionable" defined in your terms of service? If so, has its definition ever changed? And if so, can you provide the dates of such changes and the text of each definition?

Answer. Because the answers to these questions are related, we have grouped together our response to these subparts of Question No. 5.

We do not have a specific definition in our terms of service regarding "otherwise objectionable" content; rather, we describe in our terms of service and other policies a range of harmful or objectionable content (*https://policies.google.com/terms*). The term "otherwise objectionable" in the statute allows us to take action against such harmful content including spam, malware, phishing, spyware, view count manipulation, user data/permissions violations, scams and other deceptive practices, misinformation and manipulated media, hate speech or other derogatory content, and dangerous content (*e.g.,* weapon manufacturing and sales). We would be pleased to continue our ongoing dialogue with Congress on Section 230.

n. In most litigation, a defendant relies on Section 230(c)(1) for editorial decisions. If a company could only rely on 230(c)(2) for a moderation decision (as has been discussed in Congress), how would that affect your moderation practices? And how would striking "otherwise objectionable" from 230(c)(2) further affect your moderation practices?

Answer. Section 230 is one of the foundational laws that has enabled America's technology leadership and success in the Internet sector—allowing freedom of expression to flourish online. Section 230 has worked remarkably well, and we believe a cautious and thoughtful approach to potential changes is appropriate. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Google agrees that exercising good faith is important, and we also believe we already engage in good faith content moderation. We have published terms of service (*https://policies.google.com/terms*), make good faith efforts to enforce our policies, and provide opportunities to appeal a decision.

Changes to Section 230 would negatively impact our ability to remove harmful content of all types and would make our services less useful and safe. We are concerned that striking "otherwise objectionable" could put removals of spam, malware, fraud, scams, misinformation, manipulated media, and hate speech at risk. This is especially important as threats to our platforms and our users are ever-evolving, and the nature of the content we see is always changing. If we were to have specific exceptions, we would lose the ability to act in real time on troubling and dangerous content that we are seeing for the first time.

Question 6. Are your terms of service a legally binding contract with your users? How many times have you changed your terms of service in the past five years? What recourse do users of your platform have when you allege that they have violated your terms of service?

Answer. Yes, in order to use our services, users must agree to our terms of service. We want to be as transparent as possible about the changes we make to our terms of service. That is why we have archived the versions of our Google terms of service from 1999 to present. As shown in the archive, we have changed our Google terms of service two times in the past five years. We've also included comparisons of each version to the previous version to make it as easy as possible to see what has changed. Please see *https://policies.google.com/terms/archive* for more information.

As noted above in our responses to Chairman Wicker's Question No. 4 and Senator Moran's Question Nos. 16–18, if users believe their Google Accounts have been suspended or terminated in error, they can appeal the decision. Users can learn more about their rights relating to our terms of service at *https://policies.google.com/terms.*

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. RON JOHNSON TO SUNDAR PICHAI

Question 1. During the hearing, in response to both Senator Cruz's line of questioning and mine, Mr. Dorsey claimed that Twitter does not have the ability to influence nor interfere in the election.

a. Do you believe Google has the ability to influence and/or interfere in the election? To reiterate, I am not asking if you have the intent or have actively taken steps to influence/interfere, but rather if Google has the *ability.*

b. If you claim that you do not have the ability to influence or interfere in the election, can you explain Google's rational for suppressing content that Google deems to be Russian misinformation on the basis that it influences the election?

Answer. Because the answers to these questions are related, we have grouped together our response to all subparts of Question No. 1.

While many companies and organizations across a wide range of industries hypothetically could have the ability to influence an election through action or inaction (*e.g.,* a voter registration group seeking to increase voter registration or participation), Google makes no attempt to do so. Our products are built for everyone, and we design them with extraordinary care to be a trustworthy source of information without regard to politics or political viewpoint. Our users overwhelmingly trust us to deliver the most helpful and reliable information, and distorting results or moderating content for political or ideological purposes would be antithetical to our mission and contrary to our business interests.

We also have undertaken a wide range of approaches to protect the integrity of elections and prevent platform abuse, and we will continue to do so. For instance, our efforts during the 2020 election cycle focused on four different areas: elevating authoritative election-related content; combating coordinated influence operations; protecting users and campaigns; and continuing to work with law enforcement and industry patners on identifying and combating coordinated influence operations. Our teams are constantly on the lookout for malicious actors that try to game our platforms and we take strong action against coordinated influence operations. In May of this year, we also announced the launch of our Threat Analysis Group ("TAG") bulletin (*https://www.blog.google/threat-analysis-group/updates-about-government-backed-hacking-and-disinformation/*) to provide more frequent, quarterly updates about our efforts. (Q1—*https://blog.google/threat-analysis-group/tag-bulletin-q1-2020;* Q2—*https://blog.google/threat-analysis-group/tag-bulletin-q2-2020;* Q3—*https://blog.google/threat-analysis-group/tag-bulletin-q3-2020.*)

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. MARIA CANTWELL TO SUNDAR PICHAI

Foreign Disinformation. Facebook/Instagram, Twitter, and Google/YouTube have each taken concrete steps to improve defensive measures through automated detection and removal of fake accounts at creation; increased internal auditing and detection efforts; and established or enhanced security and integrity teams who can identify leads and analyze potential networks engaging in coordinated inauthentic behavior.

Social media companies have hired a lot of staff and assembled large teams to do this important work and coordinate with the FBI led Foreign Influence Task Force (FITF).

Small companies in the tech sector do not have the same level or expertise or resources, but they face some of the same and growing threats.

Likewise, public awareness and understanding of the threats foreign actors like Russia pose is key to helping fight back against them.

Question 1. What specific steps are you taking to share threat information with smaller social media companies that do not have the same level of resources to detect and stop those threats?

Answer. As noted in our response to Senator Rosen's Question No. 1 (a, b), our teams are constantly on the lookout for malicious actors that try to game our platforms, and we take strong action against coordinated influence operations. We've dedicated significant resources to help protect our platforms from such attacks by maintaining cutting-edge defensive systems and by building advanced security tools directly into our consumer products.

We help industry partners in two main ways. First, to enhance the transparency of these efforts, we regularly release reports that detail how we protect our platforms, enforce our policies, and review content, and hope that these are helpful to other companies in the digital ecosystem as well. For instance, we publicly share

data in places like our Transparency Report (*https://transparencyreport.google.com/*), including data on government removal requests, as well as information about political advertising, such as who is buying election ads on our platforms and how much money is being spent. We make this data available for public research to all who are interested in learning or using it to conduct research or improve their own content moderation efforts, including social media companies.

Second, we have collaborated with industry partners to prevent terrorists and violent extremists from exploiting our platforms. In 2017, YouTube, Facebook, Microsoft, and Twitter founded the Global Internet Forum to Counter Terrorism (GIFT) as a group of companies, dedicated to disrupting terrorist abuse of members' digital platforms. Among other important initiatives, GIFCT allows participating companies and organizations to submit hashes, or "digital fingerprints," of identified terrorist and violent extremist content to a database so that it can be swiftly removed from all participating platforms. By sharing best practices and collaborating on cross-platform tools, we have been able to bring new members to GIFCT and engage more than one hundred smaller technology companies through workshops around the world. For more information on these efforts, please see *https://gifct.org/*.

We continue to develop and learn from these collaborations over time and seek more opportunities to develop best practices jointly with partners of all sizes to help people understand what they see online and to support the creation of quality content.

Question 2. Intel Chairman Schiff has highlighted the need for social media companies to increase transparency about how social media companies have stopped foreign actors disinformation and influence operations. Where are the gaps in public disclosures of this information and what specific actions are you taking to increase transparency about malign foreign threats you have throttled?

Answer. Google is committed to transparency in our business practices, including our efforts to stop foreign disinformation and coordinated influence operations across our platforms. We were the first to launch a Transparency Report and have continued to expand and enhance our transparency efforts across numerous products and services over time. As discussed in our response to Question No. 1, in order to increase transparency about the threats we see to our platforms, we regularly release reports that detail how we protect our platforms, enforce our policies, and review content. For more information about Google's Transparency Reports, please see *https://transparencyreport.google.com/?hl=en.*

As part of our transparency reports, Google publishes information regarding government requests to remove content for each six-month period, and carefully evaluates whether content should be removed because it violates a law or product policy. For a detailed overview of requests by country/region, please see *https://transparencyreport.google.com/government-removals/overview*.

We've also invested in robust systems to detect phishing and hacking attempts (*https://security.googleblog.com/2018/08/a-reminder-about-government-backed.html*), identify influence operations launched by foreign governments, and protect political campaigns from digital attacks through our Protect Your Election program.

Our Threat Analysis Group, working with our partners at Jigsaw and Google's Trust & Safety team, identifies bad actors, disables their accounts, warns our users about them, and shares intelligence with other companies and law enforcement officials. As noted in our response to Senator Rosen's Question No. 1 (a, b), on any given day, our Threat Analysis Group tracks more than 270 targeted or government-backed attacker groups from more than 50 countries. When we find attempts to conduct coordinated influence operations on our platforms, we work with our Trust and Safety teams to swiftly remove such content from our platforms and terminate these actors' accounts. We take steps to prevent possible future attempts by the same actors, and routinely exchange information and share our findings with law enforcement and others in the industry. For example, in October 2020, the U.S. Department of Justice acknowledged Google's contributions to the fight against Iranian influence operations, in announcing the seizure of 92 domain names used by Iran's Islamic Revolutionary Guard Corps to engage in a global disinformation campaign targeting the U.S. and other countries (*https://www.justice.gov/usao-ndca/pr/united-states-seizes-domain-names-used-iran-s-islamic-revolutionary-guard-corps*).

Additionally, if we suspect that users are subject to government-sponsored attacks, we warn them. In April 2020 alone, for example, we sent 1,755 warnings to users whose accounts were targets of government-backed attackers. We share more information about these actions on our Threat Analysis Group blog (*https://blog.google/threat-analysis-group/*), which provides information about actions we take against accounts that we attribute to coordinated influence campaigns, both foreign and domestic.

With respect to YouTube, our YouTube our Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/removals*) explains how our systems and policies are actively at work identifying and removing such content. As noted in our response to Senator Rosen's Question No. 1 (a, b), in the third quarter of 2020, over 7.8 million videos were removed by YouTube for violating Community Guidelines, including policies regarding misleading and dangerous content. As threats evolve, we will continue to adapt to understand and prevent new attempts to misuse our platforms, and will continue to expand our use of cutting-edge technology to protect our users. We also will build upon our transparency efforts in the future, as they are an important component of ensuring an informed public dialogue about the role that our services play in society.

Addressing Stop Hate for Profit Recommendations. The Stop Hate for Profit, Change the Terms, and Free Press coalition—all committed to combating racism, violence, and hate online—have called on social media platforms to adopt policies and take decisive actions against toxic and hateful activities.

This includes finding and removing public and private groups focused on white supremacy, promoting violent conspiracies, or other hateful content; submitting to regular, third party, independent audits to share information about misinformation; changing corporate policies and elevating civil rights to an executive level position.

Question 3. Mr. Pichai, will you commit to making the removal of racist, violent, and hateful content an executive level priority?

Answer. Yes, and we assure you that Google strives to be a safe and inclusive space for all of our users. Improvements are happening every day, and we will continue to adapt, invent, and react as hate and extremism evolve online. We're committed to this constant improvement, and the significant human and technological investments we're making demonstrate that we're in it for the long haul. In the last year, we have spent at least $1 billion on content moderation systems and processes, and more than 20,000 people have worked in a variety of roles to help enforce our policies and moderate content.

As noted in our response to Senator Rosen's Question No. 1 (c), one of the most complex and constantly evolving areas we deal with is hate speech. That is why we systematically review and re-review all our policies to make sure we are drawing the line in the right place, often consulting with subject matter experts for insight on emerging trends. For our hate speech policy, we work with experts in subjects like violent extremism, supremacism, civil rights, and free speech from across the political spectrum.

Hate speech is a complex policy area to enforce at scale, as decisions require nuanced understanding of local languages and contexts. To help us consistently enforce our policies, we have expanded our review team's linguistic and subject matter expertise. We also deploy machine learning to better detect potentially hateful content to send for human review, applying lessons from our enforcement against other types of content, like violent extremism.

We also have recently taken a tougher stance on removing hateful and supremacist content from YouTube and have reduced borderline content by reducing recommendations of content that comes close to violating YouTube's guidelines. Since early 2019, we've increased by 46 times our daily hate speech comment removals on YouTube. And in the last quarter, of the more than 1.8 million channels we terminated for violating our policies, more than 54,000 terminations were for hate speech. This is the most hate speech terminations in a single quarter and three times more than the previous high from Q2 2019 when we updated our hate speech policy. For additional information regarding enforcement of, and improvements to, our hate speech policies, please see *https://blog.youtube/news-and-events/make-youtube-more-inclusive-platform/, https://transparencyreport.google.com/youtube-policy/featured-policies/hate-speech,* and *https://blog.youtube/news-and-events/our-ongoing-work-to-tackle-hate.*

Additionally, in October, we launched a Community Guidelines YouTube update on harmful conspiracy theories (*https://blog.youtube/news-and-events/harmful-conspiracy-theories-youtube/*), which expanded our hate speech and harassment policies to prohibit content that targets an individual or group with conspiracy theories that have been used to justify real-world violence. For example, content such as conspiracy theories saying individuals or groups are evil, corrupt, or malicious based on protected attributes (*e.g.,* age, race, religion, etc.), or hateful supremacist propaganda, including the recruitment of new members or requests for financial support for their ideology, all violate our hate speech policy (*https://support.google.com/youtube/answer/2801939*) and are subject to removal as such.

The openness of our platforms has helped creativity and access to information thrive. It's our responsibility to protect that and prevent our platforms from being used to incite hatred, harassment, discrimination, and violence. We are committed to taking the steps needed to live up to this responsibility today, tomorrow, and in the years to come.

Kenosha Wisconsin Violence. On August 25th, a man from Illinois traveled to Kenosha, Wisconsin armed with an assault rifle and fatally shot Joseph Rosenbaum and Anthony Huber, and injured another person, who were protesting the shooting of Jacob Blake, a Black resident, which left him paralyzed.

In the wake of these tragic shootings, we learned that a para-military group called the Kenosha Guard Militia, a group that organized on Facebook, called on followers to "take up arms" and "defend" the city against "evil thugs." This event post had been flagged 455 times by Facebook users, yet Facebook did not take down the group's page until after these lives were already lost.

While the Illinois shooter may not have been a member of the Kenosha Guard Militia, this brings up a very important point—that hate spread on social media platforms can lead to real life violence.

In May of this year, the *Wall Street Journal* reported that Facebook had completed internal research that said its internal algorithms "exploit the human brain's attraction to divisiveness," which could allow Facebook to feed more divisive content to gain user attention and more time on the platform. In response, the *Journal* reported that Facebook buried the research and did little to address it because it ran counter to other Facebook initiatives.

Sowing divisions in this country and further polarizing public discourse is dangerous, and can have deadly consequences.

Question 4. Mr. Pichai, your company also targets information to people based on what your data tells them they want to see, which can lead to people being stuck in an echo chamber that makes them less likely to listen to other viewpoints. What responsibility do you believe you have to stem the divisive discourse in this country?

Answer. We are committed to making Google a safe and inclusive space for people to share their viewpoints. We understand your concerns and are deeply troubled by any attempts to use our platforms to sow division.

We have put significant effort into combating harmful activity across our platforms. This includes, for instance, ranking algorithms in Search that prioritize authoritative sources. Our Search algorithm ranks pages to provide the most useful and relevant information by matching search terms against available web pages and looking at factors like the number of times the words appear and freshness of the page. A user's viewpoint is not a relevant factor. We also seek to ensure that our Search results are providing the most authoritative and responsive results by using external quality raters from across the vast majority of U.S. states. In addition, we have robust systems in place to ensure that our policies are enforced in a politically impartial way across all of our products and services, including Search.

Additionally, on Google News, we mark up links with labels that help users understand what they are about to read—whether it is local content, an op-ed, or an in-depth piece, and encourage them to be thoughtful about the content they view. Publishers who review third-party claims or rumors can showcase their work on Google News and in Google Search through fact-check labels. People come across these fact checks billions of times per year. For more information, please see *https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/*.

We also have increased transparency around news sources on YouTube, including disclosure of government funding. When a news channel on YouTube receives government funding, we make that fact clear by including an information panel under each of that channel's videos. There have been billions of impressions of information panels on YouTube around the world since June 2018. For more information, please see *https://support.google.com/youtube/answer/7630512,* and *https://blog.youtube/news-and-events/greater-transparency-for-users-around.* Our goal is to equip users with additional information to help them better understand the sources of news content that they choose to watch on YouTube. We also have taken a tougher stance on removing hateful and supremacist content and have reduced borderline content by reducing recommendations of content that comes close to violating our guidelines.

We are proud that we have a wide variety of views expressed on our platforms and are committed to ensuring we continue to enforce our policies in a fair and impartial manner.

Russian Election Interference. The U.S. Intelligence community found that foreign actors including Russia tried to interfere in the 2016 election and used social media platforms among other influence operations.

In 2017, the FBI established the Foreign Influence Task Force (FITF), which works closely with state and local partners to share information on threats and actionable leads.

The FBI has also established relationships with social media companies to enable rapid sharing of threat information. Social media companies independently make decisions regarding the content of their platforms.

The U.S. Intelligence Community warned that Russia was using a range of active measures to denigrate former Vice President Joe Biden in the 2020 election. They also warned about Iran and China.

Social media companies remain on the front lines of these threats to our democracy.

Question 5. What steps are you taking to prevent amplification of false voter fraud claims after the 2020 presidential election and for future elections? What challenges do you face trying to prevent foreign actors who seek to influence our elections?

Answer. Election and civic integrity is an issue that we take very seriously, and we have many different policies and processes to combat election-related misinformation like false voter fraud claims and related violative content. Our efforts relating to the 2020 U.S. Presidential election have focused on four different areas: elevating authoritative election-related content; combating coordinated influence operations; protecting users and campaigns; and continuing to work with law enforcement and industry partners on identifying and combating coordinated influence operations.

For example, as noted in our response to Senator Blumenthal's Question No. 1, all ads, including political ads, must comply with our publicly-available Ads policies (*https://support.google.com/adspolicy/answer/6008942*), which prohibit, among other things, dangerous or derogatory content; content that is illegal, promotes illegal activity, or infringes on the legal rights of others; and content that misrepresents the owner's origin or purpose. We put significant effort into curbing harmful misinformation on our ads platform, including prohibiting content that makes claims that are demonstrably false and could significantly undermine participation or trust in an electoral or democratic process. Given the unprecedented amount of votes that were counted after this past election day, we also implemented a sensitive event policy for political ads after the polls closed on November 3, 2020 (*https://support.google.com/adspolicy/answer/10122500*), which prohibited advertisers from running ads referencing candidates, the election, or its outcome.

Further, we do not allow content on YouTube alleging that widespread fraud or errors changed the outcome of a historical U.S. Presidential election. As December 8 was the safe harbor deadline for the U.S. Presidential election, states have all certified their election results. YouTube will remove any piece of content uploaded anytime after December 8 that misleads people by alleging that widespread fraud or errors changed the outcome of the 2020 U.S. Presidential election, in line with our approach towards historical U.S. Presidential elections. As always, news coverage and commentary on these issues can remain on our site if there's sufficient education, documentary, scientific, or artistic context, as described here: *https://blog.youtube/inside-youtube/look-how-we-treat-educational-documentary-scientific-and-artistic-content-youtube/*.

Additionally, our YouTube Community Guidelines (*https://www.youtube.com/howyoutubeworks/policies/community-guidelines/*) prohibit spam, scams, or other manipulated media, coordinated influence operations, and any content that seeks to incite violence. Since September, we've terminated over 8,000 channels and thousands of harmful and misleading elections-related videos for violating our existing policies. Over 77 percent of those removed videos were taken down before they had 100 views. And, since election day, relevant fact check information panels from third-party fact checkers were triggered over 200,000 times above relevant election-related search results, including for voter fraud narratives such as "Dominion voting machines" and "Michigan recount." For additional information, please see *https://blog.youtube/news-and-events/supporting-the-2020-us-election/*.

Similarly, on Search, we have highlighted fact checks for over three years as a way to help people make more informed judgments about the content they encounter online. For more information, see *https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/*.

We are also aware of the concerns that state-sponsored websites and broadcast channels on YouTube may provide a slanted perspective. That's why we have long taken steps to provide our users with more context and information about news sources, including state-funded sources. When a news channel on YouTube receives

government funding, we make that fact clear by including an information panel under each of that channel's videos. For more information on our information panels, please see *https://support.google.com/youtube/answer/7630512.*

Additionally, our teams are constantly on the lookout for malicious actors that try to game our platforms, and we take strong action against coordinated influence operations. If we suspect that users are subject to government-sponsored attacks, we warn them. In April 2020 alone, for example, we sent 1,755 warnings to users whose accounts were targets of government-backed attackers. For more information, please see *https://blog.google/threat-analysis-group/updates-about-government-backed-hacking-and-disinformation/*. Further, Google's Threat Analysis Group ("TAG") works to counter targeted and government-backed hacking against Google and our users. TAG tracks more than 270 targeted or government-backed groups from more than 50 countries. These groups have many goals including intelligence collection, stealing intellectual property, targeting dissidents and activists, destructive cyber attacks, or spreading coordinated disinformation. We use the intelligence we gather to protect Google infrastructure, as well as users targeted with malware or phishing. In May of last year, we announced the launch of our TAG bulletin (*https://blog.google/threat-analysis-group*) to provide more frequent, quarterly updates about our efforts.

Protecting our platforms from foreign interference is a challenge we have been tackling as a company long before the 2020 U.S. Presidential election. We've dedicated significant resources to help protect our platforms from such attacks by maintaining tools aimed at protecting our physical and network security but also detecting and preventing the artificial boosting of content, spam, and other deceptive practices aiming to manipulate our systems. As threats evolve, we will continue to adapt to understand and prevent new attempts to misuse our platforms and will continue to expand our use of cutting-edge technology to protect our users. We are committed to our ongoing efforts to strengthen protections around elections, ensure the security of users, and help combat disinformation.

Question 6. How has the U.S. Government improved information sharing about threats from foreign actors seeking to interfere in our elections since 2016? Is information that is shared timely and actionable? What more can be done to improve the cooperation to stop threats from bad actors?

Answer. Preventing the misuse of our platforms is something that we take very seriously. We're committed to stopping this type of abuse and working closely with the government and law enforcement on how we can help to combat election interference and promote election integrity and user security.

When we find attempts to conduct coordinated influence operations on our platforms, we work with our Trust and Safety teams to swiftly remove such content from our platforms and terminate these actors' accounts. We also routinely exchange information and share our findings with government agencies, and take steps to prevent possible future attempts by the same actors. For example, in October 2020, the U.S. Department of Justice acknowledged Google's contributions to the fight against Iranian influence operations, in announcing the seizure of 92 domain names used by Iran's Islamic Revolutionary Guard Corps to engage in a global disinformation campaign targeting the U.S. and other countries. For more information, please see *https://www.justice.gov/usao-ndca/pr/united-states-seizes-domain-names-used-iran-s-islamic-revolutionary-guard-corps.*

We are also focused on working with the government on identifying cyber threats. Google Cloud is working closely with the Defense Innovation Unit ("DIU") within the U.S. Department of Defense to build a secure cloud management solution to detect, protect against, and respond to cyber threats worldwide. We are honored to partner with DIU on this critical initiative to protect its network from bad actors that pose threats to our national security. For more information, please see *https://cloud.google.com/press-releases/2020/0520/defense-Innovation-unit.*

Question 7. How are you working with civil society groups like the University of Washington's Center for an Informed Public and Stanford Internet Observatory and Program?

Answer. Combating disinformation campaigns requires efforts from across the industry and the public sector, and we are proudly collaborating with technology and non-governmental organizations (NGO) partners to research and address disinformation and, more broadly, election integrity. We have a long-established policy of routinely sharing threat information with our peers and working with them to better protect the collective digital ecosystem.

For instance, our YouTube Trusted Flagger program helps provide robust tools for individuals, government agencies, and NGOs that are particularly effective at notifying YouTube of content that violates our Community Guidelines. For more infor-

mation, please see *https://support.google.com/youtube/answer/7554338.* The program provides these partners with a bulk-flagging tool and provides a channel for ongoing discussion and feedback about YouTube's approach to various content areas. The program is part of a network of over 180 academics, government partners, and NGOs that bring valuable expertise to our enforcement systems. For instance, to help address violent extremism, these partners include the International Center for the Study of Radicalization at King's College London, the Institute for Strategic Dialogue, the Wahid Institute in Indonesia, and government agencies focused on counterterrorism.

In the context of this past election season, we have engaged with the Election Integrity Partnership comprising the Stanford Internet Observatory, the University of Washington's Center for an Informed Public, Graphika, and the Atlantic Council's DFR Lab, and look forward to continued dialogue as we prepare for future elections in the U.S. and around the globe.

We are committed to continuing to work with the NGO community and others in the industry, as well as Congress and law enforcement, to strengthen protections around elections, ensure the security of users, and help combat disinformation.

Question 8. How are you raising social media users' awareness about these threats? What more can be done? How do you ensure the actions you take do not cross the line into censorship of legitimate free speech?

Answer. We are deeply concerned about any attempts to use our platforms to spread election misinformation and sow division, and have put significant efforts into curbing misinformation on our products. Our response to Senator Rosen's Question No. 1 (a, b) contains responsive information and resource links concerning steps we have taken to enforce our policies relating to election misinformation and civic integrity that we hope are helpful.

Additionally, as noted above in our response to Question No. 5, all ads, including political ads, must comply with our publicly-available Ads policies (*https://support.google.com/adspolicy/answer/6008942*), which prohibit, among other things, dangerous or derogatory content; content that is illegal, promotes illegal activity, or infringes on the legal rights of others; and content that misrepresents the owner's origin or purpose. We put significant effort into curbing harmful misinformation on our ads platform, including prohibiting content that makes claims that are demonstrably false and could significantly undermine participation or trust in an electoral or democratic process. Further, given the unprecedented amount of votes that were counted after this past election day, we also implemented a sensitive event policy for political ads after the polls closed on November 3, 2020 (*https://support.google.com/adspolicy/answer/10122500*), which prohibited advertisers from running ads referencing candidates, the election, or its outcome. Additionally, all advertisers who run U.S. election-related ads must first be verified in order to protect the integrity of the election ads that run on our platform. We're serious about enforcing these policies, and we block and remove ads that we find to be violative. For more information, please see our political content advertising policies, *https://support.google.com/adspolicy/answer/6014595.*

Not only do we remove content that violates our policies, but we also reduce borderline content and raise up authoritative voices by providing users with more information about the content they are seeing to allow them to make educated choices. As noted in our response to Question No. 4, our efforts include better ranking algorithms in Search that prioritize authoritative sources, as well as tougher policies against the monetization of misrepresentative content by publishers. On Google News, we mark up links with labels that help users understand what they are about to read—whether it is local content, an op-ed, or an in-depth piece, and encourage them to be thoughtful about the content they view. Publishers who review third-party claims or rumors can showcase their work on Google News and in Google Search through fact-check labels. People come across these fact checks billions of times per year. For more information, please see *https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/.*

We also have increased transparency around news sources on YouTube, including disclosure of government funding. When a news channel on YouTube receives government funding, we make that fact clear by including an information panel under each of that channel's videos. There have been billions of impressions of information panels on YouTube around the world since June 2018. For more information, please see *https://support.google.com/youtube/answer/7630512,* and *https://blog.youtube/news-and-events/greater-transparency-for-users-around.* Our goal is to equip users with additional information to help them better understand the sources of news content that they choose to watch on YouTube.

As to concerns of censorship, we work hard to make sure that the line between what is removed and what is allowed is drawn in the right place. We believe that

people are best served when they have access to a breadth of diverse content from a variety of sources. That is why, for example, we only remove content from search results in limited circumstances, including based on our legal obligations, copyright, webmaster guidelines, spam, and sensitive personal information like government IDs. Please see, for example, our policies relating to removals for legal obligations, *https://support.google.com/websearch/answer/9673730;* webmaster guidelines *https://developers.google.com/search/docs/advanced/guidelines/webmaster-guidelines;*voluntarily removal policies, *https://support.google.com/websearch/answer/3143948;* and policies concerning removals for copyright infringement, *https://support.google.com/transparencyreport/answer/7347743.* For each product and service we offer, we tailor our policies to distinguish between providing access to a diversity of voices and limiting harmful content and behaviors—whether those are our policies against hate speech or material that is excessively violent, unlawful, deceptive, or obscene (*e.g.,* Advertising Policies, *https://support.google.com/adspolicy/answer/6015406;* Publisher Policies, *https://support.google.com/adsense/answer/9335564;* and YouTube Community Guidelines, *https://support.google.com/youtube/answer/9288567*). We also have safeguards in place to ensure that we enforce these policies in a consistent way without bias as to the ideological viewpoint of the content.

As threats evolve, we will continue to adapt to understand and prevent new attempts to misuse our platforms and will continue to expand our use of cutting-edge technology to protect our users. There are no easy answers here, but we are deeply committed to getting this right.

Foreign Disinformation & Russian Election Interference. Since four years ago, our national security agencies and the private sector have made improvements to address foreign cyber and influence efforts that target our electoral process. However, there still needs to be more public transparency about foreign disinformation.

We need to close any gaps to stop any foreign disinformation about the 2020 election and disinformation in future elections. We cannot allow the Russians or other foreign actors to try to delegitimize election results or exacerbate political divisions any further.

Question 9. What more could be done to maximize transparency with the public about suspected foreign malign activity?

Answer. As noted in our response to Question No. 2, Google is committed to transparency in our business practices, including our efforts to stop foreign disinformation and coordinated influence operations. In order to increase transparency about the threats we see to our platforms, we regularly release reports that detail how we protect our platforms, enforce our policies, and review content. For instance, our publicly accessible Transparency Report (*https://transparencyreport.google.com*) details how we respond to removal requests from governments around the world, including those relating to national security; explains how our systems and policies are actively at work identifying and removing content in violation of our YouTube Community Guidelines; and contains information about election advertising content, including who is buying election ads on our platforms and how much money is being spent.

Additionally, when we find attempts to conduct coordinated influence operations on our platforms, we work with our Trust and Safety teams to swiftly remove such content from our platforms and terminate these actors' accounts. We take steps to prevent possible future attempts by the same actors, and routinely exchange information and share our findings with law enforcement and others in the industry. We share more information about these actions on our Threat Analysis Group blog (*https://blog.google/threat-analysis-group/*), which provides information about actions we take against accounts that we attribute to coordinated influence campaigns, both foreign and domestic. Moreover, if we suspect that users are subject to government-sponsored attacks, we warn them. In April 2020 alone, for example, we sent 1,755 warnings to users whose accounts were targets of government-backed attackers. For more information, please see *https://blog.google/threat-analysis-group/updates-about-government-backed-hacking-and-disinformation/*.

As threats evolve, we will continue to adapt to understand and prevent new attempts to misuse our platforms, and will continue to expand our use of cutting-edge technology to protect our users. We also will build upon our transparency efforts in the future, as they are an important component of ensuring an informed public dialogue about the role that our services play in society.

Question 10. How could you share more information about foreign disinformation threats among the private sector tech community and among social media platforms and with smaller companies?

Answer. Foreign disinformation is an issue we take very seriously. Managing information quality and content moderation across our products and services requires

significant resources and effort. As noted in our response to Question No. 3, in the last year, we have spent at least $1 billion on content moderation systems and processes, and more than 20,000 people have worked in a variety of roles to help enforce our policies and moderate content.

The speed at which content is created and shared, and the sophisticated efforts of bad actors who wish to cause harm, compound the challenge. Fortunately, we are not alone. As noted in our response to Question No. 1, we collaborate with technology partners to research and address misinformation and have a long-established policy of routinely sharing threat information with our peers and working with them to better protect the collective digital ecosystem.

For instance, we publicly share data in places like our Transparency Report (*https://transparencyreport.google.com/*), including data on government removal requests, as well as information about political advertising, such as who is buying election ads on our platforms and how much money is being spent. We make this data available for public research to all who are interested in learning or using it to conduct research or improve their content moderation efforts, including the tech community, social media platforms, and smaller companies.

We also have collaborated with industry partners to prevent terrorists and violent extremists from exploiting our platforms. In 2017, YouTube, Facebook, Microsoft, and Twitter founded the Global Internet Forum to Counter Terrorism (GIFCT) as a group of companies, dedicated to disrupting terrorist abuse of members' digital platforms. Among other important initiatives, GIFCT allows participating companies and organizations to submit hashes, or "digital fingerprints," of identified terrorist and violent extremist content to a database so that it can be swiftly removed from all participating platforms. By sharing best practices and collaborating on cross-platform tools, we have been able to bring new members to GIFCT and engage more than one hundred smaller technology companies through workshops around the world. For more information, please see *https://gifct.org/*.

As noted in our response to Question No. 7, we also collaboratively identify violative content through our YouTube Trusted Flagger program, which helps provide robust tools for individuals, government agencies, and NGOs that are particularly effective at notifying YouTube of content that violates our Community Guidelines. For more information, please see *https://support.google.com/youtube/answer/7554338*. The program provides these partners with a bulk-flagging tool and provides a channel for ongoing discussion and feedback about YouTube's approach to various content areas. The program is part of a network of over 180 academics, government partners, and NGOs that bring valuable expertise to our enforcement systems.

We also commission or partner with organizations specialized in tracking and documenting the work of threat actors who seek to target our products and services around the world. We typically do not share much information about these partnerships in order to protect these companies and their employees from the threat actors they monitor. Some examples of this work are public, such as our work with FireEye, a cybersecurity company, to detect a number of security incidents and influence operations.

We continue to develop and learn from these collaborations over time and seek more opportunities to develop best practices jointly with partners of all sizes to help people understand what they see online and to support the creation of quality content.

Question 11. What should the U.S. Government be doing to promote information sharing on threats and to increase lawful data-sharing about suspected foreign malign activity?

Answer. As noted in our response to Question No. 6, we recognize that U.S. government agencies face significant challenges in protecting the public against suspected foreign influence operations. That is why we work cooperatively with the government and law enforcement and provide information about our efforts in places like our Transparency Report (*https://transparencyreport.google.com/government-removals/overview*), which details how we work with government agencies around the world on removal requests, including those relating to national security. Google has a dedicated team that receives and responds to a high volume of requests for assistance from the government. We have developed a process specifically for these requests, so that Google can respond while also appropriately narrowing the scope of data disclosed. We work hard to protect the privacy of our users, while also supporting the important work of law enforcement.

Additionally, when we find attempts to conduct coordinated influence operations on our platforms, we work with our Trust and Safety teams to swiftly remove such content from our platforms and terminate these actors' accounts. We also routinely exchange information and share our findings with government agencies, and take steps to prevent possible future attempts by the same actors. For example, in Octo-

ber 2020, the U.S. Department of Justice acknowledged Google's contributions to the fight against Iranian influence operations, in announcing the seizure of 92 domain names used by Iran's Islamic Revolutionary Guard Corps to engage in a global disinformation campaign targeting the U.S. and other countries. For more information, please see *https://www.justice.gov/usao-ndca/pr/united-states-seizes-domain-names-used-iran-s-islamic-revolutionary-guard-corps.*

We are also focused on working with the government on identifying cyber threats. Google Cloud is working closely with the Defense Innovation Unit ("DIU") within the U.S. Department of Defense to build a secure cloud management solution to detect, protect against, and respond to cyber threats worldwide. We are honored to partner with DIU on this critical initiative to protect its network from bad actors that pose threats to our national security. For more information, please see *https://cloud.google.com/press-releases/2020/0520/defense-innovation-unit.*

We're committed to stopping this type of abuse, and working closely with the government and law enforcement on how we can help promote election integrity and user security.

Rohingya/Myanmar. In 2018, Facebook was weaponized against to whip up hate against the Muslim minority—the Rohingya. Myanmar held a general election last month. Prior to that election, there were concerns about the integrity of that election.

Question 12. What did you do and how are you continuing to make sure social media is not abused by any foreign or domestic actors to distort the electoral process in Myanmar and other countries?

Answer. We take very seriously any attempts to use our platforms to spread election misinformation, and have taken significant efforts to combat such activity across our platforms. This includes, for instance, ranking algorithms in Search that prioritize authoritative sources. Our Search algorithm ranks pages to provide the most useful and relevant information by matching search terms against available web pages and looking at factors like the number of times the words appear and freshness of the page.

Additionally, as noted in our response to Question No. 5, on any given day, Google's Threat Analysis Group is tracking more than 270 targeted or government-backed attacker groups from more than 50 countries. When we find attempts to conduct coordinated influence operations on our platforms, we work with our Trust and Safety teams to swiftly remove such content from our platforms and terminate these actors' accounts. We take steps to prevent possible future attempts by the same actors, and routinely exchange information and share our findings with others in the industry.

We also actively work to provide users with more information about the content they are seeing to allow them to make educated choices. On YouTube, Our Community Guidelines (*https://www.youtube.com/howyoutubeworks/policies/community-guidelines/*) prohibit spam, scams, or other manipulated media, coordinated influence operations, and any content that seeks to incite violence. Additionally, on Google News, we mark up links with labels that help users understand what they are about to read—whether it is local content, an op-ed, or an in-depth piece, and encourage them to be thoughtful about the content they view. Publishers who review third-party claims or rumors can showcase their work on Google News and in Google Search through fact-check labels. People come across these fact checks billions of times per year. For more information, please see *https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/*.

We also have increased transparency around news sources on YouTube, including disclosure of government funding. When a news channel on YouTube receives government funding, we make that fact clear by including an information panel under each of that channel's videos. There have been billions of impressions of information panels on YouTube around the world since June 2018. For more information, please see *https://support.google.com/youtube/answer/7630512,* and *https://blog.youtube/news-and-events/greater-transparency-for-users-around.* And, as discussed in response to Question No. 3, we also have taken a tougher stance on removing hateful and supremacist content and have reduced borderline content by reducing recommendations of content that comes close to violating our guidelines.

We are proud of these processes that help protect against abuse and manipulation across our products and that help ensure the integrity and transparency of elections all over the world.

Impact of S. 4534. As you are aware, Chairman Wicker and two of our Republican colleagues have offered legislation to amend Section 230 to address, among other issues, what they call "repeated instances of censorship targeting conservative voices."

That legislation would make significant changes to how Section 230 works, including limiting the categories of content that Section 230 immunity would cover and making the legal standard for removal of content more stringent. Critics of the Chairman's bill, S. 4534, suggest that these changes would inhibit companies' ability to remove false or harmful content from their platforms.

Question 13. I would like you to respond yes or no as to whether you believe that bills like the Chairman's would make it more difficult for Google to remove the following types of content—

- Bullying?
- Election disinformation?
- Misinformation or disinformation related to COVID–19?
- Foreign interference in U.S. elections?
- Efforts to engage in platform manipulation?
- Hate speech?
- Offensive content directed at vulnerable communities or other dehumanizing content?

Answer. Section 230 safeguards open access to information and free expression online. Instead of overblocking speech, the law supports platforms' ability to responsibly manage content. In this way, Section 230 is one of the foundational laws that has enabled America's technology leadership and success in the Internet sector. Changes to Section 230 could potentially make it more difficult to moderate all of the types of content listed above, and others.

Millions of small and large platforms and websites across the Internet rely on Section 230 to keep users safe by addressing harmful content and to promote free expression. Changes to Section 230 could negatively impact our ability to remove harmful content of all types and could make our services less useful and safe. We are concerned that changes could jeopardize removals of terrorist content, spam, malware, scams, misinformation, manipulated media, hate speech, content harmful to children, and other objectionable content. This is especially important as threats to our platforms and our users are ever-evolving, and the nature of the content we see is always changing.

We agree with the goal of Ranking Member Wicker's proposal—protecting free expression online. And we also agree that any proposal should address content that promotes self-harm, promotes terrorism, or is unlawful. We recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Combating "Garbage" Content. Santa Clara University Law Professor Eric Goldman, a leading scholar on Section 230, has argued that the Online Freedom and Viewpoint Diversity Act (S. 4534) wants Internet services to act as "passive" receptacles for users' content rather than content curators or screeners of "lawful but awful" third-party content.

He argues that the bill would be counterproductive because we need less of what he calls "garbage" content on the Internet, not more. Section 230 lets Internet services figure out the best ways to combat online trolls, and many services have innovated and invested more in improving their content moderation functions over the past few years.

Professor Goldman specifically points out that the bill would make it more difficult for social media companies to remove "junk science/conspiracy theories, like anti-vax content or quack COVID19 cures."

Question 14. Would S. 4534—and similar bills—hurt efforts by Google to combat online trolls and to fight what Professor Goldman calls "lawful but awful. . .garbage" content?

Answer. Yes, we believe that Section 230 strikes the appropriate balance that facilitates making more content and diverse points of view available than ever before in history, all while ensuring Internet companies can keep their platforms safe and secure for our users. Google, and millions of small and large platforms and websites across the internet, rely on Section 230 to keep users safe by addressing harmful content and to promote free expression. As noted in our response to Question No. 13, Section 230 supports our and other platforms' ability to curate content to protect users—and changes to Section 230 could jeopardize removals of terrorist content, spam, malware, scams, misinformation, manipulated media, hate speech, and content harmful to children, among other things.

The ability to remove harmful but not necessarily illegal content has been particularly important during COVID–19. In just one week, we saw 18 million malware

and phishing e-mails related to the coronavirus and more than 240 million COVID-related spam messages. Since February 2020, we've removed over 600,000 YouTube videos with dangerous or misleading coronavirus information and over 270 million coronavirus ads. Section 230 supports our development and enforcement of content rules that ensure that our platforms are safe for our users. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

The FCC's Capitulation to Trump's Section 230 Strategy. The Chairman of the Federal Communications Commission, Ajit Pai, announced recently that he would heed President Trump's call to start a rulemaking to "clarify" certain terms in Section 230.

And reports suggest that the President pulled the renomination of a sitting FCC Commissioner due to his concerns about that rulemaking, replacing him with a nominee that helped develop the Administration's petition that is the foundation of this rulemaking. This capitulation to President Trump by a supposedly independent regulatory agency is appalling.

It is particularly troubling that I—and other members of this committee—have been pressing Chairman Pai to push the envelope to interpret the agency's existing statutory authority to, among other things, use the E-Rate program to close the homework gap, which has only gotten more severe as a result of remote learning, and to use the agency's existing authority to close the digital divide on Tribal lands. And we expressed serious concern about Chairman Pai's move to repeal net neutrality, which the FCC majority based upon a highly conservative reading of the agency's statutory authority.

In contrast, Chairman Pai is now willing to take an expansive view of the agency's authority when asked to support the President's pressure campaign against social media in an attempt not to fact check or label the President's posts.

Question 15. What are your views on Chairman Pai's announced rulemaking and the FCC's legal analysis of section 230? Would you agree that his approach on this issue is in tension with his repeal of the essential consumer protections afforded by the net neutrality rules?

Answer. The open Internet has grown to become an unrivaled source of choice, competition, innovation, and the free flow of information. We appreciate that different sectors have different views about the details of 'net neutrality' legislation. But everyone agrees that an open Internet has been a good thing—the question is how to best preserve it.

We believe that Section 230 strikes the appropriate balance that facilitates making available more content and diverse points of view than ever before in history, while ensuring Internet companies can keep their platforms safe and secure for our users. Our business model depends on us being a useful and trustworthy source of information for everyone and we have strong policies across our products to protect our users. Our platforms empower a wide range of people and organizations from across the political spectrum, giving them a voice and new ways to reach their audiences. Section 230 enables Google and other platforms to strike a balance between maintaining a platform for free speech and living up to our responsibility to users. We understand that these are important issues and remain committed to working with Congress on them.

Addressing Bad Actors. I have become increasingly concerned with how easy it is for bad actors to use social media platforms to achieve their ends, and how they have been too slow to stop it. For example, a video touting antimalarial drug hydroxychloroquine as a "cure" for COVID was eventually taken down this summer—but not after garnering 17 million views on Facebook.

In May, the watchdog group Tech Transparency Project concluded that white supremacist groups are "thriving" on Facebook, despite assurances that Facebook does not allow such groups on its platform.

These are obviously troubling developments, especially in light of the millions of Americans that rely on social media services. You have to do better.

That said, I am not sure that modifying Section 230 is the solution for these and other very real concerns about your industry's behavior.

Question 16. From your company's perspective, would modifying Section 230 prevent bad actors from engaging in harmful conduct?

Answer. As noted in our responses above, Section 230 strikes the appropriate balance that facilitates making available more content and diverse points of view than ever before in history, while ensuring Internet companies can keep their platforms safe and secure for our users. Changes to Section 230 could jeopardize—rather than encourage—removals of terrorist content, spam, malware, scams, misinformation,

manipulated media, hate speech, and content harmful to children. Section 230 helps Internet companies address harmful content, including user comments, and while we've always been proponents of free speech, we've also always had rules of the road and are never going to be "neutral" about harmful content. Millions of small and large platforms and websites across the Internet rely on Section 230 to both keep users safe and to promote free expression. Under existing law, Section 230s protections for online platforms already exempt all Federal criminal law. Google also has worked closely with law enforcement and organizations such as NCMEC, Thorn, and Polaris for years. We have concerns that changes to Section 230 would negatively impact our ability to remove harmful content of all types and would make our services less useful and safe. We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Question 17. What do you recommend be done to address the concerns raised by the critics of Section 230?

Answer. It is no accident that the greatest Internet companies in the world were created in the United States. Section 230 is one of the foundational laws that has enabled the U.S. to lead the Internet globally, supporting millions of jobs and billions of dollars of economic activity—so we want to be very cautious and thoughtful about potential changes.

Our platforms empower a wide range of people and organizations from across the political spectrum, giving them a voice and new ways to reach their audiences. We have always stood for protecting free expression online, and have enforced our content moderation policies consistently and impartially, and we will continue to do so. In addition, millions of small and large platforms and websites across the Internet rely on Section 230 to keep users safe by addressing harmful content and to promote free expression. Section 230 is what permits us to curate content to protect users—and changes to Section 230 could jeopardize removals of harmful content.

We are committed to working with all stakeholders to support platforms' efforts to receive complaints, implement appropriate processes, and report out—without over-prescribing the precise manner and timelines by which they do so, or causing any unintended consequences. We recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Potential Impacts of Changes to Section 230. Section 230 has been foundational to the development of the Internet of today. Most believe that absent Section 230, we would not have the massive, worldwide public forum the Internet provides.

Of course, we all understand that this forum may not be an unmitigated good, but it is equally true that the Internet is a far more vibrant place than traditional media, because of the ability of users to contribute their thoughts and content.

Question 18. How do you expect Google would react when faced with increased possibility of litigation over user-submitted content?

Answer. Section 230 is what permits us to curate content to protect users–and changes could jeopardize removals of terrorist content, spam/malware, scams, misinformation, manipulated media, and hate speech. Without Section 230, we certainly could face an increased risk of liability and litigation costs for decisions around removal of content from our platforms. For example, YouTube might face legal claims for removing videos we determine could harm or mislead users in violation of our policies. Or we might be sued for trying to protect our users from spam and malware on Gmail and Search. We have concerns that putting potentially every decision around content moderation up to judicial review would negatively impact our ability to remove harmful content of all types and would make our services less useful and safe. As reflected in our other answers, we believe that Section 230 strikes the appropriate balance that facilitates making more content and diverse points of view available than ever before in history, all while ensuring Internet companies can keep their platforms safe and secure for our users.

Moreover, millions of small and large platforms and websites across the Internet rely on Section 230 to both keep users safe and promote free expression. Changes to 230 would disproportionately impact up-and-coming platforms without the resources to police every comment or defend every litigation. This could deter the next Google or Twitter or Facebook—the liability for third party content would be too great.

We also recognize the legitimate questions raised by this Committee on Section 230 and would be pleased to continue our ongoing dialogue with Congress.

Online Disinformation. I have serious concerns about the unchecked spread of disinformation online. From false political claims to harmful health information,

each day the problem seems to get worse and worse. And I do not believe that social media companies—who make billions of dollars from ads based in part on user views of this disinformation—are giving this problem the serious attention that it deserves.

Question 19. Do you agree that Google can and should do more to stop the spread of harmful online disinformation?

Question 20. Can you commit that Google will take more aggressive steps to stop the spread of this disinformation? What specific additional actions will you take?

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 19 and 20.

Addressing misinformation is an evolving threat, and we will continue to take action to address these issues. Due to the shifting tactics of groups promoting misinformation and conspiracy theories, we've been investing in the policies, resources, and products needed to protect our users from this harmful content.

As noted in our response to Senator Peters' Question No. 1, among other things, we launched a Community Guidelines YouTube update in October on harmful conspiracy theories (*https://blog.youtube/news-and-events/harmful-conspiracy-theories-youtube/*), which expanded our hate speech and harassment policies to prohibit content that targets an individual or group with conspiracy theories that have been used to justify real-world violence. For example, content such as conspiracy theories saying individuals or groups are evil, corrupt, or malicious based on protected attributes (*e.g.,* age, race, religion, etc.), or hateful supremacist propaganda, including the recruitment of new members or requests for financial support for their ideology, all violate our hate speech policy (*https://support.google.com/youtube/answer/2801939*) and are subject to removal as such.

As detailed in our October update, these are not the first steps we have taken to limit the reach of harmful misinformation. Nearly two years ago, we updated our YouTube recommendations system, including reducing recommendations of borderline content and content that could misinform users in harmful ways—such as videos promoting a phony miracle cure for a serious illness, claiming the earth is flat, or making blatantly false claims about historic events like 9/11. This resulted in a 70 percent drop in views coming from our search and discovery systems. Further, when we looked at QAnon content, we saw the number of views that come from non-subscribed recommendations to prominent Q-related channels dropped by over 80 percent from January 2019 to October 2020. For more information on these policies and enforcement actions, please see *https://blog.youtube/news-and-events/harmful-conspiracy-theories-youtube/, https://blog.youtube/news-and-events/continuing-our-work-to-improve/,* and *https://blog.youtube/news-and-events/our-ongoing-work-to-tackle-hate.*

In addition to removing content that violates our policies, we also reduce borderline content and raise up authoritative voices by providing users with more information about the content they are seeing to allow them to make educated choices. On YouTube, for example, there have been billions of impressions of information panels around the world since June 2018. For more information, please see *https://support.google.com/youtube/answer/9229632.* And, for over three years, we have highlighted fact checks on Search and News as a way to help people make more informed judgments about the content they encounter online. People come across these fact checks billions of times per year. For more information, please see *https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/*.

Concerning political misinformation on YouTube, since September, we've terminated over 8,000 channels and thousands of harmful and misleading elections-related videos for violating our existing policies. Over 77 percent of those removed videos were taken down before they had 100 views. And, since election day, relevant fact check information panels from third-party fact checkers were triggered over 200,000 times above relevant election-related search results, including for voter fraud narratives such as "Dominion voting machines" and "Michigan recount." For additional information, please see *https://blog.youtube/news-and-events/supporting-the-2020-us-election/*.

Our Ads policies are similarly designed to ensure a safe and positive experience for our users. For example, under our existing misrepresentation policy, we do not allow ads to run or content to monetize that promotes medically unsubstantiated claims related to COVID cures (*https://support.google.com/google-ads/answer/9811449*). Since February 2020, we've blocked or removed over 270 million coronavirus related ads across all Google advertising platforms.

As noted in our response to Question No. 8 and Senator Blumenthal's Question No. 1, all ads, including political ads, must comply with our publicly-available Ads

policies (*https://support.google.com/adspolicy/answer/6008942*), which prohibit, among other things, dangerous or derogatory content; content that is illegal, promotes illegal activity, or infringes on the legal rights of others; and content that misrepresents the owner's origin or purpose. We put significant effort into curbing harmful misinformation on our ads platform, including prohibiting content that makes claims that are demonstrably false and could significantly undermine participation or trust in an electoral or democratic process. Further, given the unprecedented amount of votes that were counted after election day this year, we also implemented a sensitive event policy for political ads after the polls closed on November 3, 2020 (*https://support.google.com/adspolicy/answer/10122500*), which prohibited advertisers from running ads referencing candidates, the election, or its outcome. Additionally, all advertisers who run U.S. election ads must first be verified in order to protect the integrity of the election ads that run on our platform. We are serious about enforcing these policies, and we block and remove ads that we find to be violative. For more information, please see our political content advertising policies, *https://support.google.com/adspolicy/answer/6014595*.

The openness of our platforms has helped creativity and access to information thrive. It's our responsibility to protect that, and prevent our platforms from being used to spread dangerous misinformation. We are committed to taking the steps needed to live up to this responsibility today, tomorrow, and in the years to come.

Question 21. Although Section 230 says that Google cannot be legally considered a publisher, Google does sell a lot of advertising and produces a lot of information. In fact, I recently released a report on the state of local journalism, and Google alone earns more advertising revenue than all of the newspaper publishers in the U.S. combined. Google is also a dominant provider of technology and services that enable online advertising and many online publishers are reliant on these services to earn revenue. What many people probably don't know is that Google also buys a lot of online advertising inventory. But you don't actually use most of that advertising inventory. Instead, you re-sell it. Of all the advertising that Google buys, what percentage do you then re-sell? Over the last two years, when you bought advertising inventory from other publishers and then re-sold it to advertisers, what was your average gross profit margin? When a competing publisher is also a customer of your ad services business, do you believe you have an obligation to help them maximize the revenue they earn from their advertising inventory?

Answer. We work to sustain a healthy, vibrant, and financially sound Internet ecosystem and build tools to help publishers make money from ad inventory, increase return on ad spend, and reach consumers at a global scale. We also strive to help users find the content they need quickly while having their privacy expectations respected and without being overwhelmed with annoying or intrusive ads. Google designs its publisher ad technology products to enable publishers to facilitate competition and increase revenues for their ad inventory, making it even easier for them to deliver valuable, trustworthy ads and the right experiences for consumers across devices and channels. We are proud of how Google's digital advertising tools help publishers generate revenue—in 2018 alone, we paid our publisher partners more than $14 billion.

When ads flow through both our buy-side and sell-side services, publishers receive most of the revenue. As detailed in our June 2020 blogpost (*https://blog.google/products/admanager/display-buying-share-revenue-publishers*), in 2019, when marketers used Google Ads or Display & Video to buy display ads on Google Ad Manager, publishers kept over 69 percent of the revenue generated. When publishers use our Ad Manager platform to sell ads directly to advertisers, they keep even more of the revenue—nearly 100 percent—paying only a minimal ad serving fee. We recently analyzed the revenue data of the top 100 news publishers globally with the highest programmatic revenue generated in Ad Manager and found that, on average, news publishers keep over 95 percent of the digital advertising revenue they generate when they use Ad Manager to show ads on their websites. For additional information, please see *https://blog.google/products/admanager/news-publishers-make-money-ad-manager*.

Google has also publicly disclosed its revenue share when publishers use AdSense (Google's ad network) to sell their ad inventory. For more information, please see our AdSense revenue share page, *https://support.google.com/adsense/answer/180195*.

Fundamentally, we're in the business of providing helpful and relevant tools and information, and we are proud that Google's investments in this space have helped publishers make money to fund their work, made it easy for businesses large and small to reach consumers, and supported the creative and diverse content we all enjoy.

Question 22. The ad server has two main roles in online advertising. First, ad servers store and make available advertising content that is transmitted to viewers. Secondly, ad servers collect metrics about advertising targeting and performance. The first function is largely a commodity service, but the second function is very strategic, since advertising data is valuable and those who collect advertising data also control how the data can be monetized. It's been reported that Google controls greater than 90 percent of the ad server market. How would you characterize Google's position in this market? Who would you describe as your strongest competitor in the ad server business?

Answer. We face formidable competition in every aspect of our business, particularly from companies that seek to connect people with online information and provide them with relevant advertising. Advertisers have lots of options when it comes to choosing an ad server, including solutions offered by Adform, AdGlare, Adslot, Addition, Amazon's Sizmek, Unilever's Celtra, Clinch, Epon, Extreme Reach, Flashtalking, Innovid, OpenX, Verizon Media, Weborama, and Zedo. Publishers likewise have many options when it comes to choosing an ad server, including solutions offered by Adform, AT&T's Xandr, Comcast's FreeWheel, PubMatic, Smart, SpotX, ironSource, Twitter's MoPub, and others. Other publishers, such as Twitter, Amazon, and Facebook, have decided to build their own in-house ad serving systems. Publishers are also able to use ad networks (without a separate ad server) to serve ads on their sites, or forgo the use of an ad server by placing an ad tag on their web page that directly connects to sell-side tools.

Additionally, the lines between different advertising technology tools have blurred in recent years. For example, many demand-side platforms (DSPs), supply-side platforms (SSPs), and ad networks include ad serving tools, which publishers and advertisers use interchangeably. Industry reports suggest that the average publisher uses four to six SSPs. Similarly, the average advertiser uses three DSPs simultaneously (and also buys ad inventory directly from publishers like Facebook, Twitter, LinkedIn, and Snapchat who have their own buying platforms). Thus, publishers and advertisers don't have to use Google's tools, and even when they do, they can either switch to competing products or even use them simultaneously. (See May 5, 2020 AdExchanger article, *https://www.adexchanger.com/platforms/google-ad-manager-policy-changes-dont-hurt-publishers-according-to-advertiser-perceptions/;* July 13, 2020 AdExchanger article, *https://www.adexchanger.com/online-advertising/google-reclaims-the-dsp-crown-in-latest-advertiser-perceptions-report/*.)

Additional information regarding the competition we face in our advertising business is available in places like our blogs, including *https://www.blog.google/technology/ads/ad-tech-industry-crowded-and-competitive/,* as well as our Forms 10–K and 10–Q, available at *https://abc.xyz/investor/*. This competitive ecosystem supports the availability of free-to-consumer content online, which has a broader, positive impact on consumers.

Question 23. What percentage of users clicked on the "show me" button on YouTube's election label on disputed content? Please provide statistics on the efficacy of this labelling, such as the average proportion of a video that is watched before and after labelling.

Answer. While only a small portion of watch time is election-related content, YouTube continues to be an important source of election news. That is why we show information panels linking both to Google's election results feature, which sources election results from The Associated Press, and to the Cybersecurity & Infrastructure Security Agency's (CISA) "Rumor Control" page for debunking election integrity misinformation, alongside these and over 200,000 other election-related videos. Collectively, these information panels have been shown over 4.5 billion times. And, since election day, relevant fact check information panels from third-party fact checkers were triggered over 200,000 times above relevant election-related search results, including for voter fraud narratives such as "Dominion voting machines" and "Michigan recount." For additional information, please see *https://blog.youtube/news-and-events/supporting-the-2020-us-election/,* and *https://blog.youtube/news-and-events/our-approach-to-election-day-on-youtube/*.

Question 24. Third parties assessed YouTube as more opaque than Facebook, Instagram, and Twitter about its policies on election misinformation for months preceding the election and weeks after it. For example, an Election Integrity Partnership review of platform policies regarding election delegitimization found on October 28th that YouTube still had non-existent policies in almost every area, as opposed to "comprehensive" ratings for Twitter and Facebook in nearly every category. Please describe the reasoning behind the lack of specificity in your election delegitimization policies leading up to the election. On December 9th, YouTube updated its policies, but only after a critical period of potential civil unrest in which

civil servants in swing states were being threatened due to election misinformation. Does YouTube plan to establish new mechanisms of policy review that would allow it to respond more effectively to such critical moments?

Answer. We've always had rules of the road for YouTube that we enforce in order to protect our users. We take a holistic approach to disinformation through several policies in our Community Guidelines (*https://www.youtube.com/howyoutube works/policies/community-guidelines/*), which explain what types of content and behaviors are not allowed, and the process by which content and users may be removed from the service. However, given the ever-evolving threats to our platforms and users, and that the nature of the content we see is always changing, it would be ineffective and impractical to attempt to address every possible harm in advance in our YouTube Community Guidelines. Instead, our policies include prohibitions against spam, deceptive practices, scams, hate speech, harassment, and harmful manipulated media. For example, our deceptive practices policy (*https://support.google.com/youtube/answer/2801973*) prohibits content that deliberately seeks to spread disinformation that could suppress voting or otherwise interfere with democratic or civic processes, such as demonstrably false content making claims of different voting days for different demographics.

Google has long had numerous systems in place, both automated and manual, to detect and address problematic content in violation of these policies. Our machine learning systems are faster and more effective than ever before and are helping our human review teams remove content with speed and volume that could not be achieved with people alone. For example, in the third quarter of 2020, more than 7.8 million videos were removed from YouTube for violating our community guidelines. Ninety-four percent of these videos were first flagged by machines rather than humans. Of those detected by machines, over 45 percent never received a single view, and just over 80 percent received fewer than 10 views. In the same period, YouTube removed more than 1.1 billion comments, 99 percent of which were detected automatically. For more information, please see our YouTube Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/removals*).

Additionally, as noted in our response to Question No. 5, since September, we've terminated over 8,000 channels and thousands of harmful and misleading elections-related videos for violating our existing policies. Over 77 percent of those removed videos were taken down before they had 100 views. And, since election day, relevant fact check information panels from third-party fact checkers were triggered over 200,000 times above relevant election-related search results, including for voter fraud narratives such as "Dominion voting machines" and "Michigan recount." In addition, we have mechanisms in place to reduce the recommendation of content that brushes right up against our policy line, including harmful misinformation. Limiting the reach of borderline content and prominently surfacing authoritative information are important ways we protect people from problematic content that doesn't violate our Community Guidelines. Since making changes to our recommendations systems, we've seen a substantial drop in borderline content and misinformation. Over 70 percent of recommendations on election-related topics came from authoritative news sources and the top recommended videos and channels for election-related content were primarily authoritative news. In fact, the top 10 authoritative news channels were recommended over 14 times more than the top 10 non-authoritative channels on election-related content. For additional information, please see *https://blog.youtube/news-and-events/supporting-the-2020-us-election/*.

Our teams work hard to ensure we are striking a balance between allowing for a broad range of political speech and making sure our platform isn't abused to incite real-world harm or broadly spread harmful misinformation. We welcome ongoing debate and discussion and will keep engaging with experts, researchers, and organizations to ensure that our policies and products are meeting that goal. And as always, we will apply learnings from this election to our ongoing efforts to protect the integrity of elections around the world.

Question 25. Mr. Pichai, how much money does Google earn in total from its ad tech businesses in the United States? What percent of those funds come from newspapers and publishers using your ad tech services?

Question 26. Mr. Pichai, could you also provide the committee a breakdown of your yearly revenues of your various ad-tech businesses?

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 25 and 26.

We generate advertising revenues primarily by delivering advertising on Google properties. Google properties revenues consist primarily of advertising revenues generated on Google.com, the Google Search app, and other Google owned and operated

properties like Gmail, Google Maps, Google Play, and YouTube. We also generate advertising revenues through the ad tech products included in our Google Network Member properties (Google Network Members includes, but is not limited to, products such as AdMob, AdSense, Google Ad Manager, Display & Video 360, Campaign Manager, etc.).

We generate most of our revenue from advertising, and ad tech is a portion of our Google Network Member advertising revenue. Our advertising revenues are disclosed on a quarterly basis in our Forms 10–K and 10–Q available at *https://abc.xyz/investor/*. In 2019, Alphabet generated total revenues of $161.9 billion, $74.8 billion of which was from the United States. Globally, our advertising revenues for 2019 were $134.8 billion, or 83 percent of total revenues. In 2019, we had gross advertising revenues of $21.5 billion globally from Google Network Member properties which includes, but is not limited to, our ad tech products, or 16 percent of total advertising gross revenues and 13 percent of total gross revenues, most of which is paid out to publishers. As noted in our response to Question No. 21, we paid more than $14 billion to our publishing partners in 2018. And when publishers use our Ad Manager platform to sell ads directly to advertisers, they keep even more of the revenue—nearly 100 percent—paying only a minimal ad serving fee. We recently analyzed the revenue data of the top 100 news publishers globally with the highest programmatic revenue generated in Ad Manager and found that, on average, news publishers keep over 95 percent of the digital advertising revenue they generate when they use Ad Manager to show ads on their websites. For additional information, please see *https://blog.google/products/admanager/news-publishers-make-money-ad-manager*.

Ad tech is a complex, highly-competitive ecosystem, and we believe it's working to the benefit of publishers, advertisers, and users, and we hope to continue succeeding by building the best products for our users.

Question 27. Mr. Pichai, how much of the $1 billion dollar pledge to publishers are you reserving for U.S. publishers? Please explain your methodology for paying publishers. How are you determining who to pay in the U.S. and internationally? Will you provide clear information to the marketplace that explains your methodology? Will you list all of the publishers you pay?

Answer. We believe strongly in connecting our users to high quality news content, and, in October 2020, we announced an initial $1 billion investment over the next three years in partnerships with news publishers and the future of news. This financial commitment—our biggest to date—will pay publishers to create and curate high-quality content for a different kind of online news experience.

Google News Showcase is a new product made up of story panels that give participating publishers the ability to package the stories that appear within Google's news products, providing deeper storytelling and more context through features like timelines, bullets, and related articles. News Showcase will also provide limited access to paywalled content in partnership with select news publishers. We've signed partnerships for News Showcase with nearly 400 leading publications across Germany, Brazil, Argentina, Canada, France, the U.K., and Australia, some of which are identified in our October 2020 blogpost (*https://blog.google/outreach-initiatives/google-news-initiative/google-news-showcase*) and December 2020 blogpost (*https://blog.google/products/news/google-news-showcase-expands*). Publishers are selected on a country-by-country basis, with publishers that have established audiences and serve a community—like local news publishers and print newspapers—receiving priority. Financial details of the licensing deals vary depending on the volume and type of content each publisher provides.

Both News Showcase and our financial investment—which will extend beyond the initial three years—are focused on contributing to the overall sustainability of our news partners around the world. We are proud that this commitment will build on our long-term support of news publishers and the future of news, and help journalism in the 21st century not just survive, but thrive.

Question 28. Mr. Pichai, my staff has been provided reports that some of your proposed agreements with news publishers around the world require the publishers to promise not to sue Google. Under these agreements news publishers would be barred, for instance, from taking legal action against Google regarding content aggregation—or they would forfeit the entire financial agreement. Are these reports accurate? Will you commit to not including covenants not to sue in agreements with American publishers?

Answer. It is not uncommon for companies to include waivers of claims in contracts or agreements to settle pending or threatened litigation. Google may have entered into these types of contracts over the years and to the extent that such terms

exist for either or both parties, they are the product of good faith negotiations by sophisticated parties.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. AMY KLOBUCHAR TO SUNDAR PICHAI

Health Data Privacy. New technologies have made it easier for people to monitor their health, but health tracking apps, wearable technology devices, and home DNA testing kits have given companies access to consumers' private health data—which is not protected under existing privacy law. In June 2019, I introduced legislation with Senator Murkowski to require the Department of Health and Human Services (HHS) to address this issue.

Question 1. Do you agree that new privacy regulations that complement existing Federal health privacy laws are required to keep up with advances in technology to protect sensitive health data?

Question 2. Do you agree that consumers should have heightened privacy for their sensitive health data, and should know where this type of data is being shared?

Answer. We support Federal comprehensive privacy legislation, and we would welcome the chance to work with you on it. We also published a framework drawing from established privacy frameworks and our practical experience. For more information, please see *https://services.google.com/fh/files/blogs/google_framework_responsible_data_protection_regulation.pdf.*

In the last year, we also have introduced new ways for users to protect their personal data, including by making our controls easier to use, continuing our advances in privacy enhancing technologies like differential privacy, and providing users options to have Google automatically delete personal data like Location History, searches, and other activity. In our work with healthcare providers helping them to deliver better care to patients in a privacy-protective way across the U.S., we implement controls designed to adhere to HIPAA and other existing data privacy and security regulations where applicable to protect patient data, as well as in accordance with our Privacy Principles (*https://safety.google/principles/*) and our Privacy Policy (*https://policies.google.com/privacy*). We recognize the legitimate questions raised by this Committee on healthcare data and privacy, and would be pleased to continue our ongoing dialogue with Congress. For more information on our approach to privacy, please see *https://health.google/* and *https://blog.google/products/admanager/additional-steps-safeguard-user-privacy/*.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. RICHARD BLUMENTHAL TO SUNDAR PICHAI

For the following questions, please provide information about your firm's content moderation decisions related to election misinformation and civic integrity covering the 2020 election period.

Question 1. Please describe what processes were used to make decisions about labeling or taking down organic and paid content related to elections or civic integrity.

Elections are a critical part of the democratic process, and we are committed to helping voters find relevant, helpful, and accurate information.

Answer. Regarding our processes for paid content, all ads, including political ads, must comply with our publicly-available Ads policies (*https://support.google.com/adspolicy/answer/6008942*), under which candidates, campaigns, and other types of political spenders are treated the same as all other advertisers. These policies prohibit, among other things, dangerous or derogatory content; content that is illegal, promotes illegal activity, or infringes on the legal rights of others; and content that misrepresents the owner's origin or purpose. We put significant effort into curbing harmful misinformation on our ads platform, including prohibiting content that makes claims that are demonstrably false and could significantly undermine participation or trust in an electoral or democratic process. For more information, please see our Misrepresentation policy, *https://support.google.com/adspolicy/answer/6020955.* We also have zero tolerance for ads that employ voter suppression tactics or undermine participation in elections—when we find those ads, we take them down. Given the unprecedented amount of votes that were counted after election day this year, we also implemented a sensitive event policy for political ads after the polls closed on November 3, 2020 (*https://support.google.com/adspolicy/answer/10122500*), which prohibited advertisers from running ads referencing candidates, the election, or its outcome. Additionally, all advertisers who run U.S. election-related ads must first be verified in order to protect the integrity of the election

ads that run on our platform. We're serious about enforcing these policies, and we block and remove ads that we find to be violative. For more information, please see our political content advertising policies, *https://support.google.com/adspolicy/answer/6014595.*

We also actively work to provide users with more information about the content they are seeing to allow them to make educated choices. On YouTube, Our Community Guidelines (*https://www.youtube.com/howyoutubeworks/policies/community-guidelines/*) prohibit spam, scams, or other manipulated media, coordinated influence operations, and any content that seeks to incite violence. Since September, we've terminated over 8,000 channels and thousands of harmful and misleading elections-related videos for violating our existing policies. Over 77 percent of those removed videos were taken down before they had 100 views. We also work to make sure that the line between what is removed and what is allowed is drawn in the right place. Our policies prohibit misleading viewers about where and how to vote. We also disallow content alleging that widespread fraud or errors changed the outcome of a historical U.S. Presidential election. In some cases, however, that has meant allowing controversial views on the outcome or process of counting votes of a current election as election officials have worked to finalize counts.

Furthermore, as December 8, 2020 was the safe harbor deadline for the U.S. Presidential election, and enough states have certified their election results to determine a President-elect, YouTube will remove any piece of content uploaded anytime after December 8 that misleads people by alleging that widespread fraud or errors changed the outcome of the 2020 U.S. Presidential election, in line with our approach towards historical U.S. Presidential elections. For example, we will remove videos claiming that a Presidential candidate won the election due to widespread software glitches or counting errors. As always, news coverage and commentary on these issues can remain on our site if there's sufficient education, documentary, scientific, or artistic context, as described here, *https://blog.youtube/inside-youtube/look-how-we-treat-educational-documentary-scientific-and-artistic-content-youtube/*.

Our publicly accessible, searchable, and downloadable Transparency Report contains information about election ad content and spending on our platforms (*https://transparencyreport.google.com/political-ads/region/US*). The report provides information about when election ads ran, how they were targeted, how many impressions they served, and the advertiser who paid for the ads. We also describe our efforts to promote election and civic integrity in recent blogs, including *https://www.blog.google/technology/ads/update-our-political-ads-policy/, https://blog.google/outreach-initiatives/civics/following-2020-us-election-google/,* and our Threat Analysis Group's blog, *https://blog.google/threat-analysis-group/*.

Our processes relating to organic content on Search apply regardless of whether or not the content relates to elections. For over three years, we have highlighted fact checks on Search as a way to help people make more informed judgments about the content they encounter online. For more information, please see *https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/*. In terms of blocking or removing content in Search results, we only remove content in limited circumstances, including based on our legal obligations, copyright, webmaster guidelines, spam, and sensitive personal information like government IDs. Please see, for example, our policies relating to removals for legal obligations (*https://support.google.com/websearch/answer/9673730*); webmaster guidelines (*https://developers.google.com/search/docs/advanced/guidelines/webmaster-guidelines*); voluntarily removal policies (*https://support.google.com/websearch/answer/3143948*); and policies concerning removals for copyright infringement (*https://support.google.com/transparencyreport/answer/7347743*). In these cases, content that is reported to us or that we identify to be in violation of our policies is filtered from our results to adhere to the law and those policies. Additionally, some of our Search features, such as featured snippets, have policies specifying what is eligible to appear (*https://support.google.com/websearch/answer/9351707*). All of these policies are intended to ensure we are not surfacing shocking, offensive, hateful, violent, dangerous, harmful, or similarly problematic material.

We are proud of these processes that help protect against abuse and manipulation across our products and help ensure the integrity and transparency of our Nation's elections.

Question 2. How many posts were reported or identified as potentially containing election misinformation or violations of civic integrity policies?

Question 3. How many posts had enforcement action taken for containing election misinformation or violations of civic integrity policies?

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 2 and 3.

Election and civic integrity is an issue that we take very seriously, and we have many different policies and processes to combat election-related misinformation and related violative content. Notably, such content may be removed for violations of a range of policies across our products, such as our misrepresentation policies, our dangerous or derogatory content policies, or our violent or graphic content policies. We regularly release reports that detail how we enforce our policies, including information on the number of removals and the reasons for those removals. For example, our YouTube Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/removals*) contains information on the volume of videos removed by YouTube, by reason the video was removed. The removal reasons correspond to our YouTube Community Guidelines (*https://www.youtube.com/howyoutubeworks/policies/community-guidelines/*).

In the third quarter of 2020, over 7.8 million videos were removed by YouTube for violating Community Guidelines, including violations of our policies regarding spam, misleading content, and scams (25.5 percent), violent or graphic content (14.2 percent), promotion of violence or violent extremism (2.5 percent), harmful or dangerous content (2.5 percent), and hateful or abusive content (1.1 percent). Our annual Bad Ads report (*https://www.blog.google/products/ads/stopping-bad-ads-to-protect-users/*) provides detailed information regarding enforcement actions we've taken to protect our ads ecosystem. We also include additional information on these enforcement actions in places like our Google Transparency Report (*https://transparencyreport.google.com/*), which shares detailed information about how the policies and actions of governments and corporations affect privacy, security, and access to information. Moreover, our quarterly Threat Analysis Group Bulletins (Q4 update here: *https://blog.google/threat-analysis-group/tag-bulletin-q4-2020/*) contains information about actions we take against accounts that we attribute to coordinated influence campaigns. We also recently reported in a blog post on "Supporting the 2020 U.S. election" (*https://blog.youtube/news-and-events/supporting-the-2020-us-election/*) that since September, we've terminated over 8,000 channels and thousands of harmful and misleading elections-related videos for violating our existing policies. Over 77 percent of those removed videos were taken down before they had 100 views.

Question 4. Who did your firm consult to draft and implement election misinformation and civic integrity policies?

Answer. We consult with a diverse set of external and internal stakeholders during policy development, which can include expert input, user feedback, and regulatory guidance. This collaborative approach taps into multiple areas of expertise within and beyond our company and is typically driven by our Trust and Safety teams, whose mission includes tackling online abuse by developing and enforcing the policies that keep our products safe and reliable. These teams include product specialists, engineers, lawyers, data scientists, and others who work together around the world and with a network of in-house and external safety and subject matter experts.

Where appropriate, these teams consult in-depth studies or research by a mix of organizations, academics, universities, or think tanks who have topical expertise in specific matters. These analysts study the evolving tactics deployed by bad actors, trends observed on other platforms, and emerging cultural issues that require further observation. Further, we engage in conversations with regulators around the world, and their perspectives and concerns directly inform our policy process.

Question 5. Who made final decisions about labeling or taking down a post related to election misinformation or civic integrity? Who did that person or those persons consult?

Answer. We enforce our content policies at scale and take tens of millions of actions every day against content that violates policies for one or more of our products. To enforce our policies at scale, we use a combination of reviewers and AI moderation systems.

Content moderation at Google and YouTube is primarily managed by Trust and Safety teams across the company. These teams are made up of engineers, content reviewers, and others who work across Google to address content that violates any of our policies. These teams also work with our legal and public policy teams, and oversee the vendors we hire to help us scale our content moderation efforts, as well as provide the native language expertise and the 24-hour coverage required of a global platform. Google employs review teams across many offices globally and across the U.S. to ensure that we have a diverse set of reviewers who are reviewing publisher sites, apps, and content.

Question 6. Does a different or specialize process exist for content from Presidential candidates, and if so, how does that process for review differ from the normal review?

Answer. As noted in our response to Question No. 1, we enforce our policies consistently, regardless of who or what is involved. Our policies apply to all users and advertisers—from voters, to politicians, to heads of state—we don't make any special exceptions.

Question 7. Based on enforcement actions taken, there a discernible difference in engagement between a labeled post and unlabeled posts? Please provide any supporting information.

Answer. On YouTube, we may provide contextual information in information panels alongside relevant topics. Such panels are displayed algorithmically based on subject matter, rather than based on a determination of whether the video contains misinformation. For example, we seek to display the same COVID information panel on all COVID-related videos. Accordingly, we cannot meaningfully compare engagement for labeled or unlabeled videos with respect to a given topic. When we provide such contextual information, we do so to help connect users to authoritative content and to provide information that can be used to help them determine for themselves the trustworthiness of the content they watch. This isn't possible everywhere, but where we have it, these features let users dig deeper on a story or piece of content.

Question 8. What was the average time to add a misinformation label to a post?

Answer. On YouTube, information panels are typically applied when appropriate by automated systems shortly after a video is uploaded.

For the following questions, please provide information about your firm's content moderation decisions related hate speech, election interference, civic integrity, medical misinformation, or other harmful misinformation over the previous year.

Question 9. How many pieces of content were reported by users to the platform related to hate speech, election interference, civic integrity, and medical misinformation, broken down by category?

Question 10. How many pieces of content were automatically identified or identified by employees related to hate speech, election interference, civic integrity, and medical misinformation, broken down by category?

Question 11. Of the content reported or flagged for review, how many pieces of content were reviewed by humans?

Question 12. How many pieces of content were subject to enforcement action? Please provide a break down for each type of enforcement action taken for each category.

Answer. Because the answers to these questions are related, we have grouped together our response to Question Nos. 9 through 12.

Our responses to Question Nos. 2 and 3 above contain additional responsive information and resource links we hope are helpful. As to the volume of removals in general, as well as the volume of removals done by our machine learning systems versus human review teams, we regularly release reports detailing this information. For example, as detailed in the YouTube Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/removals*), of the 7.8 million videos removed from YouTube in the third quarter of 2020 for violating our Community Guidelines, 94 percent of them were first flagged by machines. With respect to hate speech on YouTube, we publish detailed information about our removals on our Transparency Page (*https://transparencyreport.google.com/youtube-policy/featured-policies/hate-speech*). And, as detailed in our Bad Ads report (*https://www.blog.google/products/ads/stopping-bad-ads-to-protect-users/*), in 2019, we blocked and removed 2.7 billion bad ads, suspended nearly 1 million advertiser accounts for policy violations, and on the publisher side, terminated over 1.2 million accounts and removed ads from over 21 million web pages that are part of our publisher network for violating our policies.

Technology has helped us accelerate and scale our removal process—our sophisticated automated systems are carefully trained to quickly identify and take action against spam and violative content. Our machine learning systems are faster and more effective than ever before and are helping our human review teams remove content with speed and volume that could not be achieved with human reviewers alone. While we rely heavily on technology, reviewers also play a critical role. New forms of abuse and threats are constantly emerging that require human ingenuity to assess and develop appropriate plans for action. Our reviewers perform billions of reviews every year, working to make fair and consistent enforcement decisions in enforcing our policies and helping to build training data for machine learning models.

Question 13. For content subject to enforcement action due to violation of hate speech rules, please identify how many pieces of content targeted each type of protected category (such as race or gender) covered by your rules. Do you track this information?

Answer. As referenced in the responses above, in the third quarter of 2020 for example, over 7.8 million videos were removed by YouTube for violating Community Guidelines, and removals due to hateful or abusive content constituted 1.1 percent of the total actions taken, though we have not tracked content removal information based on demographic sub-categories such as race or gender. We publish details concerning these violations and enforcement of our hate speech policy in resources such as our YouTube Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/featured-policies/hate-speech*), which includes examples of content involving categories protected by our hate speech policy and subject to removal decisions. For additional information regarding enforcement of our hate speech policy, please see *https://blog.youtube/news-and-events/our-ongoing-work-to-tackle-hate* and *https://support.google.com/youtube/answer/6162278.*

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. EDWARD MARKEY TO SUNDAR PICHAI

Question 1. Mr. Pichai, will you commit Google and YouTube to undergo an independent civil rights audit covering such topics as potential discriminatory data uses, advertising practices, collection and use of geolocation information, and online privacy risks that disproportionately harm particular demographic populations? Will you commit to sharing the findings of this audit publicly? Please describe in detail the steps Google and YouTube will take to ensure that they implement audit recommendations in a transparent manner.

Answer. Our products are built for everyone, and we design them with extraordinary care to be a trustworthy source of information without regard to a user's demographic, socioeconomic background, or political viewpoint. Billions of people use our products to find information, and we help our users, of every background and belief, find the high-quality information they need to better understand the topics they care about.

We are also a leader in transparency concerning our privacy policies and practices, and aim to be as clear as possible to our users about how our products and policies work. We were the first platform to have a publicly-available transparency report in 2010, and since then, we have launched a number of different transparency reports to shed light on how the policies and actions of governments and corporations affect privacy, security, and access to information for our users. Our current reports cover topics such as security and privacy, content removal, political advertising on Google, and traffic and disruptions to Google. We also have a report specifically focused on YouTube community guidelines enforcement, including data on removal by the numbers, source of first detection, views, removal reason, and country/region. For example, please see our YouTube Community Guidelines Enforcement FAQs, *https://support.google.com/transparencyreport/answer/9209072.*

In the last year, we also introduced new ways for users to protect their data, including by making our controls easier to use, continuing our advances in privacy enhancing technologies like differential privacy, and by providing users options to automatically delete data like Location History, searches, and other activity. We design all our products in accordance with our Privacy Principles (*https://safety.google/principles/*) and provide clear descriptions of how we collect and use data from users in our Privacy Policy (*https://policies.google.com/privacy*).

We are also engaged in extensive discussions with civil rights experts and leadership, and we are proud that we have civil and human rights expertise on staff, internal frameworks like our AI Principles and YouTube Community Guidelines in place, and governance structures through groups like our Responsible Innovation and Trust and Safety teams, working to help build civil and human rights considerations into our work. Our civil and human rights leads will continue to develop a structure to provide the transparency that the civil rights community needs, and we have confidence that we can demonstrate our long-term commitment to getting this right.

We will continue to approach this thoughtfully. We are always open to feedback, and will continue to provide transparency about our products and policies.

Question 2. Mr. Pichai, children and teens are a uniquely vulnerable population online, and a comprehensive Federal privacy law should provide them with heightened data privacy protections. Do you agree that Congress should prohibit online

behavioral advertising, or "targeted marketing" as defined in S.748, directed at children under the age of 13?

Answer. We support Federal comprehensive privacy legislation, and we would welcome the chance to work with you on it. At Google, we are committed to ensuring that our products are safe for children and families online and are investing significant resources in this effort. For example, we offer parental supervision options through Family Link, including the option for parents to approve all apps downloaded from Google Play. We don't serve personalized ads to children using Family Link accounts across Google products. In addition to offering YouTube Kids (*https://blog.youtube/news-and-events/youtube-kids*), even on our main YouTube platform, content that is designated as "Made for Kids" will not run personalized ads and will have certain features disabled, like comments and notifications. This is a very important issue and we're committed to continue working with Congress on it.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. GARY PETERS TO SUNDAR PICHAI

Question 1. Community standards at Google and YouTube often draw the line at specific threats of violence for the removal of content, rather than conspiracy theories that may set the predicate for radicalization and future action. When it comes to conspiracy theories and misinformation, Google and YouTube often choose not to remove content, but rather to reduce the spread and to attach warnings. What testing or other analysis have Google and YouTube done that shows your work to reduce the spread of disinformation and misinformation is effective?

Answer. Managing misinformation and harmful conspiracy theories is challenging because the content is always evolving, but we take this issue very seriously. Due to the shifting tactics of groups promoting these conspiracy theories, we've been investing in the policies, resources, and products needed to protect our users from harmful content.

Among other things, we launched a Community Guidelines YouTube update in October on harmful conspiracy theories (*https://blog.youtube/news-and-events/harmful-conspiracy-theories-youtube/*), which expanded our hate speech and harassment policies to prohibit content that targets an individual or group with conspiracy theories that have been used to justify real-world violence. For example, content such as conspiracy theories saying individuals or groups are evil, corrupt, or malicious based on protected attributes (*e.g.,* age, race, religion, etc.), or hateful supremacist propaganda, including the recruitment of new members or requests for financial support for their ideology, all violate our hate speech policy (*https://support.google.com/youtube/answer/2801939*) and are subject to removal as such.

As detailed in our October update, these are not the first steps we have taken to limit the reach of harmful misinformation. Nearly two years ago, we updated our recommendations system, including reducing recommendations of borderline content and content that could misinform users in harmful ways—such as videos promoting a phony miracle cure for a serious illness, claiming the earth is flat, or making blatantly false claims about historic events like 9/11. This resulted in a 70 percent drop in views coming from our search and discovery systems. Further, when we looked at QAnon content, we saw the number of views that come from non-subscribed recommendations to prominent Q-related channels dropped by over 80 percent from January 2019 to October 2020. For more information on these policies and enforcement actions, please see *https://blog.youtube/news-and-events/harmful-conspiracy-theories-youtube/, https://blog.youtube/news-and-events/continuing-our-work-to-improve/,* and *https://blog.youtube/news-and-events/our-ongoing-work-to-tackle-hate.*

Our Ads policies are similarly designed to ensure a safe and positive experience for our users. For example, under our existing misrepresentation policy, we do not allow ads to run or content to monetize that promotes medically unsubstantiated claims related to COVID cures (*https://support.google.com/google-ads/answer/9811449*). Since this past February, we've blocked or removed over 270 million coronavirus related ads across all Google advertising platforms; additionally, we've removed 600,000 YouTube videos with dangerous or misleading coronavirus information.

In addition to removing content that violates our policies, we also reduce borderline content and raise up authoritative voices by providing users with more information about the content they are seeing to allow them to make educated choices. On YouTube, for example, there have been billions of impressions on information panels around the world since June 2018. For more information, please see *https://support.google.com/youtube/answer/9229632.* And, for over three years, we have high-

lighted fact checks on Search and News as a way to help people make more informed judgments about the content they encounter online. People come across these fact checks billions of times per year. For more information, please see *https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/*.

The openness of our platforms has helped creativity and access to information thrive. It's our responsibility to protect that, and prevent our platforms from being used to incite hatred, harassment, discrimination, and violence. We are committed to taking the steps needed to live up to this responsibility today, tomorrow, and in the years to come.

Question 2. It is clear that the existence of conspiracy theories, disinformation campaigns, and misinformation has led to violence, even if not specifically planned on your platform. Recently, Google and YouTube have taken action against the QAnon conspiracy for this reason. Why did QAnon reach that threshold now, and how will Google and YouTube address other conspiracies?

Question 2a. Is there a set number of violent incidents that must occur before Google and YouTube consider a group unfit for the platforms?

Answer. We approach QAnon the same way we would approach other content that violates our policies. We apply our four pillars of action: remove violative content; raise up authoritative content; reduce the spread of borderline content; and reward trusted creators. For more information on these four pillars, please see *https://blog.youtube/news-and-events/our-ongoing-work-to-tackle-hate.* Among other things, we remove content that violates our hate speech, harassment, and COVID misinformation policies. We have removed tens of thousands of Q-related videos that target specific groups, and have terminated hundreds of Q-related channels on YouTube. If a creator's content violates our Community Guidelines, we will issue a strike against their channel; their channel will be terminated if they receive three strikes. We also terminate entire channels if they are dedicated to posting content prohibited by our Community Guidelines or contain a single egregious violation, like child sexual abuse material.

Additionally, we reduce the spread of content that gets close to the line of violating our policies—including removing that content from the recommendations we show to our users. As described in our response to Question No. 1, this has resulted in a drop of over 80 percent of views from January 2019 to October 2020 from non-subscribed recommendations to prominent Q-related channels. We also raise up information panels to provide contextual information for QAnon content. Since 2018, we have seen 25 million impressions on our QAnon information panel. Further, we set a higher bar for what channels can make money on our site, rewarding trusted, eligible creators; we don't allow QAnon in Ads, because their content violates our dangerous or derogatory content policy. Additionally, we recently enhanced our policies that address harmful conspiracies, including QAnon, on YouTube. For more information, please see *https://blog.youtube/news-and-events/harmful-conspiracy-theories-youtube.*

All of this work has been pivotal in curbing the reach of harmful conspiracies like QAnon, and we will continue to approach this thoughtfully, balancing maintaining a platform for free speech and living up to our responsibility to users.

Question 3. YouTube policies ensure that family-friendly advertisers do not have their paid ads run before potentially harmful content, but that same content is still readily served up to viewers based on your algorithm. It is clear YouTube algorithms can identify the problematic content, yet the algorithm quickly steers users to this extremist content. How many people have to view and/or report extremist content before YouTube takes it down?

Question 3a. Why does YouTube then allow those same content creators multiple opportunities to post extremist content before they hit the "three strikes" policy?

Answer. We believe strongly in the freedom of expression and access to information—we know that the overwhelming majority of creators follow our guidelines and understand that they are part of a large, influential, and interconnected community. However, we also know that we have a responsibility to protect our users, which is why we have policies prohibiting hate speech, terrorist content, and other content that violates our policies, as well as stricter standards for who can monetize their content. Each of the products and services we offer has a different purpose, and we tailor our approach carefully to the content that should be available on each product and service.

While YouTube creates a space for ideas and expression, it is not a free-for-all. For example, it is a violation of YouTube's hate speech policy for users to post videos that promote violence against particular ethnic or religious groups (*https://support.google.com/youtube/answer/2801939*). As described in our response to Question

No. 2, creators who violate those rules may have their content removed or their accounts terminated. When we detect a video that violates our Community Guidelines, we remove the video and apply a strike to the channel. The strike restricts a creator's ability to post or create content on the platform for one week. If the creator's behavior warrants another strike within 90 days from the first, a new two-week prohibition from posting or creating content is implemented. A third strike within 90 days results in permanent removal of a channel from YouTube. Creators can appeal those strikes if they believe we are mistaken. We also terminate entire channels if they are dedicated to posting content prohibited by our Community Guidelines or contain a single egregious violation, like child sexual abuse material.

Concerning the timing of content removal, we strive to remove violative content as quickly as possible. We take down half of extremist content on YouTube within two hours, and nearly 70 percent in eight hours. Further, as detailed in the YouTube Community Guidelines Enforcement Transparency Report (*https://transparencyreport.google.com/youtube-policy/removals*), in the third quarter of 2020, more than 7.8 million videos were removed from YouTube for violating our Community Guidelines—94 percent of which were first flagged by machines rather than humans. Of those detected by machines, over 45 percent never received a single view, and just over 80 percent received fewer than 10 views.

Moreover, in 2019 we announced that we had begun reducing recommendations of borderline content on YouTube. This is content which comes close to but doesn't quite violate our policies and represents less than one percent of the content watched on YouTube. These changes have already reduced views from non-subscribed recommendations of this type of content by 70 percent in the U.S. and have been rolled out in 33 countries with more to follow.

It is also important to note that the vast majority of attempted abuse comes from bad actors trying to upload spam or adult content, as opposed to extremist content. For example, nearly 92 percent of the channels and over 45 percent of the videos that we removed in the third quarter of 2020 were removed for violating our policies on spam or adult content. In comparison, promotion of violence and violent extremism accounted for only 0.5 percent of removed channels and 2.5 percent of removed videos during the same period. For more information, please see our YouTube Community Guidelines Enforcement Transparency Report, *https://transparencyreport.google.com/youtube-policy/removals.*

We are proud of our efforts to prevent the spread of this type of content and are working to do everything we can to ensure users are not exposed to extremist content.

Question 4. While I appreciate that Google and YouTube continue to evolve and learn about threats of violence on the platforms, would you agree that as groups evolve and change their tactics you will always be one step behind extremist groups that seek to use social media to recruit and plan violent acts? How do you address this problem?

Answer. As described in our response to Question No. 3, we strive to remove content that violates our policies as quickly as possible. To enforce our policies at the scale of the web, we use a combination of human reviewers and cutting-edge machine learning to combat violent and extremist content. We estimate that we spent at least $1 billion over the past year on content moderation systems and processes, and we continue to invest aggressively in this area. In the last year, more than 20,000 people have worked in a variety of roles to help enforce our policies and moderate content. We're also constantly innovating to improve our machine learning and algorithms to spot content in violation of our policies. And, we partner with a network of academics, industry groups, and subject matter experts to help us better understand emerging issues.

These improvements are happening every day, and we will need to adapt, invent, and react as hate and extremism evolve online. We're committed to this constant improvement, and the significant human and technological investments we're making demonstrate that we're in it for the long haul.

We also recognize the value of collaborating with industry partners to prevent terrorists and violent extremists from exploiting our platforms. That is why in 2017, YouTube, Facebook, Microsoft and Twitter founded the Global Internet Forum to Counter Terrorism (GIFCT) as a group of companies dedicated to disrupting terrorist abuse of members' digital platforms.

Although our companies have been sharing best practices around counter-terrorism for several years, GIFCT provided a more formal structure to accelerate and strengthen this work and present a united front against the online dissemination of terrorist and violent extremist content.

YouTube and GIFCT's other founding members signed on to the Christchurch Call to Eliminate Terrorist and Violent Extremist Content Online (*https://www*

.christchurchcall.com/). Building on the Christchurch Call, GIFCT developed a new content incident protocol for GIFCT member companies to quickly share digital hashes of content and respond efficiently after a violent attack. This protocol has been tested and proven effective, for example, following the attack on a synagogue in Halle, Germany (October 2019) and following a shooting in Glendale, Arizona here in the United States (May 2020).

GIFCT has evolved to be a standalone organization with an independent Executive Director, Nicholas J. Rasmussen, formerly Director of the National Counterterrorism Center and dedicated staff. For more information, please see *https://gifct.org/about/story/#june-2020—appointment-of-executive-director-and-formation-of-the-independent-advisory-committee-1.* We remain committed to the GIFCT and hold a position on the independent GIFCT's Operating Board within the new governance framework of the institution.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. KYRSTEN SINEMA TO SUNDAR PICHAI

COVID–19 Misinformation. The United States remains in the midst of a global pandemic. More than 227,000 Americans have died of COVID–19, including nearly 6,000 in my home state of Arizona. COVID has impacted the health, employment, and education of Arizonans, from large cities to tribal lands like the Navajo Nation. And at the time of this hearing, the country is facing another significant surge in cases.

The persistent spread of COVID–19 misinformation on social media remains a significant concern to health officials. Digital platforms allow for inflammatory, dangerous, and inaccurate information—or outright lies—to spread rapidly. Sometimes it seems that misinformation about the virus spreads as rapidly as the virus itself. This misinformation can endanger the lives and livelihoods of Arizonans.

Social distancing, hand washing, testing, contact tracing, and mask wearing should not be partisan issues, nor should they be the subject of online misinformation.

Question 1. What has Google done to limit the spread of dangerous misinformation related to COVID–19 and what more can it do?

Answer. Since the outbreak of COVID–19, our efforts have focused on keeping people informed with trusted and authoritative information, supporting people as they adapt to the current situation, and contributing to recovery efforts. To help ensure that people are well informed, we have taken multiple steps to organize and provide accurate and verifiable information on the pandemic. These efforts to fight misinformation across our platforms include our Homepage "Do the Five" promotion, amplifying authoritative voices through ad grants (*https://support.google.com/google-ads/answer/9803410*), and launching our COVID–19 site (*https://www.google.com/intl/en_us/covid19/*), which includes coronavirus information, insights, and resources.

A number of the policies and product features that were used for the COVID–19 crisis were already in place before the crisis began, and others were underway. For example, our ranking systems on Google Search and YouTube have been designed to elevate authoritative information in response to health-related searches for years. Before 2020, YouTube's advertiser content guidelines (*https://support.google.com/youtube/answer/6162278*) already prohibited "harmful health or medical claims or practices," and our work to update our YouTube recommendation systems to decrease the spread of misinformation, including, but not limited to, health-related misinformation, was announced in January 2019. For more information, please see *https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html.*

Since the outbreak of COVID–19, we also implemented, and have enforced, a COVID misinformation policy (*https://support.google.com/youtube/answer/9891785*) to facilitate removal of COVID–19-related misinformation on YouTube.

With respect to COVID-related ads, our Ads policies (*https://support.google.com/adspolicy/answer/6008942*) are designed not only to abide by laws, but also to ensure a safe and positive experience for our users. This means that our policies prohibit some content that we believe to be harmful to users and the overall advertising ecosystem. This includes policies that prohibit ads for counterfeit products, dangerous products or services, or dishonest behavior, and any content that seeks to capitalize on the pandemic, or lacks reasonable sensitivity towards the COVID–19 global health crisis. For more information on these policies, please see *https://support.google.com/google-ads/answer/9811449.* In addition, our dangerous or derogatory content policy (*https://support.google.com/adspolicy/answer/6015406*) prohibits content in Google Ads that would advocate for physical or mental harm, such

as content that denies the efficacy of vaccines, as well as content that relates to a current, major health crisis and contradicts authoritative scientific consensus. As a result, content contradicted by scientific consensus during COVID–19 such as origin theories, claims the virus was created as a bioweapon, as well as claims the virus is a hoax or government-funded are not permitted on our platform.

And these efforts to limit the spread of COVID misinformation are working. There have been over 400 billion impressions on our information panels for coronavirus related videos and searches, and, since February, we've removed 600,000 coronavirus videos and removed or blocked over 270 million coronavirus-related ads globally across all Google advertising platforms—including Shopping ads—for policy violations including price-gouging, capitalizing on global medical supply shortages, and making misleading claims about cures.

We are proud of our efforts to combat health misinformation and address this unprecedented public health crisis. We will continue to work hard and do everything we can to help our communities in addressing this global pandemic.

Spreading Accurate Information. Arizonans need accurate, scientifically based information to help get through this pandemic. Many Arizonans get their news from sources such as Google. As a result, your companies can play a role in helping people receive accurate information that is relevant to their communities and can aid them in their decisions that keep their families healthy and safe.

For example, earlier this month, the CDC issued a report illustrating that COVID–19 cases fell dramatically in Arizona after prevention and control measures were put into place. I shared this information on social media, and this is the type of information we should emphasize to help save lives.

Question 2. What more can Google do to better amplify accurate, scientifically-based health information to ensure that Arizonans understand how best to protect themselves from the pandemic?

Answer. Since the outbreak of COVID–19, we have worked to surface trusted and authoritative information and partner with health organizations and governments in order to bring our users information they can rely on in a rapidly changing environment.

With Search, for example, in partnership with the CDC and other health authorities, we have promoted important guidance to prevent the spread of COVID–19. We have introduced a comprehensive experience for users seeking information relating to COVID–19 that provides easy access to information from health authorities alongside new data and visualizations (*https://blog.google/products/search/connecting-people-covid-19-information-and-resources/*). This new format organizes the search results page to help people easily navigate resources and makes it possible to add more information as it becomes available over time. This experience came as a complement to pre-existing work on Google Search and Google News to recognize sensitive events and contexts, and our systems are designed to elevate authoritative sources for those classes of queries.

Across YouTube, we similarly elevate authoritative sources such as the CDC and other authorities to help users get the latest COVID–19 information. With anti-vaccination content, for example, we elevate reliable information across both Google and YouTube regarding medical topics (including vaccination) from trustworthy sources, such as health authorities.

Another way we connect users to authoritative content is by providing contextual information that can be used to help them determine for themselves the trustworthiness of the content they are provided. On YouTube, for example, we've included fact check information panels on COVID–19 videos, that feature information on COVID–19 symptoms, prevention, and treatment, and links to the CDC and other health authorities. These panels provide fresh context during fast-moving situations such as COVID–19 by highlighting relevant, third-party fact-checked articles above search results for relevant queries. For more information, please see *https://blog.youtube/news-and-events/expanding-fact-checks-on-youtube-To-united-states, https://support.google.com/youtube/answer/9795167,* and *https://support.google.com/youtube/answer/9004474.* In addition, YouTube elevates content from authoritative channels such as news organizations or health authorities when our systems detect that a user's search is health-related.

We are committed to our responsibility to provide relevant and authoritative context to our users, and to continue to reduce the spread of harmful misinformation across our products.

Scientific Evidence-based COVID Information. Our best sources of information related to the pandemic are doctors, researchers, and scientists. We should be relying on their expertise to help stop the spread of the virus and help our country recover from its devastating impacts.

Question 3. Who determines whether content on Google is scientifically supported and evidence based?

Answer. As noted in our response to Question Nos. 1 and 2, we have invested heavily to ensure that we surface authoritative content, and have taken active steps to detect and remove COVID–19 related misinformation that contradicts guidance from health authorities and may result in real-world harm.

For example, to ensure Search algorithms meet high standards of relevance and quality, we have a rigorous process that involves both live tests and thousands of trained external Search Quality Raters from around the world. Our Search Quality Rater Guidelines (*https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf*) provide that there is a higher standard when a user is looking for things like specifical medical information or advice. In that case, we work to provide content from authoritative sources like health professionals and medical organizations. The Guidelines explicitly state, for example, that "medical advice should be written or produced by people or organizations with appropriate medical expertise or accreditation," and that "information pages on scientific topics should be produced by people or organizations with appropriate scientific expertise and represent well-established scientific consensus on issues where such consensus exists." For additional information on Search Quality Raters and how ratings work, please see *https://blog.google/products/search/raters-experiments-improve-google-search.*

In terms of removing content, we rely on a mix of automated and manual efforts to spot problematic content. Our automated systems are carefully trained to quickly identify and take action against spam and violative content. This includes flagging potentially problematic content for reviewers, whose judgement is needed for the many decisions that require a more nuanced determination. The context in which a piece of content is created or shared is an important factor in any assessment about its quality or its purpose, and we are attentive to educational and scientific contexts where the content might otherwise violate our policies.

Moreover, as the COVID–19 situation has evolved, we have partnered closely with the CDC and other health authorities to ensure that our policy enforcement is effective in preventing the spread of harmful misinformation relating to COVID–19. Our YouTube policies prohibit, for example, content that explicitly disputes the efficacy of CDC and other health authority advice regarding social distancing that may lead people to act against that guidance. For more information, please see our COVID–19 misinformation policy, *https://support.google.com/youtube/answer/9891785.*

We are proud of our efforts to combat health misinformation and address this unprecedented public health crisis, and will continue to work hard and do everything we can to help our communities in addressing this global pandemic.

COVID Scams. Arizonans and Americans have been inundated with fraudulent offers and scams, using social media to spread inaccurate information and perpetrate criminal scams. I've been using my own social media to help warn Arizonans about common scams related to economic assistance, false coronavirus "cures", and where they can report scams to Federal and state authorities.

Question 4. What has Google done to limit the spread of scams and report criminal activity and what more can be done to protect seniors, veterans, and others who have been targeted by fraudsters?

Answer. As people around the world are staying at home more due to COVID–19, many are turning to new apps and communications tools to work, learn, access information, and stay connected with loved ones. While these digital platforms are helpful in our daily lives, they can also introduce new online security risks. Bad actors are creating new attacks and scams every day that attempt to take advantage of the fear and uncertainty surrounding the pandemic—and we are committed to working to constantly stay ahead of those threats.

Our security systems have detected a range of new scams, such as phishing e-mails posing as messages from charities and NGOs battling COVID–19, directions from "administrators" to employees working from home, and even notices spoofing healthcare providers. For example, in just one week, we saw 18 million daily malware and phishing e-mails related to COVID–19—in addition to more than 240 million COVID-related daily spam messages. Our systems have also spotted malware-laden sites that pose as sign-in pages for popular social media accounts, health organizations, and even official coronavirus maps. As to government-backed hacking activity, our Threat Analysis Group continually monitors for such threats and is seeing new COVID–19 messaging used in attacks. For more information, please see *https://blog.google/threat-analysis-group/,* and *https://cloud.google.com/blog/products/identity-security/protecting-against-cyber-threats-during-covid-19-and-beyond.*

In many cases, these threats are not new—rather, they are existing malware campaigns that have simply been updated to exploit the heightened attention on COVID–19. To protect against these attacks, we have put proactive monitoring in place for COVID–19-related malware and phishing across our systems and workflows and have built advanced security protections into Google products to automatically identify and stop threats before they ever reach users. For example, our machine learning models in Gmail already detect and block more than 99.9 percent of spam, phishing, and malware. Our built-in security protocols also protect users by alerting them before they enter fraudulent websites, by scanning apps in Google Play before downloads, and more.

When we identify a threat, we add it to the Safe Browsing API, which protects users in Chrome, Gmail, and all other integrated products. Safe Browsing helps protect over four billion devices every day by showing warnings to users when they attempt to navigate to dangerous sites or download dangerous files. Further, in G Suite, advanced phishing and malware controls are turned on by default, ensuring that all G Suite users automatically have these proactive protections in place.

Because we have a longstanding and unwavering commitment to security and want to help users stay secure everywhere online, not just on our products, we've also provided tips, tools, and resources relating to online security in our Safety Center (*https://safety.google/securitytips-covid19/*) and public blogs, including *https://blog.google/technology/safety-security/helping-you-avoid-covid-19-security-risks/, https://cloud.google.com/blog/products/identity-security/protecting-against-cyber-threats-during-covid-19-and-beyond,* and *https://cloud.google.com/blog/products/identity-security/blocking-coronavirus-cyber-threats.*

Finally, to help facilitate reporting of COVID-related scams to law enforcement authorities, we have worked closely with the Department of Justice (DOJ) and included its COVID–19 fraud site in our COVID–19 Safety Center (*https://safety.google/securitytips-covid19/*). Moreover, we have received COVID–19 fraud-related data from DOJ to review for policy violations, and we have also submitted proactive criminal referrals to DOJ for potential COVID–19-related criminal activity.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. JACKY ROSEN TO SUNDAR PICHAI

Question 1. Adversaries like Russia continue to amplify propaganda—on everything from the election to the coronavirus to anti-Semitic conspiracy theories—and they do it on your platform, weaponizing division and hate to destroy our democracy and our communities. The U.S. intelligence community warned us earlier this year that Russia is now actively inciting white supremacist violence, which the FBI and Department of Homeland Security say poses the most lethal threat to America. In recent years, we have seen white supremacy and anti-Semitism on the rise, much of it spreading online. What enables these bad actors to disseminate their hateful messaging to the American public are the algorithms on your platforms, effectively rewarding efforts by foreign powers to exploit divisions in our country.

Question 1a. Are you seeing foreign manipulation or amplification of white supremacist and anti-Semitic content, and if so, how are your algorithms stopping this? Are your algorithms dynamic and nimble enough to combat even better and more personalized targeting that can be harder to identify?

Question 1b. Have you increased or modified your efforts to quell Russian disinformation in the wake of recently revealed efforts by Russia and Iran to weaponize stolen voter data to exploit divisions in our nation? How have you or will you adjust your algorithms to reduce the influence of such content—knowing that these countries' newly obtained data will allow for even better targeting, making their deception harder to identify?

Answer. Because the answers to these questions are related, we have grouped together our response to these subparts of Question No. 1.

We are deeply concerned about any attempts to use our platforms to sow division and hate. That's why our teams are constantly on the lookout for malicious actors that try to game our platforms, and we take strong action against coordinated influence operations. We've dedicated significant resources to help protect our platforms from such attacks by maintaining cutting-edge defensive systems and by building advanced security tools directly into our consumer products. As examples of how our systems and policies are actively at work identifying and removing such content, in the third quarter of 2020, over 7.8 million videos were removed by YouTube for violating Community Guidelines, including violations of our policies regarding spam, misleading content, and scams (25.5 percent), violent or graphic content (14.2 per-

cent), promotion of violence or violent extremism (2.5 percent), harmful or dangerous content (2.5 percent), and hateful or abusive content (1.1 percent). For more information, please see our YouTube Community Guidelines Enforcement Transparency Report, *https://transparencyreport.google.com/youtube-policy/removals.*

On any given day, Google's Threat Analysis Group is also tracking more than 270 targeted or government-backed attacker groups from more than 50 countries. When we find attempts to conduct coordinated influence operations on our platforms, we work with our Trust and Safety teams to swiftly remove such content from our platforms and terminate these actors' accounts. We take steps to prevent possible future attempts by the same actors, and routinely exchange information and share our findings with others in the industry. For example, in October 2020, the Department of Justice acknowledged Google's contributions to the fight against Iranian influence operations in announcing the seizure of 92 domain names used by Iran's Islamic Revolutionary Guard Corps to engage in a global disinformation campaign targeting the U.S. and other countries (*https://www.justice.gov/usao-ndca/pr/united-states-seizes-domain-names-used-iran-s-islamic-revolutionary-guard-corps*). Additionally, if we suspect that users are subject to government-sponsored attacks, we warn them. In April 2020 alone, for example, we sent 1,755 warnings to users whose accounts were targets of government-backed attackers. For more information about these actions on our Threat Analysis Group blog, please see *https://blog.google/threat-analysis-group/*.

While some tools may work for violent extremism and terrorism-related content in a scalable way, the problem is very different for misleading or inauthentic content. Many times, the misleading content looks identical to content uploaded by genuine activists. As noted in our response to Senator Peters' Question No. 4, that is why we use a combination of human reviewers and cutting-edge machine learning. Technology has helped us accelerate and scale our removal of content that violates our policies, but we also rely on highly-trained individuals from our Trust and Safety and Security teams, who work closely with machine learning tools and our algorithms, to ensure our platforms are protected and there is adherence to our policies.

On YouTube, we also employ a sophisticated spam and security-breach detection system to identify anomalous behavior and attempts to manipulate our systems. We have also increased transparency around news sources on YouTube, including disclosure of government funding. When a news channel on YouTube receives government funding, we make that fact clear by including an information panel under each of that channel's videos. Our goal is to equip users with additional information to help them better understand the sources of news content that they choose to watch on YouTube. For more information, please see *https://blog.youtube/news-and-events/greater-transparency-for-users-around.*

As threats evolve, we will continue to adapt to understand and prevent new attempts to misuse our platforms and will continue to expand our use of cutting-edge technology to protect our users. There are no easy answers here, but we are deeply committed to getting this right.

Question 1c. Are you consulting outside groups to validate moderator guidelines on hate speech, including what constitutes anti-Semitic content? Are you collecting data on hate speech content? If so, what are you doing with that data to combat hate speech on your platforms?

Answer. As described in our response to Senator Blumenthal's Question No. 4, we consult with a diverse set of external and internal stakeholders during policy development, including expert input, user feedback, and regulatory guidance. One of the most complex and constantly evolving areas we deal with is hate speech. We systematically review and re-review all our policies to make sure we are drawing the line in the right place, often consulting with subject matter experts for insight on emerging trends. For our hate speech policy, we work with experts in subjects like violent extremism, supremacism, civil rights, and free speech from across the political spectrum.

Hate speech is a complex policy area to enforce at scale, as decisions require nuanced understanding of local languages and contexts. To help us consistently enforce our policy, we have expanded our review team's linguistic and subject matter expertise. We also deploy machine learning to better detect potentially hateful content to send for human review, applying lessons from our enforcement against other types of content, like violent extremism. As noted in our response to Senator Peters' Question No. 1, we have also recently taken a tougher stance on removing hateful and supremacist content and have reduced borderline content by reducing recommendations of content that comes close to violating our guidelines. Since early 2019, we've increased by 46 times our daily hate speech comment removals on YouTube. And in the last quarter, of the more than 1.8 million channels we terminated for violating our policies, more than 54,000 terminations were for hate speech.

This is the most hate speech terminations in a single quarter and three times more than the previous high from Q2 2019 when we updated our hate speech policy. For additional information regarding enforcement of, and improvements to, our hate speech policies, please see *https://blog.youtube/news-and-events/make-youtube-more-inclusive-platform/, https://transparencyreport.google.com/youtube-policy/featured-policies/hate-speech,* and *https://blog.youtube/news-and-events/our-ongoing-work-to-tackle-hate.*

Question 1d. When advertisers purchase an ad campaign on YouTube, Google then takes the advertisement and plays it with videos determined by an algorithm to be the best fit for the ad. Google pays the video's creator a small fee each time a user plays or clicks on the ad. What specific steps is Google taking to ensure that the creators of videos containing hateful content do not receive advertisement-related fees from Google?

Answer. It is critical that our monetization systems reward trusted creators who add value to YouTube. We have longstanding guidelines (*https://support.google.com/youtube/answer/6162278*) that prohibit ads from running (and thus, no fees are paid) on videos that include hateful content, and we enforce these policies rigorously. Channels that repeatedly brush up against our hate speech policies will be suspended from the YouTube Partner program, meaning they can't run ads on their channel or use other monetization features like Super Chat. In order to protect our ecosystem of creators, advertisers, and viewers, we also tightened our advertising criteria in 2017 (*https://blog.youtube/news-and-events/additional-changes-to-youtube-partner*). After thorough analysis and conversations with creators, we changed certain eligibility requirements for monetization, which significantly improved our ability to identify creators who contribute positively to the community, while also preventing potentially inappropriate videos from monetizing content. For more information about these actions, please see *https://blog.youtube/news-and-events/our-ongoing-work-to-tackle-hate.*

Question 2. Recently, there have been high profile cybersecurity breaches involving private companies, government agencies, and even school districts—including in my home state of Nevada. A few months ago, a hacker subjected Clark County School District—Nevada's largest school district and our country's fifth largest, serving more than 320,000 students—to a ransomware attack. In the tech industry, there was a notable breach of Twitter in July, when hackers were able to access an internal IT administrator tool used to manage accounts. Dozens of verified accounts with high follower counts—including those of President Obama, Bill Gates, and Jeff Bezos—were used to send out a tweet promoting a Bitcoin scam. What we learned from this breach is stunning. . .the perpetrators were inside the Twitter network in one form or another.

Question 2a. How often do your staff attend cybersecurity training? Do you hire outside cybersecurity firms to look at your systems, offering a fresh look and catching overlooked flaws?

Question 2b. Now that many schools have migrated to using Google products for distance education, how are you ensuring that students, teachers, and schools are adequately protected from cyberattacks?

Answer. As a company, cybersecurity is a critical priority, and we are proud to have a strong security culture. All Google employees undergo security training as part of the orientation process and receive ongoing security training throughout their Google careers. During orientation, new employees agree to our Code of Conduct, which highlights our commitment to keep customer information safe and secure. Depending on their job/role, additional training on specific aspects of security may be required. For instance, the information security team instructs new engineers on topics like secure coding practices, product design, and automated vulnerability testing tools. Engineers also attend technical presentations on security-related topics and receive a security newsletter that covers new threats, attack patterns, mitigation techniques, and more. In addition, we host regular internal conferences to raise awareness and drive innovation in security and data privacy, which are open to all employees. Security and privacy is an ever-evolving area, and we recognize that dedicated employee engagement is a key means of raising awareness. We host regular "Tech Talks" focusing on subjects that often include security and privacy. In addition, we bring in outside experts from third-party vendors and law firms to assist with training our employees on relevant topics to make sure all our training needs are met. We also regularly undergo independent, third-party verification of our security, privacy, and compliance controls.

As the world continues to adapt to the changes brought on by the COVID–19 pandemic, cyber threats are evolving as well. As noted in our response to Senator Sinema's Question No. 4, bad actors are creating new attacks and scams every day

that attempt to take advantage of the fear and uncertainty surrounding the pandemic—it's our job to stay ahead of those threats. Our teams work every day to make our products safe no matter what users are doing—browsing the web, managing their inbox, or seeing family on Google Meet. Keeping users safe online means continuously protecting the security and privacy of their information. That is why protections are automatically built into each user's Google Account and every Google product: Safe Browsing protects more than 4 billion devices; Gmail blocks more than 100 million phishing attempts every day; and Google Play Protect scans over 100 billion apps every day for malware and other issues. Further, in G Suite, advanced phishing and malware controls are turned on by default, ensuring that all G Suite users automatically have these proactive protections in place. G Suite administrators can also look at Google-recommended defenses on our advanced phishing and malware protection page (*https://support.google.com/a/answer/9157861*), and may choose to enable the security sandbox, a virtual environment where Gmail scans or runs attachments (*https://support.google.com/a/answer/7676854*).

Because we have a longstanding and unwavering commitment to security and want to help users stay secure everywhere online, not just on our products, we've also provided tips, tools, and resources relating to online security in our Safety Center (*https://safety.google/securitytips-covid19/*) and public blogs, including *https://blog.google/technology/safety-security/helping-you-avoid-covid-19-security-risks/, https://cloud.google.com/blog/products/identity-security/protecting-against-cyber-threats-during-covid-19-and-beyond,* and *https://cloud.google.com/blog/products/identity-security/blocking-coronavirus-cyber-threats.* For more information about privacy and security in G Suite for Education, please see our Privacy and Security Center, *https://edu.google.com/why-google/privacy-security/*. Safeguarding user security—including that of students, teachers, and administrators—is an obligation we take very seriously, and we will continue to invest appropriate technical resources in this area.

Question 3. The COVID–19 pandemic has shined a light on our Nation's digital divide and on the technological inequalities facing millions of American students, including those in Nevada. Lack of access to broadband disproportionately affects low-income communities, rural populations, and tribal nations—all of which are present in my state. In addition to broadband access, many students still do not have regular access to a computer or other connected device, making online learning incredibly difficult, and sometimes impossible.

Google stepped up during the pandemic to help close the digital divide. You provided Chromebook tablets to students lacking devices, including to thousands in Clark County, and also updated Google Classroom products to help students and school districts around the world adapt to online education.

Question 3a. As classes continue to meet online, or in a hybrid model, what more can Google do to help students and teachers?

Answer. We recognize that families and educators are relying on digital platforms to provide access to online learning and educational tools—especially during COVID–19—and Google is proud to help students continue their education from home during the pandemic. From the very beginning, Google has been committed to providing students, teachers, parents, and IT administrators with the tools young learners need to be successful.

Since March 2020, Google has offered free access to its advanced Google Meet features through its Google Classroom solution (*https://edu.google.com/products/classroom/*) that is used by thousands of school districts, charter schools, private and parochial schools, as well as home schoolers.

To aid teachers and those assisting students at home, Google launched a website that lists resources and tips for teaching classes remotely through a new Teach from Home hub (*https://teachfromanywhere.google/intl/en/#for-teachers*) with information and resources. This hub includes tutorials, step-by-step guides, and inspiration for distance learning during school closures.

In addition, our G Suite for Education solution is free, or can be upgraded to an enterprise solution (*https://edu.google.com/products/gsuite-for-education/*), and helps more than 120 million teachers and students around the world work and learn together. We also created a dedicated Distance Learning Fund through Google.org to help educators and parents access tools and resources needed to provide learning opportunities for students (*https://www.blog.google/outreach-initiatives/education/helping-educators-and-students-stay-connected/*). The Fund supports Khan Academy, Wide Open Schools by Common Sense Media, and DonorsChoose.

Google also has made it easy to turn school-based Chromebooks into take-home devices for students. Through Chromebook resellers, multiple school districts have purchased Chromebooks to distribute to students, and we are proud that our prod-

ucts provide a functional and accessible avenue to remote learning. Policies and permissions for Chromebooks can be set by IT administrators using Chrome Education Upgrade through the Google Admin console, making it simple for schools to deploy and manage thousands of devices. There's no need to manually install software or login to a device to apply settings—admins can simply flip a switch online and every device updates its applications and settings automatically. Moreover, educators can easily integrate the collaborative power of Google's educational tools into their learning management systems. For more information, please see *https://edu.google.com/products/gsuite-for-education/*.

Even if students don't have WiFi access, they can still access their Google Drive and edit and save files offline. That said, we are acutely aware of the fact that millions of students globally don't have connectivity at home, which is what inspired us to create Rolling Study Halls (*https://edu.google.com/why-google/our-commitment/rolling-study-halls/*), a program that equips school buses across the U.S. with WiFi, devices, and onboard educator support. This program has been expanded not only by Google, but also by numerous school districts and other providers.

Finally, Google's broader efforts in bringing broadband access to rural communities is key to closing the digital divide. Google Fiber's Community Connections program (*https://fiber.google.com/community/*) offers organizations such as libraries, community centers, and nonprofits free Internet access. To keep up with the rising demand for bandwidth, the FCC has worked with industry leaders like Google to create the CBRS rules (*https://www.cbrsalliance.org/resource/what-is-cbrs/*) for shared spectrum as a new model for adding capacity at a low cost. By aligning on industry standards, Google is helping the CBRS ecosystem bring better wireless Internet to more people in more places. As the foundation for Google's suite of products and services for CBRS (*https://www.google.com/get/spectrumdatabase/#cbrs*), Google's Spectrum Access System (SAS) controls fundamental access to CBRS. Google's SAS is purpose-built to support dense networks across operators and to scale on demand—from a small in-building network to the largest nationwide deployment. For more information on how Google is bringing affordable Internet and choice to consumers, please see *https://www.google.com/get/spectrumdatabase/sas/*.

As remote learning has evolved, so have we. We're continuing to work with partners and local communities to see what else we can do to help support students without access at home. And we will continue to update our resource hub (*https://edu.google.com/latest-news/covid-19-support-resources/*) so that educators and IT professionals can find the latest materials, resources, and training. Google is committed to helping students and educators across the country, and would be pleased to discuss the best ways Google can continue to serve our communities.

Question 3b. How does Google plan to remain engaged in K–12 education after we get through the pandemic? In particular, what role can you play in closing not only the urban/rural divide, but also the racial divide in access to technologies and the Internet?

Answer. Even before the pandemic, teachers increasingly assigned schoolwork that required access to the internet. Google knows that millions of students lack connectivity at home. This "Homework Gap" disproportionately impacts low-income students, especially in more remote or rural areas, where they face additional burdens like long bus commutes. To help ease this gap, Google piloted a program of Rolling Study Halls in North Carolina and South Carolina (*https://edu.google.com/why-google/our-commitment/rolling-study-halls/*). As noted in our response to Question No. 3.a, this program equips school buses with WiFi, devices, and onboard educator support. Since early results indicate promising gains in reading and math proficiency, and increased digital fluency, Google is expanding the program to reach thousands more students across 16 more school districts across 12 states (Alabama, Colorado, Georgia, Kansas, Minnesota, New Mexico, Oregon, Pennsylvania, South Carolina, Tennessee, Texas, and Virginia), focused on rural communities.

Moreover, Google's G Suite for Education has helped teachers and students connect and collaborate, even when they are not able to be in the same classroom. But even when all students are able to return to the classroom, G Suite for Education tools will help students turn in their best work, help teachers create assignments and grade work all in one place, and help schools stay on top of their daily activities, all in a secure and safe online environment.

Google is also committing nearly $3 million to help close the racial equity gaps in computer science education and increase Black+ representation in STEM fields. And, in connection with our recently-announced Workplace Commitments, we will ensure that $310 million in funding goes toward diversity, equity, and inclusion initiatives and programs focused on increasing access to computer science education and careers. We are committed to staying engaged in K–12 education after we get

through the pandemic and would be pleased to discuss other opportunities for further engagement.

Question 3c. There has been a surge in demand across the globe for affordable laptops and Chromebooks, which has created months-long shipment delays. How is Google working with its manufacturers on this shortage issue?

Google is committed to helping students and educators across the country, and is proud of the positive impact that technology has played in that effort. Chromebooks have played a crucial role in remote learning, and we continue to work to ensure that everyone who needs access to an affordable computer has it.

Question 4. One of my top priorities in Congress is supporting the STEM workforce and breaking down barriers to entering and succeeding in STEM fields. This includes ensuring we have a diverse STEM workforce that includes people of color and women. In the past several years, tech companies have begun releasing diversity reports and promising to do better at hiring Black and Latino workers, including women. In overall employment, Google is doing much better today in building a diverse workforce. However, while overall diversity is increasing, only 2.4 percent of Google tech employees in 2020 were Black.

I know that tech companies in Nevada understand that by increasing the number of women and people of color in tech careers, we diversify the qualified labor pool that the U.S. relies on for innovation. This will help us maintain our global competitiveness and expand our economy, and I hope your companies redouble your efforts to this effect.

Question 4a. Can you discuss the full set of 2020 data on women and the people of color who work at your companies, and would you please discuss what you are doing to increase these numbers in 2021?

Answer. Google is committed to continuing to make diversity, equity, and inclusion part of everything we do—from how we build our products to how we build our workforce. Our recent diversity report (*https://diversity.google/*) shows how we've taken concrete actions to steadily grow a more representative workforce, to launch programs that support our communities globally, and to build products that better serve all of our users.

Among other efforts, we have recommitted to our company-wide objective in 2020: advance a diverse, accessible, and inclusive Google. Earlier this year, we announced our goal to improve leadership representation of underrepresented groups by 30 percent by 2025, increasing our investment in diverse talent markets such as Atlanta, Washington D.C., Chicago, and London.

Recently, we also expanded on our commitments, including setting a goal to spend $100 million with Black-owned businesses, as part of our broader commitment to spend a minimum of $1 billion with diverse-owned suppliers in the U.S., every year starting in 2021; committing to adding an additional 10,000 Googlers across our sites in Atlanta, Washington D.C., Chicago and New York; and building off our earlier commitment to increase leadership representation of underrepresented groups by adding a goal to more than double Black+ representation in the U.S. at all other levels by 2025. And in connection with our recently-announced Workplace Commitments, we will ensure that $310 million in funding goes toward diversity, equity, and inclusion initiatives and programs focused on increasing access to computer science education and careers; continuing to build a more representative workforce; fostering a respectful, equitable, and inclusive workplace culture; and helping businesses from underrepresented groups to succeed in the digital economy and tech industry.

But of course we know that there is much more to be done: only a holistic approach to these issues will produce meaningful, sustainable change. We must and will continue our work to expand the talent pool externally, and improve our culture internally, if we want to create equitable outcomes and inclusion for everyone. We understand the importance of this issue and remain committed to diversity, equity, and inclusion.

Question 4b. What are you doing more broadly to support STEM education programs and initiatives for women and people of color, including young girls of color?

Answer. We recognize the importance of supporting STEM education programs and initiatives for women and people of color, including young girls of color. Toward that end, Google is committing nearly $3 million to help close the racial equity gaps in computer science education and increase Black+ representation in STEM fields. This starts with making sure Black students have access to opportunities early on in their education. To that end, we're expanding our CS First curriculum to 7,000 more teachers who reach 100,000+ Black students (*https://csfirst.withgoogle.com/s/en/home*), scaling our Applied Digital Skills program to reach 400,000 Black middle and high school students (*https://applieddigitalskills.withgoogle.com/s/en/*

home), and making a $1 million Google.org grant to the DonorsChoose #SeeMe campaign, to help teachers access materials to make their classrooms more inclusive (*https://www.donorschoose.org/iseeme*).

Beyond the classroom, we're increasing our exploreCSR awards (*https://research.google/outreach/explore-csr*) to 16 more universities to address racial gaps in computing science research and academia, and we're also supporting Black in AI (*https://blackinai2020.vercel.app/*) with $250,000 to help increase Black representation in the field of AI.

These efforts build on our other education initiatives (*https://www.blog.google/inside-google/googlers/she-word/education-equity-team/*), including CodeNext, focused on cultivating the next generation of Black and Latinx tech leaders, and TechExchange, which partners with historically Black colleges and universities (HBCUs) and Hispanic-serving Institutions (HSIs) to bring students to Google's campus for four months to learn about topics from product management to machine learning. For more information, please see *https://blog.google/inside-google/company-announcements/commitments-racial-equity*. Code Next has now launched "Connect," a free, fully virtual computer science education program for Black and Latinx high school students that provides the skills and tech social capital needed to pursue long and high achieving careers in technology. For more information, please see *https://codenext.withgoogle.com/#welcome*.

Question 5. To continue being the most innovative country in the world, we need to maintain a workforce that can innovate. By 2026, the Department of Labor projects there will be 3.5 million computing-related jobs, yet our current education pipeline will only fill 19 percent of those openings. While other countries have prioritized STEM education as a national security issue, collaborating with non-profits and industry, the United States has mostly pursued an approach that does not meaningfully include such partnerships. The results of such a strategy are clear. A recent study found that less than half of K–12 students are getting any cyber related education, despite a growing demand for cyber professionals, both in national security fields and in the private sector.

Question 5a. What role can Google play in helping the United States boost its competitiveness in STEM fields, so that our economy can better compete with others around the globe?

Answer. As stated in our response to Question No. 4.b, we recognize the critical importance of supporting STEM education programs and initiatives. In addition to the initiatives detailed in our response to Question No. 4.b, we have also launched the Rising STEM Scholars Initiative with a $10 million contribution from Google.org. Through a partnership with Equal Opportunity Schools, UC Berkeley's Graduate School of Education, Kingmakers of Oakland, and Donorschoose.org, we'll collaborate with districts, schools, administrators, educators, students, and families to place and support 3,000 students of color and low income students in Bay Area AP STEM and CS classrooms. We'll also provide money for educators to get resources for their classrooms and find ways to inspire students to take AP courses. For more information, please see *https://blog.google/outreach-initiatives/google-org/10-million-increase-diversity-bay-area-stem-classrooms/*.

Google is also investing in students' cyber-related education. More than 65 percent of young people will work in jobs that don't currently exist. Learning computer science skills helps students thrive in a rapidly changing world. Yet our research with Gallup (*https://edu.google.com/latest-news/research*) shows that many students aren't getting the Computer Science education they need—and teachers don't have sufficient resources to provide it. Code with Google helps to ensure that every student has access to the collaborative, coding, and technical skills that unlock opportunities in the classroom and beyond–no matter what their future goals may be. For more information, please see *https://edu.google.com/code-with-google/*. Additionally, to help school districts provide more STEM opportunities to students, we offer a bundle of STEM tools on Chromebooks that are designed to help students become inventors and makers. For additional information, please see *https://edu.google.com/youchromebook/*.

We are proud of, and will continue, our work to support education through these products, programs, and philanthropy.

Response to Written Questions Submitted by Hon. Roger Wicker to Mark Zuckerberg

Question 1. During the COVID–19 pandemic, countless bad actors have propagated incorrect and unsafe information about the virus, including taking advantage

of unsuspecting Internet users by profiting off sales of unproven or fake COVID–19 "cures."

What steps has Facebook taken to crack down on this type of illegal behavior on its platform?

How many Facebook pages and groups have been removed upon identifying fraudulent COVID–19 claims, including the sale of illegal drugs through the platform?

How has Facebook Marketplace adjusted its algorithms and review processes to ensure illicit substances and unproven COVID products are not offered through the platform?

Answer. Facebook is supporting the global public health community's work to keep people safe and informed during the COVID–19 public health crisis. We're also working to address the pandemic's long-term impacts by supporting industries in need and making it easier for people to find and offer help in their communities. We've been prioritizing ensuring everyone has access to accurate information, removing harmful content, supporting health and economic relief efforts, and keeping people connected.

Under our Regulated Goods policy, we've also taken steps to protect against exploitation of this crisis for financial gain by banning content that attempts to sell or trade medical masks, hand sanitizer, surface-disinfecting wipes, and COVID–19 test kits. We also prohibit influencers from promoting such sales through branded content. From March through October 2020, we removed over 14 million pieces of content globally from Facebook and Instagram related to COVID–19 that violated our medical supply sales standards. Of these, over 370,000 were removed in the U.S.

In removing content that has the potential to contribute to real-world harm, we are also focusing on our policies related to commerce listings. We prohibit people from making health or medical claims related to COVID–19 in product listings on commerce surfaces, including those listings that guarantee a product will prevent someone from contracting COVID–19. We also prohibit the buying or selling of drugs and prescription products. When someone creates a listing on Marketplace, before it goes live, it is reviewed against our Commerce Policies using automated tools, and in some cases, further manual review. When we detect that a listing violates our policies, we reject it.

Question 2. What does "good faith" in Section 230 mean?

Is there any action you could take that could not be justified as done in "good faith"? Do you agree bad faith content moderation is not covered by Section 230?

If content is removed pretextually, or if terms and conditions are applied inconsistently depending on the viewpoint expressed in the content, is that removing content in good faith?

Answer. As we understand it, "good faith," as that term is used in Section 230(c)(2)(A), and as courts have been interpreting it for years, relates to a platform's subjective intent when it removes or restricts content. At Facebook, we are clear and transparent about what our standards are, and we seek to apply them to all of our users consistently.

Decisions about whether to remove content are based on whether the content violates our terms and policies, including our Community Standards. Our Community Standards are global, and all reviewers use the same guidelines when making decisions.

Question 3. Why wouldn't a platform be able to rely on terms of service to address categories of potentially harmful content outside of the explicit categories in Section 230(c)(2)? Why should platforms get the additional protections of Section 230 for removal of yet undefined categories of speech?

Does Section 230s "otherwise objectionable" catchall offer immunity for content moderation decisions motivated by political bias?

If the "otherwise objectionable" catchall does *not* offer such immunity, what limiting principle supports the conclusion that the catchall does not cover politically-biased moderation?

If the "otherwise objectionable" catchall *does* offer such immunity now, how would you rewrite Section 230 to deny immunity for politically-biased content moderation while retaining it for moderation of content that is harmful to children?

Answer. As we understand it, "otherwise objectionable," as the term is used in Section 230(c)(2)(A), is a standard that courts have interpreted for many years. At Facebook, our Community Standards—which are public—include restrictions around content that is harmful to members of our community, including bullying, harassment, hate speech, and incitement to violence.

At Facebook, we are a platform for ideas across the political and ideological spectrum, and we moderate content according to our published Community Standards to help keep users on the platform safe, reduce objectionable content, and ensure

users participate on the platform responsibly. We are clear and transparent about what our standards are, and we seek to apply them to all of our users consistently. The political affiliation of the user generating the content has no bearing on content removal assessments.

Regarding content that is harmful to children, Facebook's Community Standards prohibit coordinating harm and criminal activity, including posting content that sexually exploits or endangers children. When we become aware of apparent child exploitation, we report it to the National Center for Missing and Exploited Children ("NCMEC"), in compliance with applicable law. We work hard to identify and remove such content; over the past three years, we've found over 99 percent of the violating content we actioned before users reported it to us. And we certainly think it is important to make sure that platforms are serious about the illegal activity on their platforms.

Facebook supported SESTA/FOSTA, and we were very pleased to be able to work successfully with a bipartisan group of Senators on a bill that protects women and children from the harms of sex trafficking. We would welcome the opportunity to work with the Committee on proposals to modify Section 230 in ways that focus on bad actors, while being mindful not to disincentivize platforms from trying to find the illegal activity in the first place.

Question 4. Are your terms of service easy to understand and transparent about what is and is not permitted on your platform?

What notice and appeals process do you provide users when removing or labeling third-party speech?

What redress might a user have for improper content moderation beyond your internal appeals process?

In what way do your terms of service ensure against politically-biased content moderation and in what way do your terms of service limit your ability to moderate content on your platform?

How would you rewrite your terms of service to protect against politically-biased content moderation?

Do you think that removing content inconsistent with your terms of service and public representations is removal of content "in good faith"?

Answer. With respect to our Terms of Service, we believe that people should have clear, simple explanations of how online services work and use personal information. In June 2019, we updated our Terms of Service to clarify how Facebook makes money and to better explain the rights people have when using our services. The updates did not change any of our commitments or policies—they solely explained things more clearly. These updates are also part of our ongoing commitment to give people more transparency and control over their information.

When it comes to content moderation, we strive to enforce our policies consistently, without regard to political affiliation. Suppressing content on the basis of political viewpoint or preventing people from seeing what matters most to them directly contradicts Facebook's mission and our business objectives. Content reviewers assess content based on our Community Standards. We have made our detailed reviewer guidelines public to help people understand how and why we make decisions about the content that is and is not allowed on Facebook.

We also make appeals or "disagree with decision" feedback available for certain types of content that is removed from Facebook, when we have resources to review the appeals or feedback. We recognize that we sometimes make enforcement errors on both what we allow and what we remove, and that mistakes may cause significant concern for people. That's why we allow the option to request review of the decision when we can. This type of feedback will also allow us to continue improving our systems and processes so we can work with our partners and content reviewers to prevent similar mistakes in the future.

Facebook also recognizes that we should not make so many important decisions about free expression and safety on our own. With our size comes a great deal of responsibility, and while we have always taken advice from experts to inform our policies on how best to keep our platforms safe, until now, we have made the final decisions about what should be allowed on our platforms and what should be removed. And these decisions often are not easy to make; many judgments do not have obvious—or uncontroversial—outcomes, and yet they may have significant implications for free expression.

That's why we have created and empowered a new group, the Oversight Board, to exercise independent judgment over some of the most difficult and significant content decisions. In doing so, we've sought input from both critics and supporters of Facebook. We expect this Oversight Board to make some decisions that we, at Facebook, will not always agree with—but that's the point: Board Members are autonomous in their exercise of independent judgment. Facebook will implement the

Board's decisions unless doing so could violate the law, and we will respond constructively and in good faith to policy guidance put forth by the Board.

The Board won't be able to hear every case we or the public might want it to hear, but we look forward to working with the Board to ensure that its scope grows over time. As it does, we know the Board will play an increasingly important role in setting precedent and direction for content policy at Facebook. And in the long term, we hope its impact extends well beyond Facebook, and that it serves as a springboard for similar approaches to content governance in the online sphere.

Question 5. Please provide a list of all instances in which a prominent individual promoting liberal or left-wing views has been censored, demonetized, or flagged with extra context by your company.

Please provide a list of all instances in which a prominent individual promoting conservative or right-wing views has been censored, demonetized, or flagged with extra context by your company.

How many posts by government officials from Iran or China have been censored or flagged by your company?

How many posts critical of the Iranian or Communist Chinese government have been flagged or taken down?

Answer. As a general matter, when we identify or learn of content that violates our policies, we remove that content regardless of who posted it. The political affiliation of the user generating the content has no bearing on that content assessment. Rather, decisions about whether to remove content are based on our Community Standards, which direct all reviewers when making decisions. We seek to write actionable policies that clearly distinguish between violating and non-violating content, and we seek to make the decision-making process for reviewers as objective as possible.

In terms of moderation decisions, we have removed content posted by individuals and entities across the political spectrum. For example, we have taken down ads submitted on behalf of the Biden campaign and the Democratic National Committee, and organizations like the SEIU. We have also taken down ads submitted on behalf of the Trump campaign and the Republican National Committee, and organizations like the America First Action PAC.

We also remove content linked to coordinated inauthentic behavior campaigns, including those connected to state actors. When it comes to our influence operations investigations, we are often focused on the behavior, as opposed to the content, because that is the best way to stop the abuse; hence, our investigative work and enforcement are often location-and content-agnostic. We define coordinated inauthentic behavior as coordinated efforts to manipulate public debate for a strategic goal, where fake accounts are central to the operation. Our approach to coordinated inauthentic behavior and influence operations more broadly is grounded in behavior- and actor-based enforcement. This means that we are looking for specific violating behaviors exhibited by violating actors, rather than violating content (which is predicated on specific violations of our Community Standards, such as misinformation and hate speech). For a comprehensive overview of how we respond to inauthentic behavior, see *https://about.fb.com/news/2019/10/inauthentic-behavior-policy-update/*. For our most recent report sharing our findings about the coordinated inauthentic behavior we detected and removed from our platform, see our October 2020 Coordinated Inauthentic Behavior Report at *https://about.fb.com/news/2020/11/october-2020-cib-report/*.

Question 6. Should algorithms that promote or demote particular viewpoints be protected by Section 230? Why or why not?

Answer. On Facebook, people see posts from their friends, Pages they've chosen to follow, and Groups they've joined, among others, in their News Feed. On a given day, the number of eligible posts in a user's News Feed inventory can number in the thousands, so we use an algorithm to personalize how this content is organized. The goal of the News Feed algorithm is to predict what pieces of content are most relevant to the individual user, and rank (*i.e.,* order) those pieces of content accordingly every time a user opens Facebook, to try and bring those posts that are the most relevant to a person closer to the top of their News Feed. This ranking process has four main elements: the available inventory (all of the available content from the people, Pages, and Groups a person has chosen to connect with); the signals, or data points, that can inform ranking decisions (*e.g.,* who posted a particular piece of content); the predictions we make, including how likely we think a person is to comment on a story, share with a friend, etc.; and a relevancy score for each story. We've also taken steps to try and minimize the amount of divisive news content people see in News Feed, including by reducing the distribution of posts containing clickbait headlines.

As for our content moderation, we are clear and transparent about what our standards are, and we apply them to all of our users. We are a platform for ideas across the political and ideological spectrum, and we moderate content according to our published Community Standards in order to keep users on the platform safe, reduce objectionable content, and ensure users participate on the platform responsibly.

The debate about Section 230 shows that people of all political persuasions are unhappy with the status quo. People want to know that companies are taking responsibility for combatting harmful content—especially illegal activity—on their platforms. They want to know that when platforms remove content, they are doing so fairly and transparently. And they want to make sure that platforms are held accountable.

Section 230 made it possible for every major Internet service to be built and ensured that important values like free expression and openness were part of how platforms operate. Changing it is a significant decision. However, we believe Congress should update the law to make sure it's working as intended. We support the ideas around transparency and industry collaboration that are being discussed in some of the current bipartisan proposals, and we look forward to a meaningful dialogue about how we might update the law to deal with the problems we face today.

Do you think the use of an individual company's algorithms to amplify the spread of illicit or harmful materials like online child sexual exploitation should be protected by Section 230?

Answer. As discussed in the response to your Question 3, Facebook's Community Standards prohibit coordinating harm and criminal activity, including posting content that sexually exploits or endangers children. When we become aware of apparent child exploitation, we report it to NCMEC, in compliance with applicable law. We work hard to identify and remove such content; over the past three years, we've found over 99 percent of the violating content we actioned before users reported it to us. And we certainly think it is important to make sure that platforms are serious about addressing the illegal activity on their platforms.

Facebook supported SESTA/FOSTA, and we were very pleased to be able to work successfully with a bipartisan group of Senators on a bill that protects women and children from the harms of sex trafficking. We would welcome the opportunity to work with the Committee on proposals to modify Section 230 in ways that focus on bad actors, while being mindful not to disincentivize platforms from trying to find the illegal activity in the first place.

Question 7. Should platforms that knowingly facilitate or distribute Federal criminal activity or content be immune from civil liability? Why?/Why not?

Answer. Please see the responses to your Questions 3 and 6. Facebook has a variety of policies that prohibit the use of our platform for illegal activity or to share illegal content, including our policy against coordinating harm or publicizing crime, prohibitions on the sale of illegal goods, IP protections, and our policy against child sexual abuse material. We enforce these policies through a combination of human and automated review. We will continue working to improve our systems for finding violating content across a variety of categories, including illegal activity. As we did in the case of SESTA/FOSTA, and as indicated above, we would welcome the opportunity to work with the Committee on proposals to modify Section 230 in ways that focus on bad actors who intentionally facilitate wrongdoing, while being mindful not to disincentivize platforms from trying to find the illegal activity in the first place.

If your company has actual knowledge of content on your platform that incites violence, and your company fails to remove that content, should Federal law immunize your company from any claims that might otherwise be asserted against your company by victims of such violence? Are there limitations or exceptions to such immunity that you could propose for consideration by the Committee?

Answer. Please see the response to your previous question. Facebook prohibits incitement to violence on our platform. We remove content, disable accounts, and work with law enforcement when we believe there is a genuine risk of physical harm or direct threat to public safety.

Should platforms that are willfully blind to Federal criminal activity or content on their platforms be immune from civil liability? Why? Why not?

Answer. Please see the response to your previous question. We certainly think it is important to make sure that platforms are serious about addressing the illegal activity on their platforms. Facebook has a variety of policies that prohibit the use of our platform for illegal activity or to share illegal content, including our policy against coordinating harm or publicizing crime, prohibitions on the sale of illegal goods, IP protections, and our policy against child sexual abuse material. We enforce these policies rigorously through a combination of human and automated review.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. JOHN THUNE TO MARK ZUCKERBERG

Question 1. We have a public policy challenge to connect millions of Americans in rural America to broadband. I know you share in our commitment to connect every American household with broadband not only because it's the right thing to do but because it will add millions of new users to your platforms, which of course, means increase profits. What role should Congress and your companies play in ensuring that we meet all the broadband demands in rural America?

Answer. Facebook's ability to build communities and bring the world closer together depends on people being connected. Communities come in all sizes and across all regions, but many aren't currently being served by traditional methods of connectivity. Although hundreds of thousands have been connected in rural areas through programs like the FCC's High-Cost initiative, in many tribal and other rural areas it is still difficult for Internet service providers ("ISPs") to make the business case to serve sparsely populated, expansive geographic areas with difficult terrain. That leads to continued diminished access to broadband Internet for rural Americans. Through Facebook's connectivity efforts, we're working to help change that.

We're focused on developing next-generation technologies that can help bring the cost of connectivity down to reach the unconnected and increase capacity and performance for everyone. We know that there is no silver bullet for connecting the world; no single technology or program will get the job done. Rather than look for a one-size-fits-all solution, we are investing in a building block strategy—designing different technologies for specific use cases which are then used together to help connect people.

The COVID–19 pandemic in particular has underscored the importance of Internet connectivity. While many people have shifted their lives online, there are still more than 18 million Americans who lack reliable Internet access. To help, we have partnered with the Information Technology Disaster Resource Center ("ITDRC") and NetHope to provide Internet connectivity to communities most impacted by COVID–19. We also work with ISPs—including wireless ISPs—in rural areas. The goal of these partnerships is to better understand the unique barriers these communities face in getting online and to create the programs and infrastructure needed to increase the availability and affordability of high-quality internet access.

Question 2. Local news remains one of the most trusted news sources for individuals. Does Facebook's algorithm differentiate at all between news reported by a national or international source, and that of a local outlet?

Answer. We want Facebook to be a place where people can discover more news, information, and perspectives, and we are working to build products that help. Through our News Feed algorithm, we work hard to both actively reduce the distribution of clickbait, sensationalism, and misinformation and to boost news and information that keeps users informed, and we know the importance to users of staying informed about their local communities. As part of that effort, Facebook prioritizes local news on News Feed, so that people can see topics that have a direct impact on their community and discover what's happening in their local area.

We identify local publishers as those whose links are clicked on by readers in a tight geographic area. If a story is from a publisher in a user's area, and the user either follows the publisher's Page or the user's friend shares a story from that outlet, it might show up higher in News Feed. For more information, please visit *https://about.fb.com/news/2018/01/news-feed-fyi-local-news/*.

Our guiding principle is that journalism plays a critical role in our democracy. When news is deeply reported and well-sourced, it gives people information they can rely on to make good decisions. To that end, in January 2019, Facebook announced a $300 million investment in news programs, partnerships, and content, focused on supporting local news outlets and philanthropic efforts.

Question 3. The PACT Act would require your platforms to take down content that a court has ruled to be illegal. Do you support a court order-based takedown rule?

Answer. We support efforts aimed at greater transparency and external accountability. A court order-based takedown rule would provide this.

Question 4. Section 230 was initially adopted to provide a "shield" for young tech start-ups against the risk of overwhelming legal liability. Since then, however, some tech platforms like yours have grown larger than anyone could have imagined. Often a defense we hear from Section 230 proponents is that reform would hurt current and future start-ups. The PACT Act requires greater reporting from tech platforms on moderation decisions, largely exempts small business. However, your com-

panies are no longer start-ups, but rather some of the most powerful and profitable companies in the world.

Do tech giants need "shields" codified by the U.S. government? Have you outgrown your need for Section 230 protections?

Answer. Section 230 is a foundational law that allows us to provide our products and services to users. At a high level, Section 230 does two things. First, it encourages free expression. Without Section 230, platforms could potentially be held liable for everything people say. They would likely remove more content to avoid legal risk and would be less likely to invest in technologies that enable people to express themselves in new ways. Second, it allows platforms to remove harmful content. Without Section 230, platforms could face liability for doing even basic moderation, such as removing bullying and harassment that impact the safety and security of their communities.

Section 230 made it possible for every major Internet service to be built and ensured important values like free expression and openness were part of how platforms operate. As the Internet keeps growing and evolving, the core principles of Section 230 will continue to be crucial for innovation—for small platforms that don't have the same capabilities when it comes to content moderation, for large ones that host billions of pieces of content across the globe, and for the American tech sector as a whole if we are going to maintain our edge in innovation. But that doesn't mean it shouldn't be updated to reflect the way the Internet has changed in the last 25 years, and that's why we support thoughtful reform to make sure the law is working as intended.

Question 5. Last year, I introduced the Filter Bubble Transparency Act to address the filter bubble phenomena, in which social media users are only shown content they agree with. This is believed to be leading to ideological isolation and increased polarization, as illustrated in a recent documentary called "The Social Dilemma". In response to that documentary, Mr. Zuckerberg, your company stated that "polarization and populism have existed long before Facebook" and that the platform "takes steps to reduce content that could drive polarization."

Mr. Zuckerberg, do you believe the filter bubble exists, and do you believe Facebook's use of algorithms is contributing to polarization?

Answer. We know that one of the biggest issues social networks face is that, when left unchecked, people will engage disproportionately with more sensationalist and provocative content. At scale this type of content can undermine the quality of public discourse and lead to polarization. In our case, it can also degrade the quality of our services. Our research suggests that no matter where we draw the line for what is allowed, as a piece of content gets close to that line, people will engage with it more on average—even when they tell us afterwards they don't like the content. That is why we've invested heavily and have taken steps to try and minimize the amount of divisive news content people see in News Feed, including by reducing the distribution of posts containing clickbait headlines.

On Facebook, people see posts from their friends, Pages they've chosen to follow, and Groups they've joined, among others, in their News Feed. On a given day, the number of eligible posts in a user's News Feed inventory can number in the thousands, so we use an algorithm to personalize how this content is organized. The goal of the News Feed algorithm is to predict what pieces of content are most relevant to the individual user, and rank (*i.e.,* order) those pieces of content accordingly every time a user opens Facebook, to try and bring those posts that are the most relevant to a person closer to the top of their News Feed. This ranking process has four main elements: the available inventory (all of the available content from the people, Pages, and Groups a person has chosen to connect with); the signals, or data points, that can inform ranking decisions (*e.g.,* who posted a particular piece of content); the predictions we make, including how likely we think a person is to comment on a story, share with a friend, etc.; and a relevancy score for each story, which informs its position in News Feed.

We frequently make changes to the algorithm that drives News Feed ranking in an effort to improve people's experience on Facebook. For example, in 2018, we responded to feedback from our community that public content—posts from businesses, brands, and media—was crowding out the personal moments that lead us to connect more with each other. As a result, we moved from focusing only on helping people find relevant content to helping them have more meaningful social interactions. This meant that people began seeing more content from their friends, family, and Groups. We also reduce the distribution of some problematic types of content, including content that users may find spammy or low-quality, such as clickbait headlines, misinformation as confirmed by third-party fact-checkers, and links to low-quality webpages like ad farms.

Facebook is a platform that reflects the conversations already taking place in society. We are keenly aware of the concern that our platform is contributing to polarization, and we have been working to understand the role that we play in discourse and information diversity. The data on what causes polarization and "filter bubbles" is mixed. Some independent research has shown that social media platforms provide more information diversity than traditional media, and our own research indicates that most people on Facebook have at least some friends who claim an opposing political ideology—probably because Facebook helps people maintain ties with people who are more distantly connected to them than their core community—and that the content in News Feed reflects that added diversity. We want Facebook to be a place where people can discover more news, information, and perspectives, and we are working to build products that help. And because we want Facebook to be a place where people can express themselves, we must also preserve our community's sense of safety, privacy, dignity, and authenticity via our Community Standards, which define what is and isn't allowed on Facebook. We remove content that violates our Community Standards, such as hate speech, bullying, and harassment.

With respect to your legislation, S. 2763, we believe we are compliant with the bill's proposed requirement that users be provided with the opportunity to choose a chronological feed. Users who do not wish to consume ranked News Feed have access to a control to view content chronologically from those they follow in the "Most Recent" News Feed view (see *https://www.facebook.com/help/218728138156311*).

Question 6. As discussed during the hearing, please provide for the record a complete list of U.S. newspaper articles that Facebook suppressed or limited the distribution of over the past five years, as Facebook did with the October 14, 2020 *New York Post* article entitled "Smoking-Gun E-mail Reveals How Hunter Biden Introduced Ukrainian Businessman to VP Dad." For each article listed, please also provide an explanation why the article was suppressed or the distribution was limited.

Answer. People often tell us they don't want to see misinformation. People also tell us that they don't want Facebook to be the arbiter of truth or falsity. That's why we work with over 80 independent, third-party fact-checkers who are certified through the non-partisan International Fact-Checking Network ("IFCN") to help identify and review false news. If content is deemed by a fact-checker to be False, Altered, or Partly False, according to our public definitions, its distribution will be reduced, and it will appear lower in News Feed. We also implement an overlaid warning screen on top of fact-checked content. People who try to share the content will be notified of the fact-checker's reporting and rating, and they will also be notified if content they have shared in the past has since been rated false by a fact-checker.

We also work to take fast action to prevent misinformation from going viral, especially given that quality reporting and fact-checking takes time. In 2019, we announced that, if we identify signals that a piece of content is false, we will temporarily reduce its distribution in order to allow sufficient time for our independent, third-party fact-checkers to review and determine whether to apply a rating. Quick action is critical in keeping a false claim from going viral, so we take this step to provide an extra level of protection against potential misinformation. These temporary demotions expire after seven days if the content has not been rated by an independent fact-checker.

We believe it's important for the fact-checking process to be transparent, so Page and domain owners will receive a notification when content they shared is rated by a fact-checking partner. Page owners can also review all violations, including Community Standards violations, in their Page Quality tab. Additionally, the third-party fact-checkers with which we work are all signatories to the IFCN's Code of Principles, which requires transparency of sources and methodology and a commitment to open and honest corrections. To that end, our partners' fact-checking articles are publicly available and easily accessible at their websites. For a list of our third-party fact-checkers in the U.S., please visit *https://www.facebook.com/journalismproject/programs/third-party-fact-checking/partner-map*.

Regarding the October 14 *New York Post* story, for the past several months, the U.S. intelligence community has urged voters, companies, and the Federal government to remain vigilant in the face of the threat of foreign influence operations seeking to undermine our democracy and the integrity of our electoral process. For example, the Director of National Intelligence, the Head of the FBI, and the bipartisan leaders of the Senate Select Committee on Intelligence reminded Americans about the threat posed by foreign influence operations emanating from Russia and Iran. Along with their public warnings, and as part of the ongoing cooperation that tech companies established with government partners following the 2016 election, the FBI also privately warned tech companies to be on high alert for the potential

of hack-and-leak operations carried out by foreign actors in the weeks leading up to the Presidential election. We took these risks seriously.

Given the concerns raised by the FBI and others, we took steps consistent with our policies to slow the spread of suspicious content and provide fact-checkers the opportunity to assess it. However, at no point did we take any action to block or remove the content from the platform. People could—and did—read and share the *Post*'s reporting while we had this temporary demotion in place. Consistent with our policy, after seven days, we lifted the temporary demotion on this content because it was not rated false by an independent fact-checker.

Question 7. Justice Thomas recently observed that "[p]aring back the sweeping immunity courts have read into § 230 would not necessarily render defendants liable for online misconduct. It simply would give plaintiffs a chance to raise their claims in the first place. Plaintiffs still must prove the merits of their cases, and some claims will undoubtedly fail." Do you agree with him? Why shouldn't lawsuits alleging that a tech platform has violated a law by exercising editorial discretion be evaluated on the merits rather than being dismissed because a defendant invokes Section 230 as a broad shield from liability?

Answer. We want to engage productively in conversations about how to update Section 230 to make sure it's working as intended. As discussed in response to your Question 4, Section 230 made it possible for every major Internet service to be built and ensured that important values like free expression and openness were part of how platforms operate. Changing it is a significant decision. However, we believe Congress should update the law to make sure it's working as intended. We support the ideas around transparency and industry collaboration that are being discussed in some of the current bipartisan proposals, and we look forward to a meaningful dialogue about how we might update the law to deal with the problems we face today.

Question 8. What does "good faith" in Section 230 mean? Is there any action you could take that could not be justified as done in "good faith"? Do you agree bad faith content moderation is not covered by Section 230? If content is removed pretextually, or if terms and conditions are applied inconsistently depending on the viewpoint expressed in the content, is that removing content in good faith?

Answer. As we understand it, "good faith," as that term is used in Section 230(c)(2)(A), and as courts have been interpreting it for years, relates to a platform's subjective intent when it removes or restricts content. At Facebook, we are clear and transparent about what our standards are, and we seek to apply them to all of our users consistently.

Decisions about whether to remove content are based on whether the content violates our terms and policies, including our Community Standards. Our Community Standards are global, and all reviewers use the same guidelines when making decisions.

Question 9. Mr. Pichai noted in the hearing that without the "otherwise objectionable" language of Section 230, the suppression of teenagers eating tide pods, cyberbullying, and other dangerous trends would have been impossible. Could the language of Section 230 be amended to specifically address these concerns, by including the language of "promoting self-harm" or "unlawful" without needing the "otherwise objectionable" language that provides online platforms a blank check to take down any third-party speech with which they disagree?

Answer. We certainly understand the tension between free expression and safety; we grapple with this tension every day. We think there are other kinds of content that are also harmful that may not be covered by proposals like the ones you reference, such as incitement to violence, bullying, and harassment. We do not think we should be subject to costly litigation for removal of harmful content. As Mr. Zuckerberg said during the hearing, we would worry that some of the proposals that suggest getting rid of the phrase "otherwise objectionable" from Section 230 would limit our ability to remove bullying and harassing content from our platforms, which we think would make them worse places for people.

However, we do believe Congress should update the law to make sure it's working as intended. We support the ideas around transparency and industry collaboration that are being discussed in some of the current bipartisan proposals, and we look forward to a meaningful dialogue about how we might update the law to deal with the problems we face today.

Question 10. What other language would be necessary to address truly harmful material online without needing to rely on the vague term "otherwise objectionable?"

Answer. Please see the response to your Question 9.

Question 11. Why wouldn't a platform be able to rely on terms of service to address categories of potentially harmful content outside of the explicit categories in

Section 230(c)(2)? Why should platforms get the additional protections of Section 230 for removal of yet undefined categories of speech?

Answer. As we understand it, "otherwise objectionable," as the term is used in Section 230(c)(2)(A), is a standard that courts have interpreted for many years. At Facebook, our Community Standards—which are public—include restrictions around content that is harmful to members of our community, including bullying, harassment, hate speech, and incitement to violence.

At Facebook, we are a platform for ideas across the political and ideological spectrum, and we moderate content according to our published Community Standards to help keep users on the platform safe, reduce objectionable content, and ensure users participate on the platform responsibly. We are clear and transparent about what our standards are, and we seek to apply them to all of our users consistently. The political affiliation of the user generating the content has no bearing on content removal assessments.

Regarding content that is harmful to children, Facebook's Community Standards prohibit coordinating harm and criminal activity, including posting content that sexually exploits or endangers children. When we become aware of apparent child exploitation, we report it to the National Center for Missing and Exploited Children ("NCMEC"), in compliance with applicable law. We work hard to identify and remove such content; over the past three years, we've found over 99 percent of the violating content we actioned before users reported it to us. And we certainly think it is important to make sure that platforms are serious about the illegal activity on their platforms.

Facebook supported SESTA/FOSTA, and we were very pleased to be able to work successfully with a bipartisan group of Senators on a bill that protects women and children from the harms of sex trafficking. We would welcome the opportunity to work with the Committee on proposals to modify Section 230 in ways that focus on bad actors, while being mindful not to disincentivize platforms from trying to find the illegal activity in the first place.

Question 12. Does Section 230s "otherwise objectionable" catchall offer immunity for content moderation decisions motivated by political bias?

If the "otherwise objectionable" catchall does *not* offer such immunity, what limiting principle supports the conclusion that the catchall does not cover politically-biased moderation?

If the "otherwise objectionable" catchall *does* offer such immunity now, how would you rewrite Section 230 to deny immunity for politically-biased content moderation while retaining it for moderation of content that is harmful to children?

Answer. Please see the response to your previous question.

Question 13. Are your terms of service easy to understand and transparent about what is and is not permitted on your platform?

Answer. We believe that people should have clear, simple explanations of how online services work and use personal information. In June 2019, we updated our Terms of Service to clarify how Facebook makes money and better explain the rights people have when using our services. The updates did not change any of our commitments or policies—they solely explained things more clearly. These updates are also part of our ongoing commitment to give people more transparency and control over their information.

When it comes to content moderation, we strive to enforce our policies consistently, without regard to political affiliation.

Question 14. What notice and appeals process do you provide users when removing or labeling third-party speech?

Answer. We strive to enforce our policies consistently, without regard to political affiliation. Suppressing content on the basis of political viewpoint or preventing people from seeing what matters most to them directly contradicts Facebook's mission and our business objectives. Content reviewers assess content based on our Community Standards. We have made our detailed reviewer guidelines public to help people understand how and why we make decisions about the content that is and is not allowed on Facebook.

We also make appeals or "disagree with decision" feedback available for certain types of content that is removed from Facebook when we have resources to review the appeals or feedback. We recognize that we sometimes make enforcement errors on both what we allow and what we remove, and that mistakes may cause significant concern for people. That's why we need to allow the option to request review of the decision when we can. This type of feedback will also allow us to continue improving our systems and processes so we can work with our partners and content reviewers to prevent similar mistakes in the future.

Facebook also recognizes that we should not make so many important decisions about free expression and safety on our own. With our size comes a great deal of responsibility and while we have always taken advice from experts to inform our policies and on how best to keep our platforms safe, until now, we have made the final decisions about what should be allowed on our platforms and what should be removed. And these decisions often are not easy to make—many judgments do not have obvious, or uncontroversial, outcomes and yet they may have significant implications for free expression.

That's why we have created and empowered a new group, the Oversight Board, to exercise independent judgment over some of the most difficult and significant content decisions. In doing so, we've sought input from both critics and supporters of Facebook. We expect this Oversight Board to make some decisions that we, at Facebook, will not always agree with—but that's the point: they are autonomous in their exercise of independent judgment. Facebook will implement the Board's decisions unless doing so could violate the law, and will respond constructively and in good faith to policy guidance put forth by the Board.

The board won't be able to hear every case we or the public might want it to hear, but we look forward to working with the board to ensure that its scope grows over time. As it does, we know the board will play an increasingly important role in setting precedent and direction for content policy at Facebook. And in the long term, we hope its impact extends well beyond Facebook, and serves as a springboard for similar approaches to content governance in the online sphere.

Question 15. What redress might a user have for improper content moderation beyond your internal appeals process?

Answer. Please see the response to your previous question.

Question 16. In what way do your terms of service ensure against politically-biased content moderation and in what way do your terms of service limit your ability to moderate content on your platform?

Answer. Please see the responses to your Questions 13 and 14.

Question 17. How would you rewrite your terms of service to protect against politically-biased content moderation?

Answer. Please see the responses to your Questions 13 and 14.

Question 18. Do you think that removing content inconsistent with your terms of service and public representations is removal of content "in good faith"?

Answer. Please see the responses to your Questions 8, 13, and 14.

Question 19. As it stands, Section 230 has been interpreted not to grant immunity if a publishing platform "ratifies" illicit activity. Do you agree? How do you think "ratification" should be defined?

Answer. Section 230 protects platforms from liability related to content created by others, not by itself. Platforms do not (and should not) lose Section 230 protection for content created by others simply because they choose to speak for themselves in certain circumstances.

Question 20. Do you agree that a platform should not be covered by Section 230 if it adds its own speech to third-party content?

Answer. Please see the response to your previous question.

Question 21. When a platform adds its own speech, does it become an information content provider under Section 230(f)(3)?

Answer. Please see the response to your Question 19.

Question 22. Should algorithms that promote or demote particular viewpoints be protected by Section 230? Why or why not?

A*nswer.* On Facebook, people see posts from their friends, Pages they've chosen to follow, and Groups they've joined, among others, in their News Feed. On a given day, the number of eligible posts in a user's Feed inventory can number in the thousands, so we use an algorithm to personalize how this content is organized. The goal of the News Feed algorithm is to predict what pieces of content are most relevant to the individual user, and rank (*i.e.,* order) those pieces of content accordingly every time a user opens Facebook, to try and bring those posts that are the most relevant to a person closer to the top of their News Feed. This ranking process has four main elements: the available inventory (all of the available content from the people, Pages, and Groups a person has chosen to connect with); the signals, or data points, that can inform ranking decisions (*e.g.,* who posted a particular piece of content); the predictions we make, including how likely we think a person is to comment on a story, share with a friend, etc.; and a relevancy score for each story. We've also taken steps to try and minimize the amount of divisive news content people see in News Feed, including by reducing the distribution of posts containing clickbait headlines.

As for our content moderation, we are clear and transparent about what our standards are, and we apply them to all of our users. We are a platform for ideas across the political and ideological spectrum, and we moderate content according to our published Community Standards in order to keep users on the platform safe, reduce objectionable content, and ensure users participate on the platform responsibly.

The debate about Section 230 shows that people of all political persuasions are unhappy with the status quo. People want to know that companies are taking responsibility for combatting harmful content—especially illegal activity—on their platforms. They want to know that when platforms remove content, they are doing so fairly and transparently. And they want to make sure that platforms are held accountable.

Section 230 made it possible for every major Internet service to be built and ensured important values like free expression and openness were part of how platforms operate. Changing it is a significant decision. However, we believe Congress should update the law to make sure it's working as intended. We support the ideas around transparency and industry collaboration that are being discussed in some of the current bipartisan proposals, and we look forward to a meaningful dialogue about how we might update the law to deal with the problems we face today.

Question 23. Do you think the use of an individual company's algorithms to amplify the spread of illicit or harmful materials like online child sexual exploitation should be protected by Section 230?

Answer. Facebook's Community Standards prohibit coordinating harm and criminal activity, including posting content that sexually exploits or endangers children. When we become aware of apparent child exploitation, we report it to the National Center for Missing and Exploited Children (NCMEC), in compliance with applicable law. We work hard to identify and remove such content; over the past three years, we've found over 99 percent of the violating content we actioned before users reported it to us. And we certainly think it is important to make sure that platforms are serious about addressing the illegal activity on their platforms.

Facebook supported SESTA/FOSTA, and we were very pleased to be able to work successfully with a bipartisan group of Senators on a bill that protects women and children from the harms of sex trafficking. We would welcome the opportunity to work with the Committee on proposals to modify Section 230 in ways that focus on bad actors, while being mindful not to disincentivize platforms from trying to find the illegal activity in the first place.

Question 24. Should platforms that knowingly facilitate or distribute Federal criminal activity or content be immune from civil liability? Why or why not?

Answer. Please see the responses to your Questions 11, 12, 22, and 23. Facebook has a variety of policies that prohibit the use of our platform for illegal activity or to share illegal content, including our policy against coordinating harm or publicizing crime, prohibitions on the sale of illegal goods, IP protections, and our policy against child sexual abuse material. We enforce these policies through a combination of human and automated review. We will continue working to improve our systems for finding violating content across a variety of categories, including illegal activity. As we did in the case of SESTA/FOSTA, and as indicated above, we would welcome the opportunity to work with the Committee on proposals to modify Section 230 in ways that focus on bad actors who intentionally facilitate wrongdoing, while being mindful not to disincentivize platforms from trying to find the illegal activity in the first place.

Question 25. If your company has actual knowledge of content on your platform that incites violence, and your company fails to remove that content, should Federal law immunize your company from any claims that might otherwise be asserted against your company by victims of such violence? Are there limitations or exceptions to such immunity that you could propose for consideration by the Committee?

Answer. Please see the responses to your Questions 22 and 23. Facebook prohibits incitement to violence on our platform. We remove content, disable accounts, and work with law enforcement when we believe there is a genuine risk of physical harm or direct threat to public safety.

Question 26. Should platforms that are willfully blind to Federal criminal activity or content on their platforms be immune from civil liability? Why or why not?

Answer. Please see the responses to your Questions 22 and 23. We certainly think it is important to make sure that platforms are serious about addressing the illegal activity on their platforms. Facebook has a variety of policies that prohibit the use of our platform for illegal activity or to share illegal content, including our policy against coordinating harm or publicizing crime, prohibitions on the sale of illegal

goods, IP protections, and our policy against child sexual abuse material. We enforce these policies rigorously through a combination of human and automated review.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. JERRY MORAN TO MARK ZUCKERBERG

Question 1. How much money does your company spend annually on content moderation in general?

Answer. We're spending as much—if not more—on safety and security as the entire revenue of our company at the time of our IPO earlier this decade. And we now have over 35,000 people working in this area, about 15,000 of whom review content.

Question 2. How many employees does your company have that are involved with content moderation in general? In addition, how many outside contractors does your company employ for these purposes?

Answer. As discussed in the response to your Question 1, we have over 35,000 people working on safety and security, about 15,000 of whom review content. The majority of our content reviewers are people who work full-time for our partners and work at sites managed by these partners. We have a global network of partner companies so that we can quickly adjust the focus of our workforce as needed. This approach gives us the ability to, for example, make sure we have the right language or regional expertise. Our partners have a core competency in this type of work and are able to help us adjust as new needs arise or when a situation around the world warrants it.

Question 3. How much money does your company currently spend on defending lawsuits stemming from users' content on your platform?

Answer. Defending lawsuits related to users' content on our platform requires a substantial amount of resources, including litigation costs and employee time, both in the U.S. and elsewhere.

Question 4. Without Section 230s liability shield, would your legal and content moderation costs be higher or lower?

Answer. Broadly speaking, Section 230 does two things. First, it encourages free expression. Without Section 230, platforms could potentially be held liable for everything people say. Platforms would likely moderate more content to avoid legal risk and would be less likely to invest in technologies that enable people to express themselves in new ways. Second, it allows platforms to moderate content. Without Section 230, platforms could face liability for doing even basic moderation, such as removing hate speech and harassment that impact the safety and security of their communities. Repealing Section 230 entirely would likely substantially increase many companies' costs associated with legal challenges and content moderation.

Question 5. How many liability lawsuits have been filed against your company based on user content over the past year?

Answer. Please see the response to your Question 3.

Question 6. Please describe the general breakdown of categories of liability, such as defamation, involved in the total number of lawsuits over the past year.

Answer. Lawsuits based on user content may include claims that we took down content improperly, or that we left up content that we should have taken down (for example, because it was allegedly illegal, defamatory, or otherwise harmful).

Question 7. Of the total number of liability lawsuits based on user content, how many of them did your company rely on Section 230 in its defense?

Answer. We don't have a precise number. We may invoke Section 230 in our defense when a claim seeks to treat Facebook as the publisher or speaker of information provided by a user or other entity.

Question 8. Of the liability lawsuits based on user content in which your company relies on Section 230 in its defense, what categories of liability in each of these lawsuits is your company subject to?

Answer. Please see the responses to your Questions 6 and 7.

Question 9. In a defamation case based on user content, please describe the typical procedural steps your company takes to litigate these claims.

Answer. The lawsuits the company faces—and therefore the procedural steps the company takes to defend them—vary based on the nature of the claims, facts alleged, relevant legal and procedural standards, and fora, among other factors.

Question 10. Of the claims that have been dismissed on Section 230 grounds, what is the average cost of litigation?

Answer. The costs of litigation are often substantial, even when the suits are dismissed on Section 230 grounds.

Question 11. I understand the U.S.-Mexico-Canada Agreement (USMCA) contains similar intermediary liability protections that Section 230 established domestically. The recent trade deal with Japan also included similar provisions.

If Congress were to alter Section 230, do you expect litigation or free trade agreement compliance issues related to the United States upholding trade agreements that contain those provisions?

Answer. Facebook is not in a position to comment on the type of litigation described in this question.

Question 12. How does the inclusion of Section 230-like protections in the aforementioned trade deals affect your business operations in the countries party to said trade deals? Do you expect fewer defamation lawsuits and lower legal costs associated with intermediary liability in those countries due to these trade deals?

Answer. It's too early to tell. These countries and their respective legal systems implement and enforce international agreements differently and we have yet to see how they will do so here.

Question 13. In countries that do not have Section 230-like protections, are your companies more vulnerable to litigation or liability as a result?

Answer. In countries that do not have laws analogous to Section 230, Facebook has faced litigation based on content moderation decisions or seeking to hold Facebook responsible for content posted by our users. These cases involve substantial litigation costs.

Question 14. How do your content moderation and litigation costs differ in these countries compared to what you might expect if Section 230-like protections were in place?

Answer. Please see the response to your previous question.

Question 15. As American companies, does Section 230s existence provide you any liability protection overseas in countries that do not have similar protections for tech companies?

Answer. Unless specified in trade agreements, or unless overseas courts apply California law, we are not aware of any liability protection that Section 230 provides in countries other than the U.S.

Question 16. To differing extents, all of your companies rely on automated content moderation tools to flag and remove content on your platforms.

What is the difference in effectiveness between automated and human moderation?

Answer. To enforce our Community Standards, we have introduced tools that allow us to proactively detect and remove certain violating content using advances in technology, including artificial intelligence, machine learning, and computer vision. We do this by analyzing specific examples of bad content that have been reported and removed to identify patterns of behavior. Those patterns can be used to teach our software to proactively identify similar content.

These advances in technology mean that we can now remove bad content more quickly, identify and review more potentially harmful content, and increase the capacity of our review team. To ensure the accuracy of these technologies, we constantly test and analyze our systems, technology, and AI to ensure accuracy. All content goes through some degree of automated review, and we use human reviewers to check some content that has been flagged by that automated review or reported by people that use Facebook. We also use human reviewers to perform reviews of content that was not flagged or reported by people to check the accuracy and efficiency of our automated review systems. The percentage of content that is reviewed by a human varies widely depending on the type and context of the content, and we don't target a specific percentage across all content on Facebook.

Question 17. What percentage of decisions made by automated content moderation systems are successfully appealed, and how does that compare to human moderation decisions?

Please describe the limitations and benefits specific to automated content moderation and human content moderation.

Answer. For information about automated and human content moderation, please see the response to your previous question.

With respect to appeals, we release our Community Standards Enforcement Report (available at *https://transparency.facebook.com/community-standards-enforcement*) on a quarterly basis to report on our progress and demonstrate our continued commitment to making Facebook safe and inclusive. This report shares metrics on how Facebook is performing in preventing and removing content that violates our Community Standards. We also share data in this report on our process for appealing and restoring content to correct mistakes in our enforcement decisions.

Question 18. In your written testimonies, each of you note the importance of tech companies being transparent with their users.

Have you already, or do you plan to make public the processes that your automated moderation system undertakes when making decisions about content on your platform?

Given the complexity of the algorithms that are now governing a portion of the content across your platforms, how have you or how do you plan to explain the functions of your automated moderation systems in a simple manner that users can easily understand?

Answer. An algorithm is a formula or set of steps for solving a particular problem. At Facebook, we use algorithms to offer customized user experiences and to help us achieve our mission of building a global and informed community. For example, we use algorithms to help generate and display search results (see *https://about.fb.com/news/2018/11/inside-feed-how-search-works/*), to prioritize the content people follow with their personalized News Feed (see *https://about.fb.com/news/2018/05/inside-feed-news-feed-ranking/*), and to serve ads that may be relevant to them.

As a company, we are committed to helping our users understand how we use algorithms. We publish a series of blog posts called News Feed FYI (see *https://about.fb.com/news/category/news-feed-fyi/*) that highlight major updates to News Feed and explain the thinking behind them. Also, in 2019, we launched a feature called "Why am I seeing this post?" (see *https://about.fb.com/news/2019/03/why-am-i-seeing-this/*). This feature directly responded to user feedback asking for more transparency around why certain content appears in News Feed and easier access to News Feed controls. Through their News Feed Preferences and our See First tool, users can choose to see posts from certain friends and Pages higher up in their News Feed. Controls also include Snooze, which keeps the content from a selected person, Page, or Group out of a user's News Feed for a limited time.

We also maintain a blog focused exclusively on our artificial intelligence work at *https://ai.facebook.com/*. Most recently, we published a series of posts on our use of artificial intelligence to help protect users from harmful content:

- How we use AI to help detect misinformation (*https://ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/*)
- How we train AI to detect hate speech (*https://ai.facebook.com/blog/training-ai-to-detect-hate-speech-in-the-real-world/*)
- How AI is getting better at detecting hate speech (*https://ai.facebook.com/blog/how-ai-is-getting-better-at-detecting-hate-speech/*)
- How we use super-efficient AI models to detect hate speech (*https://ai.facebook.com/blog/how-facebook-uses-super-efficient-ai-models-to-detect-hate-speech/*)

Question 19. How has COVID–19 impacted your company's content moderation systems?

Is there a greater reliance on automated content moderation?

Please quantify how content moderation responsibilities have shifted between human and automated systems due to COVID–19.

Answer. Throughout the COVID–19 crisis, we've worked to keep both our workforce and the people who use our platforms safe. In March, we announced that we would temporarily send our content reviewers home. Since then we've made some changes to keep our platform safe during this time, including increasing the use of automation, carefully prioritizing user reports, and temporarily altering our appeals process. We also asked some of our full-time employees to review content related to real-world harm like child safety, suicide, and self-injury.

Question 20. Last year, the Department of Justice's Antitrust Division held a workshop that brought together academics and executives from leading companies, including buyers and sellers of advertising inventory. The discussion explored the practical considerations that industry participants face and the competitive impact of technological developments such as digital and targeted advertising in media markets, including dynamics between local broadcast and online platforms for advertisement expenditures.

Separately, the FCC has attempted to update its local broadcast ownership rules following its 2018 quadrennial review, including permitting the ownership of two TV stations in local markets. However, this recent attempt by the FCC to modernize the local media marketplace has been halted by the Third Circuit's decision to reject the FCC's update of broadcast ownership restrictions.

For purposes of understanding your companies' general views on the local media marketplace, do your companies compete with local broadcast stations for digital advertising revenue?

Do you think Federal regulations determining acceptable business transactions in local media marketplaces should be updated to account for this evolving and increasing competition for digital advertising purchases?

Answer. The advertising sector is incredibly dynamic, and competition for advertising spend is increasingly fierce. Companies have more options than ever when deciding where to advertise. Unlike a few decades ago, when companies had more limited options, today there are more choices, different channels and platforms, and hundreds of companies offering them.

Facebook competes for advertisers' budgets with online and offline advertisers and with a broad variety of advertising players. This includes the intense competitive pressure that Facebook faces for advertising budgets from offline channels (such as print, radio, and broadcast), established digital platforms (such as Google, Amazon, Twitter, and Pinterest), and newer entrants that have attracted a large user base from scratch (such as Snap and TikTok). The landscape is also highly dynamic, with offline advertising channels (such as television and radio) benefiting from industry-wide digitalization and new technologies to offer their own ad targeting and measurement products.

Advertisers can and do shift spend in real time across ad platforms to maximize their return on investment. As a result of this competition and choice, advertisers spread their budgets across multiple outlets and channels, including Facebook.

Facebook is able to provide nearly all of its consumer services free of charge because it is funded by advertising that is relevant and useful. Millions of Americans use Facebook to connect with the people, organizations, and businesses they care about. Research has shown that though Facebook offers these services at no cost, they offer significant value—a huge consumer surplus.

Question 21. Earlier this year, you stated that due to increased reliance on automated review of some content due to COVID–19 means "we may be a little less effective in the near term. . ."

Please state the current status of your content moderation preparedness, especially in regard to your election preparations. Did Facebook experience content moderation difficulties during the election due to changes made due to COVID–19?

Answer. We're gratified that, thanks to the hard work of election administrators across the country, the voting process went relatively smoothly. Facebook worked hard to do our part in protecting the integrity of the 2020 election, and we're proud of the work we've done to support our democracy. For example, we ran the largest voting information campaign in American history. Based on conversion rates we calculated from a few states we partnered with, we estimate that we helped 4.5 million people register to vote across Facebook, Instagram, and Messenger—and helped about 100,000 people sign up to be poll workers. We launched a Voting Information Center to connect people with reliable information on deadlines for registering and voting and details about how to vote by mail or vote early in person, and we displayed links to the Voting Information Center when people posted about voting on Facebook. More than 140 million people have visited the Voting Information Center on Facebook and Instagram since it launched. We are encouraged that more Americans voted in 2020 than ever before, and that our platform helped people take part in the democratic process.

We also worked to tackle misinformation and voter suppression. We displayed warnings on more than 150 million pieces of content that our third-party fact-checkers debunked. We partnered with election officials to remove false claims about polling conditions, and we put in place strong voter suppression policies that prohibit explicit or implicit misrepresentations about how or when to vote, as well as attempts to use threats related to COVID–19 to scare people into not voting. We removed calls for people to engage in voter intimidation that used militarized language or suggested that the goal is to intimidate, exert control, or display power over election officials or voters. In addition, we blocked new political and issue ads during the final week of the campaign, as well as all political and issue ads after the polls closed on election night.

We also instituted a variety of measures to help in the days and weeks after voting ended:

- We used the Voting Information Center to prepare people for the possibility that it could take a while to get official results. This information helped people understand that there was nothing illegitimate about not having a result on election night.
- We partnered with Reuters and the National Election Pool to provide reliable information about election results. We displayed this in the Voting Information Center, and we notified people proactively as results became available. We added labels to any post by a candidate or campaign trying to declare victory

before the results were in, stating that official results were not yet in and directing people to the official results.

- We attached informational labels to content that sought to delegitimize the outcome of the election or discuss the legitimacy of voting methods, for example, by claiming that lawful methods of voting lead to fraud. This label provided basic reliable information about the integrity of the election and voting methods.
- We enforced our violence and harm policies more broadly by expanding our definition of high-risk targets to include election officials in order to help prevent any attempts to pressure or harm them, especially while they were fulfilling their critical obligations to oversee the vote counting.
- We strengthened our enforcement against militias, conspiracy networks, and other groups that could have been used to organize violence or civil unrest in the period after the election. We removed thousands of these groups from our platform.

Since 2016, we've built an advanced system combining people and technology to review the billions of pieces of content that are posted to our platform every day. State-of-the-art AI systems flag content that may violate our policies, users report content to us they believe is questionable, and our own teams review content. We've also been building a parallel viral content review system to flag posts that may be going viral—no matter what type of content it is—as an additional safety net. This helps us catch content that our traditional systems may not pick up. We used this tool throughout this election, and in countries around the world, to detect and review Facebook and Instagram posts that were likely to go viral and take action if that content violated our policies.

While the COVID–19 pandemic continues to disrupt our content review workforce, we are seeing some enforcement metrics return to pre-pandemic levels. Our proactive detection rates for violating content are up from Q2 across most policies, due to improvements in AI and expanding our detection technologies to more languages. Even with a reduced review capacity, we still prioritize the most sensitive content for people to review. We recently published our Community Standards Enforcement Report for the third quarter of 2020, available at *https://transparency.facebook.com/community-standards-enforcement.*

We also coordinated with state attorneys general and other federal, state, and local law enforcement officials responsible for election protection. If they identified potential voter interference, we investigated and took action where warranted. These efforts were part of our ongoing coordination with law enforcement and election authorities at all levels to protect the integrity of elections.

Question 22. In your written testimony, you detail a number of steps Facebook is taking to ensure the 2020 General Election was not marred by misinformation or false claims. Additionally, according to a report in the Wall Street Journal, Facebook planned for possible unrest in the United States during and after the election. Part of this planning, according to the report, includes readying "tools designed for. . .'at-risk' countries."

Please further detail what "tools" Facebook prepared to deploy during the election, including those described in your written testimony.

Did Facebook deploy any of the above tools during the 2020 General Election? If so, please describe the tools used and the circumstances in which they were used.

Please describe circumstances in the election that would have required the use of the tools you have detailed.

Answer. Please see the response to your Question 21. We also developed temporary measures to address the challenge of uncertainty in the period after Election Day.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. MIKE LEE TO MARK ZUCKERBERG

Question 1. Mr. Zuckerberg, during the hearing I asked each of the witnesses to provide me with one example of a high-profile person or entity from a liberal ideology that your company has censored and what particular action you took. In response, you told me that, "I can get you a list." Could you please provide the list?

Answer. As a general matter, when we identify or learn of content that violates our policies, we remove that content regardless of who posted it. The political affiliation of the user generating the content has no bearing on that content assessment. Rather, decisions about whether to remove content are based on our Community Standards, which direct all reviewers when making decisions. We seek to write ac-

tionable policies that clearly distinguish between violating and non-violating content, and we seek to make the review process for reviewers as objective as possible.

In terms of moderation decisions, we have removed content posted by individuals and entities across the political spectrum. For example, we have taken down ads submitted on behalf of the Biden campaign and the Democratic National Committee, and organizations like the SEIU. We also have taken down ads submitted on behalf of the Trump campaign and the Republican National Committee, and organizations like the America First Action PAC.

Question 2. Mr. Zuckerberg, after Twitter made the decision to start fact-checking President Trump's tweets, you noted your disagreement with the policy stating on May 28, 2020: *"I don't think that Facebook or Internet platforms in general should be arbiters of truth. Political speech is one of the most sensitive parts in a democracy, and people should be able to see what politicians say . . . In terms of political speech, again, I think you want to give broad deference to the political process and political speech."*[1] I agree with this statement. But on September 3, 2020, you announced new policies fighting "misinformation." How is blocking what you deem "misinformation" consistent with your prior stance of wanting to give "broad deference" to the political process and political speech?

Answer. Freedom of expression is a founding principle for Facebook. Giving people a voice to express themselves has been at the heart of everything we do. We think people should be able to see for themselves what politicians are saying. We don't believe that it's an appropriate role for us to referee political debates and prevent a politician's speech from reaching its audience and being subject to public debate and scrutiny. That's why direct speech from politicians is generally not eligible for our third-party fact-checking program.

Our commitment to free speech, however, does not mean that politicians can say whatever they want on Facebook. They can't spread misinformation about where, when, or how to vote, for example, or incite violence. And when a politician shares previously debunked content, including links, videos, and photos, we demote that content, display related information from fact-checkers, and reject its inclusion in advertisements. When it comes to ads, while we won't remove politicians' ads based solely on the outcome of a fact-check, we still require them to follow our Advertising Policies.

On September 3, 2020, we announced additional steps to help secure the integrity of the U.S. elections by encouraging voting, connecting people to reliable information, and reducing the risks of post-election confusion. We did not change our policies regarding fact-checking direct speech from politicians as part of that announcement. For more information, please visit *https://about.fb.com/news/2020/09/additional-steps-to-protect-the-us-elections/*.

Question 3. Mr. Zuckerberg, last month Facebook tagged a Michigan ad that criticized Joe Biden and Senator Gary Peters as "missing context."[2] The next day the ad was shut down entirely.[3] Facebook relied on a supposed "fact-check" from PolitiFact, which said that the ad makes "predictions we can't fact-check."[4]

a. Don't political ads by their very nature lack context?

Answer. We do not allow advertisers to run ads that contain content that has been debunked by third-party fact-checkers, including content rated False, Partly False, Altered, or Missing Context. Third-party fact-checkers can review and rate public Facebook and Instagram posts, including ads, articles, photos, videos, and text-only posts. While Facebook is responsible for setting rating guidelines, the fact-checkers independently review and rate content. Missing Context is an available rating option for fact-checkers for content that may mislead without additional information or context. For example, this rating may be used for: clips from authentic video or audio or cropping of authentic photos that lack the full context from the original content, but that have not otherwise been edited or manipulated; media edited to omit or reorder the words someone said that changes, but that does not

---

[1] Zuckerberg, Mark. 2020, May 28. *Interview with Andrew Ross Sorkin, CNBC's "Squawk Box,"* comments at 0:37. *https://www.cnbc.com/2020/05/28/zuckerberg-facebook-twitter-should-not-fact-check-political-speech.html*

[2] American Principles Project (@approject). 2020, Sept. 15. *"Facebook just censored our PAC's $4 million ad campaign in Michigan . . ."* [Tweet]. *https://twitter.com/approject/status/1305901992108318721*

[3] Schilling, Terry (@schilling1776). 2020, Sept. 16. *"We received final word from @PolitiFact today that the @approject ads appeal has been rejected. Here's why . . ."* [Tweet]. *https://twitter.com/Schilling1776/status/1306302305508249603*

[4] PolitiFact. 2020, Sept. 15. *Ad watch: Conservative PAC claims Gary Peters would 'destroy girls' sports'. https://www.politifact.com/article/2020/sep/15/ad-watch-peters-supports-ending-discrimination-bas/*

reverse the meaning of the statement; hyperbole or exaggeration that is technically false but based on a real event or statement; content that presents a conclusion not supported by the underlying facts; claims stated as fact that are plausible but unproven; and more. For more information, please visit *https://www.facebook.com/business/help/341102040382165?id=673052479947730.*

b. Vice President Biden has run ads that say Donald Trump will defund Social Security,[5] a statement PolitiFact rated "Mostly False."[6] Joe Biden has also run ads saying President Trump is "attack[ing] democracy itself"[7] and that the GOP are doing "everything they can to invalidate the election."[8] It's my understanding that Facebook took no action against these ads. Can you explain this? Do you think these statements "lack context"?

Answer. As discussed in the response to your Question 2, direct speech from politicians, including advertisements, is generally not eligible for our third-party fact-checking program.

Question 4. Mr. Zuckerberg, on October 14, Facebook spokesperson Andy Stone announced on Twitter that Facebook would be "reducing its distribution"[9] of the NY Post article regarding Hunter Biden and Ukraine. He later noted that this was Facebook's "standard process."[10]

c. Does Facebook "reduce distribution" of every news article before your fact checker reviews it? If not, why in this case did you "reduce distribution" of the NY Post article?

d. What particular metrics do you use to "reduce distribution" of articles or publications prior to conducting a fact-check? And are these metrics publicly available?

Answer. In 2019, we announced that if we identify signals that a piece of content is false, we will temporarily reduce its distribution in order to allow sufficient time for our independent, third-party fact-checkers to review and determine whether to apply a rating. Quick action is critical in keeping a false claim from going viral, and so we take this step to provide an extra level of protection against potential misinformation. These temporary demotions expire after seven days if the content has not been rated by an independent fact-checker.

For the past several months, the U.S. intelligence community has urged voters, companies, and the Federal government to remain vigilant in the face of the threat of foreign influence operations seeking to undermine our democracy and the integrity of our electoral process. For example, the Director of National Intelligence, the Head of the FBI, and the bipartisan leaders of the Senate Select Committee on Intelligence reminded Americans about the threat posed by foreign influence operations emanating from Russia and Iran. Along with their public warnings, and as part of the ongoing cooperation that tech companies established with government partners following the 2016 election, the FBI also privately warned tech companies to be on high alert for the potential of hack-and-leak operations carried out by foreign actors in the weeks leading up to November 3rd. We took these risks seriously.

Regarding the October 14 *New York Post* story, given the concerns raised by the FBI and others, we took steps consistent with our policies to slow the spread of suspicious content and provide fact-checkers the opportunity to assess it. However, at no point did we take any action to block or remove the content from the platform. People could—and did—read and share the *Post*'s reporting while we had this temporary demotion in place. Consistent with our policy, after seven days, we lifted the temporary demotion on this content because it was not rated false by an independent fact-checker.

Question 5. Mr. Zuckerberg, you've noted that Facebook is a "platform for all ideas." In response to Twitter's decision in 2017 to block Sen. Blackburn's ad for the

---

[5] Joe Biden, Biden Victory Fund. 2020, Aug. 27–28. *Donald Trump said that if he's re-elected, he'll defund Social Security—we can't let that happen . . .* [Facebook Ad]. *https://www.facebook.com/ads/library/?id=309439267043399*

[6] PolitiFact. 2020, Aug. 12. *Did Trump say he will terminate Social Security if re-elected? https://www.politifact.com/factchecks/2020/aug/12/social-security-works/did-trump-say-he-will-terminate-social-security-if/*

[7] Joe Biden, Biden for President. 2020, Oct. 23–27. *Donald Trump has repeatedly attacked democracy and those who fight for it . . .* [Facebook Ad]. See in text and in video at 0:37. *https://www.facebook.com/ads/library/?id=767615284086821*

[8] Joe Biden, Biden Victory Fund. 2020, Oct. 26–27. *In the first debate, Donald Trump cast doubt on the validity . . .* [Facebook Ad]. *https://www.facebook.com/ads/library/?id=1110668466017273*

[9] Stone, Andy (@andymstone). 2020, Oct. 14. *While I will intentionally not link to the New York Post, I want be clear . . .* [Tweet]. *https://twitter.com/andymstone/status/1316395902479872000*

[10] Stone, Andy (@andymstone). 2020 Oct. 14. *This is part of our standard process as we laid out here . . .* [Tweet] *https://twitter.com/andymstone/status/1316425399384059904*

ad's pro-life language, your COO, Sheryl Sandberg, argued that Facebook would have let the ad run. She noted: "When you cut off speech for one person, you cut off speech for all people."[11] Does Facebook's censorship of the NY Post story equate to "cutting off speech for all people"? And how is this consistent with your position that Facebook is a "platform for all ideas"?

Answer. As an initial matter, and as explained above, with respect to the *New York Post* story, we took steps consistent with our policies to slow the spread of suspicious content and provide fact-checkers the opportunity to assess it. At no point did we take any action to block or remove the content from the platform. People could—and did—read and share the *Post*'s reporting while we had this temporary demotion in place. Consistent with our policy, after seven days, we lifted the temporary demotion on this content because it was not rated false by an independent fact-checker.

Facebook is a platform for ideas across the political and ideological spectrum, but people also tell us they don't want to see misinformation. That's why we work to reduce the spread of viral misinformation on our platform by working with independent, third-party fact-checkers.

Question 6. In 2017, the FCC passed the Restoring Internet Freedom Order, which reversed the FCC's net neutrality order. Facebook, Google, and Twitter each opposed that decision calling for a "free and open internet." Let me be clear. I still support the FCC's decision to not regulate the Internet as a public utility under Title II. But Mr. Zuckerberg, I found your comments particularly interesting. In 2017, you noted: "*Net neutrality is the idea that the Internet should be free and open for everyone. If a service provider can block you from seeing certain content or can make you pay extra for it, that hurts all of us and we should have rules against it . . . If we want everyone in the world to have access to all the opportunities that come with the internet, we need to keep the Internet free and open.*"[12] Again, I find the idea that we would regulate the Internet as a public utility to be bad policy, but you indicated that service providers that "can block you from seeing certain content" denigrate a free and open internet. By that logic, would you say that Facebook is now not contributing to a free and open Internet due to the blocking of certain viewpoints?

Answer. Freedom of expression is one of our core values, and we believe that the Facebook community is richer and stronger when a broad range of viewpoints is represented. We are committed to encouraging dialogue and the free flow of ideas by designing our products to give people a voice. We also know that people will not come to Facebook to share and connect with one another if they do not feel that the platform is a safe and respectful environment. In that vein, we have Community Standards that are public and that outline what is and is not allowed on Facebook. Suppressing content on the basis of political viewpoint or preventing people from seeing what matters most to them is directly contrary to Facebook's mission and our business objectives.

We base our policies on principles of voice, safety, dignity, authenticity, and privacy. Our policy development is informed by input from our community and from experts and organizations outside Facebook so we can better understand different perspectives on safety and expression, as well as the impact of our policies on different communities globally. Based on this feedback, as well as changes in social norms and language, our standards evolve over time.

Decisions about whether to remove content are based on whether the content violates our Community Standards. Discussing controversial topics or espousing a debated point of view is not at odds with our Community Standards. In fact, we believe that such discussion is important in helping bridge division and promote greater understanding.

Question 7. Congress is in the midst of a debate over future reforms to Section 230. This is an important discussion that Congress should have.

a. In making decisions to moderate third-party content on your platform, do you rely solely on Section 230? In other words, could you still moderate third-party content without the protections of Section 230?

b. If the provisions of Section 230 were repealed or severely limited, how would your content moderation practices shift?

Answer. Broadly speaking, Section 230 does two things. First, it encourages free expression. Without Section 230, platforms could potentially be held liable for every-

---

[11] Sandberg, Sheryl. 2017, Oct. 17. *Axios Exclusive Interview with Sheryl Sandberg,* comments found at 7:31. *https://www.axios.com/exclusive-interview-with-facebooks-sheryl-sandberg-1513306121-64e900b7-55da-4087-afee-92713cbbfa81.html*

[12] Zuckerberg, Mark. 2017, July 12. *Today people across the U.S. are rallying together to save net neutrality* . . . [Facebook Post]. *https://www.facebook.com/zuck/posts/10103878724831141*

thing people say. Platforms would likely moderate more content to avoid legal risk and would be less likely to invest in technologies that enable people to express themselves in new ways. Second, it allows platforms to moderate content. Without Section 230, platforms could face liability for doing even basic moderation, such as removing hate speech and harassment that impact the safety and security of their communities. Repealing Section 230 entirely would likely substantially increase many companies' costs associated with legal challenges and content moderation.

Question 8. How many content posts or videos are generated by third-party users per day on Facebook, Twitter, and YouTube?

c. How many decisions on average per day does your platform take to moderate content? Are you able to provide data on your takedown numbers over the last year?

d. Do you ever make mistakes in a moderation decision? If so, how do you become aware of these mistakes and what actions do you take to correct them?

e. What remedies or appeal process do you provide to your users to appeal an action taken against them? On average, how long does the adjudication take until a final action is taken? How quickly do you provide a response to moderation decision appeals from your customers?

f. Can you provide approximate numbers, by month or week, for the times you tookdown, blocked, or tagged material from November 2019 to November 2020?

Answer. Billions of pieces of content are posted to our platform every day. Content reviewers take action on content that is flagged after it is assessed against our Community Standards. Our Community Standards are global, and all reviewers use the same guidelines when assessing content. We seek to write actionable policies that clearly distinguish between violating and non-violating content, and we seek to make the assessment process for reviewers as objective as possible.

We recognize that our policies are only as good as the strength and accuracy of our enforcement—and our enforcement is not perfect. We make mistakes because our processes involve both people and machines, and neither are infallible. We are always working to improve. One way in particular that we become aware of mistakes is through user feedback, such as when users appeal our content moderation decisions.

Every week, we audit a sample of reviewer decisions for accuracy and consistency. We also audit our auditors. When a reviewer makes mistakes or misapplies our policies, we follow up with appropriate training and review the mistakes with our Community Operations team to prevent similar mistakes in the future.

With respect to our appeals process, we generally provide our users with the option to disagree with our decision when we have removed their content for violating our policies or when they have reported content and we have decided it does not violate our policies. In some cases, we then re-review our decisions on those individual pieces of content.

In order to request re-review of a content decision we made, users are often given the option to "Request Review" or to provide feedback by stating they "Disagree with Decision." We try to make the opportunity to request this review or give this feedback clear, either via a notification or interstitial, but we are always working to improve.

Transparency in our appeals process is important, so we now include in our Community Standards Enforcement Report how much content people appealed and how much content was restored upon appeal. Gathering and publishing those statistics keeps us accountable to the broader community and enables us to continue improving our content moderation. For more information, see *https://transparency.facebook.com/community-standards-enforcement.*

Our Community Standards Enforcement Report also shares metrics on how Facebook is performing in preventing and removing content that violates our Community Standards. The report specifies how much content we took action on during the specified period, as well as how much of it we found before users reported it to us.

Question 9. The first major case to decide the application of Section 230 was *Zeran v. AOL.*[13] In *Zeran,* Judge Wilkinson recognized the challenges of conferring "distributor liability" to a website because of the sheer number of postings. That was 1997. If we imposed a form of "distributor liability" on your platforms that would likely mean that your platform would be liable for content if you acquired knowledge of the content. I think there is an argument to be made that you "acquire knowledge" when a user "flags" a post, video, or other form of content.

g. How many "user-generated" flags do your companies receive daily?

h. Do users ever flag posts solely because they disagree with the content?

---

[13] Kenneth M. Zeran v. America Online, Inc. 129 F. 3d 327 (4th Cir. 1997)

i. If you were liable for content that was "flagged" by a user, how would that affect content moderation on your platform?

Answer. Facebook encourages users to report content to us that violates our Community Standards, including if it contains or relates to nudity, violence, harassment, terrorism, or suicide or self-injury. Facebook's Community Operations team receives millions of reports each week, and they work hard to review those reports and take action when content violates our policies. User reports are an important signal, and we rely on our community to help identify content that violates our policies. However, not every piece of reported content is determined to violate our policies upon review. In some cases, users may report posts because they disagree with the content or based on other objections that do not constitute violations of our Community Standards. In such cases, we give users control over what they see and who they interact with by enabling them to block, unfriend, or unfollow the other user.

Facebook publishes a quarterly Community Standards Enforcement Report to track our progress; for more information regarding Facebook's content moderation efforts, please visit *https://transparency.facebook.com/community-standards-enforcement.*

Broadly speaking, Section 230 does two things. First, it encourages free expression. Without Section 230, platforms could potentially be held liable for everything people say. Platforms would likely moderate more content to avoid legal risk and would be less likely to invest in technologies that enable people to express themselves in new ways. Second, it allows platforms to moderate content. Without Section 230, platforms could face liability for doing even basic moderation, such as removing hate speech and harassment that impact the safety and security of their communities.

Question 10. Section 230 is often used as a legal tool to have lawsuits dismissed in a pre-trial motion.

j. How often is your company sued under a theory that you should be responsible for the content posted by a user of your platform? How often do you use Section 230 as a defense in these lawsuits? And roughly how often are those lawsuits thrown out?

Answer. We do not have a precise number reflecting how often we're sued under different legal theories, but defending lawsuits related to users' content on our platform requires a substantial amount of resources, including litigation costs and employee time, both in the U.S. and elsewhere.

We may invoke Section 230 in our defense against such suits when a claim seeks to treat Facebook as the publisher or speaker of information provided by a user or other entity.

k. If Section 230 was eliminated and a case seeking to make your platform liable for content posted by a third party went to the discovery phase, roughly how much more expensive would that case be as opposed to its dismissal pre-discovery?

Answer. Broadly speaking, Section 230 does two things. First, it encourages free expression. Without Section 230, platforms could potentially be held liable for everything people say. Platforms would likely moderate more content to avoid legal risk and would be less likely to invest in technologies that enable people to express themselves in new ways. Second, it allows platforms to moderate content. Without Section 230, platforms could face liability for doing even basic moderation, such as removing hate speech and harassment that impact the safety and security of their communities. Repealing Section 230 entirely would likely substantially increase many companies' costs associated with legal challenges and content moderation.

Question 11. Section 230s Good Samaritan provision contains the term "otherwise objectionable."

l. How do you define "otherwise objectionable"?

m. Is "otherwise objectionable" defined in your terms of service? If so, has its definition ever changed? And if so, can you provide the dates of such changes and the text of each definition?

n. In most litigation, a defendant relies on Section 230(c)(1) for editorial decisions. If a company could only rely on 230(c)(2) for a moderation decision (as has been discussed in Congress), how would that affect your moderation practices? And how would striking "otherwise objectionable" from 230(c)(2) further affect your moderation practices?

Answer. As we understand it, "otherwise objectionable," as the term is used in Section 230, is a standard that courts have interpreted for many years. At Facebook, our Community Standards—which are public—include restrictions around content that is harmful to members of our community, including bullying, harassment, hate speech, and incitement to violence.

At Facebook, we are a platform for ideas across the political and ideological spectrum, and we moderate content according to our published Community Standards in order to keep users on the platform safe, reduce objectionable content, and ensure users participate on the 1platform responsibly. We are clear and transparent about what our standards are, and we seek to apply them to all of our users consistently. The political affiliation of the user generating the content has no bearing on content removal assessments.

Facebook's Community Standards prohibit coordinating harm and criminal activity, including posting content that sexually exploits or endangers children. When we become aware of apparent child exploitation, we report it to the National Center for Missing and Exploited Children (NCMEC), in compliance with applicable law. We work hard to identify and remove such content; over the past three years, we've found over 99 percent of the violating content we actioned before users reported it to us. And we certainly think it is important to make sure that platforms are serious about the illegal activity on their platforms.

For example, Facebook supported SESTA/FOSTA, and we were very pleased to be able to work successfully with a bipartisan group of Senators on a bill that protects women and children from the harms of sex trafficking. We would welcome the opportunity to work with the Committee on proposals to modify Section 230 in ways that focus on bad actors who intentionally facilitate wrongdoing, while being mindful not to disincentivize platforms from trying to find the illegal activity in the first place.

Question 12. Are your terms of service a legally binding contract with your users? How many times have you changed your terms of service in the past five years? What recourse do users of your platform have when you allege that they have violated your terms of service?

Answer. We believe that people should have clear, simple explanations of how online services work and use personal information. In June 2019, we updated our Terms of Service to clarify how Facebook makes money and better explain the rights people have when using our services. The updates did not change any of our commitments or policies—they solely explained things more clearly. These updates are part of our ongoing commitment to give people more transparency and control over their information. June 2019 is the last time we updated our Terms.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. RON JOHNSON TO MARK ZUCKERBERG

Question 1. During the hearing, in response to both Senator Cruz's line of questioning and mine, Mr. Dorsey claimed that Twitter does not have the ability to influence nor interfere in the election.

a. Do you believe Facebook has the ability to influence and/or interfere in the election? To reiterate, I am not asking if you have the intent or have actively taken steps to influence/interfere, but rather if Facebook has the *ability?*

Answer. It is the more than 160 million Americans who voted in this election that decided the election's outcome. We are proud that Facebook is one of the places that individuals could go to learn about candidates and issues, and about the electoral process more generally, and we take extremely seriously the responsibility that comes with these uses of our product.

We took a number of steps to help protect the integrity of the democratic process, including combating foreign interference, bringing transparency to political ads, limiting the spread of misinformation, and—against the backdrop of a global pandemic—providing citizens access to reliable information about voting.

We are confident in the actions that we took to protect the safety and security of this election. But it's also important for independent voices to weigh in. That's why we launched a new independent research initiative with more than a dozen academics to look specifically at the role Facebook and Instagram played in the election. The results—whatever they may be—will be published next year, unrestricted by Facebook and broadly available. This research won't settle every debate about social media and democracy, but we hope that it will shed more light on the relationship between technology and our elections.

b. If you claim that you do not have the ability to influence or interfere in the election, can you explain Facebook's rational for suppressing content that Facebook deems to be Russian misinformation on the basis that it influences the election?

Answer. Inauthentic behavior, including foreign influence campaigns, has no place on Facebook. If we find instances of coordinated inauthentic behavior conducted on behalf of a foreign actor, regardless of whether or not such behavior targets a candidate or political party, we apply the broadest enforcement measures, including the

removal of every on-platform property connected to the operation itself and the people and organizations behind it. We also report publicly about such takedowns in a monthly report, available at *https://about.fb.com/news/tag/coordinated-inauthentic-behavior/*.

Question 2. In Senator Rosen's testimony, you stated that Congress could hold Facebook accountable by monitoring the percentage of users that see harmful content before Facebook acts to take it down. While this is important, it does not address the problem of Facebook biasedly and inconsistently enforcing content moderation policies of political speech.

c. In regards to this issue, what role do you think Congress should have in holding Facebook accountable?

d. Do you have an example of a mechanism by which Congress can currently hold Facebook accountable on this issue? If there are none, can you please at a minimum acknowledge that there are none?

Answer. Facebook is a platform for ideas across the political and ideological spectrum. Suppressing content on the basis of political viewpoint directly contradicts Facebook's mission and our business objectives.

We are committed to free expression and err on the side of allowing content. When we make a mistake, we work to make it right. And we are committed to constantly improving our efforts so we make as few mistakes as possible. Decisions about whether to remove content are based on whether the content violates our Community Standards, not political affiliation or viewpoint. Discussing controversial topics or espousing a debated point of view is not at odds with our Community Standards. We believe that such discussion is important in helping bridge division and promote greater understanding.

We don't always get it right, but we try to be consistent. The reality is that people have very different ideas and views about where the line should be. Democrats often say that we don't remove enough content, and Republicans often say we remove too much. Indeed, people can reasonably disagree about where to draw the lines. We need a more accountable process that people feel is legitimate and that gives platforms certainty.

Question 3. During a Senate Commerce Committee hearing in the Summer of 2019, I asked your company's representative about how the "suggestions for you" feature decides which accounts should be prompted after a user follows a new account. My staff found at the time that, no matter which user was requesting to follow the Politico account, they were all prompted the same liberal suggestions such as Senator Sanders, Senator Warren, and MSNBC. The user had to scroll through dozens of suggested accounts before finding any semblance of a non-liberal publication, the *Wall Street Journal.*

Following the hearing, I sent your company a letter asking why these suggestions were made based on your stated data policy which says the suggestions are based off of accounts that the user follows and likes. In your company's response it said, "*These suggestions are generated by Instagram automatically, using machine learning systems that consider a variety of signals, such as the accounts you follow and your likes. Our employees don't determine the ranking of any specific piece of content. . .*"

We later met in person to discuss this matter, and you changed your tune by saying that these suggestions are based on who follows and likes Politico's account, and not so much what the user likes and follows. You also made clear that POLITICO itself has nothing to do with which accounts are suggested when a user clicks to follow their account. It is Instagram that has control over that.

e. Mr. Zuckerberg, you have claimed on multiple occasions that Facebook and Instagram are neutral platforms. However, more than a year later since the hearing in 2019, the suggested results for my staff continue to have the same liberal bias as they did then. Do you still contend that Facebook and Instagram are neutral platforms?

Answer. Facebook is first and foremost a technology company. We do not create or edit the content that our users post on our platform. While we are a platform for ideas across the political and ideological spectrum, we do moderate content in good faith according to our published Community Standards in order to keep users on the platform safe, reduce objectionable content, and ensure users participate on the platform responsibly.

Freedom of expression is one of our core values, and we believe that the Facebook community is richer and stronger when a broad range of viewpoints is represented. We are committed to encouraging dialogue and the free flow of ideas by designing our products to give people a voice. We also know that people will not come to Facebook to share and connect with one another if they do not feel that the platform

is a safe and respectful environment. In that vein, we have Community Standards that outline what is and is not allowed on Facebook.

Recommendations for accounts that a user may want to follow or like are based on a variety of signals, including the accounts that the user already follows and likes, and the other users that follow and like those accounts. These suggestions are generated automatically using machine learning systems. Facebook employees do not determine the rankings or recommendations for any specific piece of content.

f. Does a publication, like Politico for example, have the ability to pay Facebook to be included more frequently in the "suggestions for you" feature?

Answer. Publishers cannot pay to appear in the "Suggested for You" feature.

As discussed in the letter you reference, the purpose of the "Suggested for You" feature is to help people find accounts that may interest them. These suggestions are generated by Facebook and Instagram automatically, using machine learning systems that consider a variety of signals, such as the accounts people follow and their likes. Our employees don't determine the ranking of any specific account in Suggested for You.

Question 4. During Mr. Dorsey's testimony he said that Twitter should "enable people to choose algorithms created by 3rd parties to rank and filter their own content," in reference to Dr. Stephen Wolfram's research.

g. Which of the methods described in his research and testimony have you deployed on your platform?

h. What other methods would you like to see put in place?

i. What is preventing you from implementing more methods such as these?

Answer. An algorithm is a formula or set of steps for solving a particular problem. At Facebook, we use algorithms to offer customized user experiences and to help us achieve our mission of building a global and informed community. For example, we use algorithms to help generate and display search results (see *https://about.fb.com/news/2018/11/inside-feed-how-search-works/*), to determine the order of posts that are displayed in each person's personalized News Feed (see *https://about.fb.com/news/2018/05/inside-feed-news-feed-ranking/*), and to serve ads that may be relevant to them.

On Facebook, people see posts from their friends, Pages they've chosen to follow, and Groups they've joined, among others, in their News Feed. On a given day, the number of eligible posts in a user's Feed inventory can number in the thousands, so we use an algorithm to personalize how this content is organized. The goal of the News Feed algorithm is to predict what pieces of content are most relevant to the individual user, and rank (*i.e.,* order) those pieces of content accordingly every time a user opens Facebook, to try and bring those posts that are the most relevant to a person closer to the top of their News Feed. This ranking process has four main elements: the available inventory (all of the available content from the people, Pages, and Groups a person has chosen to connect with); the signals, or data points, that can inform ranking decisions (*e.g.,* who posted a particular piece of content); the predictions we make, including how likely we think a person is to comment on a story, share with a friend, etc.; and a relevancy score for each story, which informs its position in News Feed.

We frequently make changes to the algorithms that drive News Feed ranking in an effort to improve people's experience on Facebook. For example, in 2018, we responded to feedback from our community that public content—posts from businesses, brands, and media—was crowding out the personal moments that lead us to connect more with each other. As a result, we moved from focusing only on helping users find relevant content to helping them have more meaningful social interactions. This meant that users began seeing more content from their friends, family, and Groups. We also reduce the distribution of some problematic types of content, including content that users may find spammy or low-quality, such as clickbait headlines, misinformation as confirmed by third-party fact-checkers, and links to low-quality webpages like ad farms.

To help people on Facebook better understand what they see from friends, Pages, and Groups in News Feed, including how and why that content is ranked in particular ways, we publish a series of blog posts called News Feed FYI (see *https://about.fb.com/news/category/news-feed-fyi/*), which highlight major updates to News Feed and explain the thinking behind them. Also, in 2019, we launched a feature called "Why am I seeing this post?" (see *https://about.fb.com/news/2019/03/why-am-i-seeing-this/*). This feature directly responded to user feedback asking for more transparency around why certain content appears in News Feed and easier access to News Feed controls. Through their News Feed Preferences, users can choose to see posts from certain friends and Pages higher up in their News Feed. Controls

also include Snooze, which keeps the content from a selected person, Page, or Group out of a user's News Feed for a limited time.

Users who do not wish to consume ranked News Feed also have access to a control to view content purely chronologically from those they follow in the 'Most Recent' Feed view (see *https://www.facebook.com/help/218728138156311*). Additionally, we promoted a series of educational initiatives and campaigns to help people learn about the technology that underlies our various products and features, which includes AI and machine learning, through our series called "Inside Feed" (see *https://about.fb.com/news/category/inside-feed/*).

Question 5. Do you agree that Facebook competes with local newspapers and broadcasters for local advertising dollars?

Answer. The advertising sector is incredibly dynamic, and competition for advertising spend is increasingly fierce. Companies have more options than ever when deciding where to advertise. Unlike a few decades ago, when companies had more limited options, today there are more choices, different channels and platforms, and hundreds of companies offering them.

Facebook competes for advertisers' budgets with online and offline advertisers and with a broad variety of advertising players. This includes the intense competitive pressure that Facebook faces for advertising budgets from offline channels (such as print, radio, and broadcast), established digital platforms (such as Google, Amazon, Twitter, and Pinterest), and newer entrants that have attracted a large user base from scratch (such as Snap and TikTok). The landscape is also highly dynamic, with offline advertising channels (such as television and radio) benefiting from industry-wide digitalization and new technologies to offer their own ad targeting and measurement products.

Advertisers can and do shift spend in real time across ad platforms to maximize their return on investment. As a result of this competition and choice, advertisers spread their budgets across multiple outlets and channels, including Facebook.

Facebook is able to provide nearly all of its consumer services free of charge because it is funded by advertising that is relevant and useful. Millions of Americans use Facebook to connect with the people, organizations, and businesses they care about. Research has shown that though Facebook offers these services at no cost, they offer significant value—a huge consumer surplus.

j. Should Congress allow local news affiliates, such as local newspapers and local broadcast stations, to jointly negotiate with Facebook for fair market compensation for the content they create when it is distributed over your platform?

Answer. The antitrust laws promote competition and innovation, and they have stood the test of time. The laws are flexible, and they can meet the challenges of today.

Mobile technology has fundamentally changed the way people discover and consume news, and this has resulted in real challenges for publishers. We understand these challenges and have worked with publishers to adapt to digital transformation. But *how* news is distributed on Facebook warrants further discussion. News organizations voluntarily post their content on Facebook because it helps them reach new and larger audiences, and ultimately those audiences drive additional revenue for them.

To date, we have built tools to help publishers increase their subscribers by driving people from Facebook links to publisher websites. Among other benefits, Facebook provides publishers with free, organic distribution of news (and other content), which grows the audience and revenue for news publishers; customized tools and products to help news publishers monetize their content; and initiatives to assist publishers to innovate with online news content, along with bringing indirect value to publishers such as brand awareness and community-building.

Publishers are also able to decide when a reader sees a paywall on content they've found via Facebook. Publishers control the relationship with their readers with subscription payments taking place directly on their owned and operated websites. We do not take any cut of the subscription revenue because we want that money to go toward funding quality journalism. Helping publishers reach new audiences has been one of our most important goals.

Beyond distribution and revenue tools already mentioned, we've focused on meaningful collaboration with publishers. In 2017, we launched the Facebook Journalism Project (see *https://www.facebook.com/journalismproject*), an initiative focused on building innovative and sustainable solutions to support journalism. In 2019, we announced a $300 million commitment (see *https://www.facebook.com/journalismproject/facebook-supports-local-news*) to news programs, partnerships, and content—with a specific focus on local news. And later that year we launched Facebook News

(see *https://www.facebook.com/news*), a section of Facebook dedicated solely to reliable and informative news content.

During the COVID–19 pandemic, we announced a $100 million investment (see *https://www.facebook.com/journalismproject/coronavirus-update-news-industry-support*) to support the news industry—$25 million in emergency grant funding for local news through the Facebook Journalism Project, and $75 million in additional marketing spend to move money over to news organizations around the world.

We've also focused on supporting the global fact-checking community's work—we partnered with the Independent Fact-Checking Network to launch a $1 million grant program (see *https://www.facebook.com/journalismproject/coronavirus-grants-fact-checking*) to increase their capacity during this time.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. MARIA CANTWELL TO MARK ZUCKERBERG

Foreign Disinformation. Facebook/Instagram, Twitter, and Google/YouTube have each taken concrete steps to improve defensive measures through automated detection and removal of fake accounts at creation; increased internal auditing and detection efforts; and established or enhanced security and integrity teams who can identify leads and analyze potential networks engaging in coordinated inauthentic behavior.

Social media companies have hired a lot of staff and assembled large teams to do this important work and coordinate with the FBI led Foreign Influence Task Force (FITF).

Small companies in the tech sector do not have the same level or expertise or resources, but they face some of the same and growing threats.

Likewise, public awareness and understanding of the threats foreign actors like Russia pose is key to helping fight back against them.

Question 1. What specific steps are you taking to share threat information with smaller social media companies that do not have the same level of resources to detect and stop those threats?

Answer. We work with others in the industry to limit the spread of violent extremist content on the Internet. For example, in 2017, we established the Global Internet Forum to Counter Terrorism (GIFCT) with others in the industry with the objective of disrupting terrorist abuse on digital platforms. Since then, the consortium has grown and collaborates closely on critical initiatives focused on tech innovation, knowledge-sharing, and research.

Question 2. Intel Chairman Schiff has highlighted the need for social media companies to increase transparency about how social media companies have stopped foreign actors disinformation and influence operations. Where are the gaps in public disclosures of this information and what specific actions are you taking to increase transparency about malign foreign threats you have throttled?

Answer. When we find instances of coordinated inauthentic behavior conducted on behalf of a government entity or by a foreign actor, in which the use of fake accounts is central to the operation, we apply the broadest enforcement measures, including the removal of every on platform property connected to the operation itself and the people and organizations behind it. We regularly share our findings about the networks we find and remove for coordinated inauthentic behavior.

Our teams continue to focus on finding and removing deceptive campaigns around the world—whether they are foreign or domestic. We have shared information about our findings with law enforcement, policymakers, and industry partners. And we publish regular reports on the coordinated inauthentic behavior we detect and remove from our platforms. Our October 2020 report can be found at *https://about.fb.com/news/2020/11/october-2020-cib-report/*.

Addressing Stop Hate for Profit Recommendations. The Stop Hate for Profit, Change the Terms, and Free Press coalition—all committed to combating racism, violence, and hate online—have called on social media platforms to adopt policies and take decisive actions against toxic and hateful activities.

This includes finding and removing public and private groups focused on white supremacy, promoting violent conspiracies, or other hateful content; submitting to regular, third party, independent audits to share information about misinformation; changing corporate policies and elevating a civil rights to an executive level position.

Question 3. Mr. Zuckerberg, you have taken some steps to address these recommendations from the organizations that have made it their mission to get Facebook to take a stronger role in stopping hate on your platform, which I appre-

ciate. Will you commit to continuing to meet with the experts at these anti-hate organizations to learn how to more quickly detect, remove, and stop hateful speech?

Answer. In developing and iterating on our policies, including our policy specific to hate speech, we consult with outside academics and experts from across the political spectrum and around the world and we look forward to doing so in the future.

Kenosha Wisconsin Violence. On August 25th, a man from Illinois traveled to Kenosha, Wisconsin armed with an assault rifle and fatally shot Joseph Rosenbaum and Anthony Huber, and injured another person, who were protesting the shooting of Jacob Blake, a Black resident, which left him paralyzed.

In the wake of these tragic shootings, we learned that a para-military group called the Kenosha Guard Militia, a group that organized on Facebook, called on followers to "take up arms" and "defend" the city against "evil thugs". This event post had been flagged 455 times by Facebook users, yet Facebook did not take down the group's page until after these lives were already lost.

While the Illinois shooter may not have been a member of the Kenosha Guard Militia, this brings up a very important point—that hate spread on social media platforms can lead to real life violence.

In May of this year, the *Wall Street Journal* reported that Facebook had completed internal research that said its internal algorithms "exploit the human brain's attraction to divisiveness", which could allow Facebook to feed more divisive content to gain user attention and more time on the platform. In response, the *Journal* reported that you buried the research and did little to address it because it ran counter to other Facebook initiatives.

Sowing divisions in this country and further polarizing public discourse is dangerous, and can have deadly consequences.

Question 4. Mr. Zuckerberg, you admitted it was a mistake to not to remove the Kenosha Guard page and event that encouraged violence. But you knew at the time that your algorithms help fuel the flames of these para-military organizations by amplifying divisiveness. It should have been an easy decision to remove the content. What do you believe is Facebook's responsibility to stop amplification of divisive content? Do you have concerns that Facebook is helping to divide our country?

Answer. Under our Violence and Incitement policy, we remove content, disable accounts, and work with law enforcement when we believe there is a genuine risk of physical harm or direct threats to public safety. We also try to consider the language and context in order to distinguish casual statements from content that constitutes a credible threat to public or personal safety.

Over the last several years we've continued to update and refine our Violence and Incitement policy. The most recent update to the policy was made after the events in Kenosha in August. In this update, we developed a framework whereby we can identify certain locations that are more at risk for violence or intimidation by the threat of violence, the same as we identify schools, polling places, and houses of worship, and remove more implicit calls and statements to bring weapons to that location. This policy could have meant the Kenosha Guard Event Page violated our Violence and Incitement policy had it been live at the time, but either way, the Kenosha Guard Page and the Event Page it hosted violated our policy addressing Militarized Social Movements, and the militia group's main Page was removed on that basis.

Indeed, earlier in August, we updated our policies to address militia organizations a week before the horrible events in Kenosha, and since then, we have identified over 600 groups that we consider militarized social movements and banned them from operating Pages, Groups, and Instagram accounts for their organizations. Following the violence that took place in Kenosha, we removed the shooter's Facebook and Instagram account and took action against organizations and content related to Kenosha. We have found no evidence that suggests the shooter followed the Kenosha Guard Page or that he was invited to the Event Page they organized.

Russian Election Interference. The U.S. Intelligence community found that foreign actors including Russia tried to interfere in the 2016 election and used social media platforms among other influence operations.

In 2017, the FBI established the Foreign Influence Task Force (FITF), which works closely with state and local partners to share information on threats and actionable leads.

The FBI has also established relationships with social media companies to enable rapid sharing of threat information. Social media companies independently make decisions regarding the content of their platforms.

The U.S. Intelligence Community warned that Russia was using a range of active measures to denigrate former Vice President Joe Biden in the 2020 election. They also warned about Iran and China.

Social media companies remain on the front lines of these threats to our democracy.

Question 5. What steps are you taking to prevent amplification of false voter fraud claims after the 2020 presidential election and for future elections? What challenges do you face trying to prevent foreign actors who seek to influence our elections?

Answer. We're gratified that, thanks to the hard work of election administrators across the country, the voting process went relatively smoothly. Facebook worked hard to do our part in protecting the integrity of the 2020 election, and we're proud of the work we've done to support our democracy. For example, we ran the largest voting information campaign in American history. Based on conversion rates we calculated from a few states we partnered with, we estimate that we helped 4.5 million people register to vote across Facebook, Instagram, and Messenger—and helped about 100,000 people sign up to be poll workers. We launched a Voting Information Center to connect people with reliable information on deadlines for registering and voting and details about how to vote by mail or vote early in person, and we displayed links to the Voting Information Center when people posted about voting on Facebook. More than 140 million people have visited the Voting Information Center on Facebook and Instagram since it launched. We are encouraged that more Americans voted in 2020 than ever before, and that our platform helped people take part in the democratic process.

We also worked to tackle misinformation and voter suppression. We displayed warnings on more than 150 million pieces of content that our third-party fact-checkers debunked. We partnered with election officials to remove false claims about polling conditions, and we put in place strong voter suppression policies that prohibit explicit or implicit misrepresentations about how or when to vote, as well as attempts to use threats related to COVID–19 to scare people into not voting. We removed calls for people to engage in voter intimidation that used militarized language or suggested that the goal was to intimidate, exert control, or display power over election officials or voters. In addition, we blocked new political and issue ads during the final week of the campaign, as well as all political and issue ads after the polls closed on election night.

We also instituted a variety of measures to help in the days and weeks after voting ended:

- We used the Voting Information Center to prepare people for the possibility that it could take a while to get official results. This information helped people understand that there was nothing illegitimate about not having a result on election night.
- We partnered with Reuters and the National Election Pool to provide reliable information about election results. We displayed this in the Voting Information Center, and we notified people proactively as results became available. We added labels to any post by a candidate or campaign trying to declare victory before the results were in, stating that official results were not yet in and directing people to the official results.
- We attached informational labels to content that sought to delegitimize the outcome of the election or discuss the legitimacy of voting methods, for example, by claiming that lawful methods of voting lead to fraud. This label provided basic reliable information about the integrity of the election and voting methods.
- We enforced our violence and harm policies more broadly by expanding our definition of high-risk targets to include election officials, in order to help prevent any attempts to pressure or harm them, especially while they were fulfilling their critical obligations to oversee the vote counting.
- We strengthened our enforcement against militias, conspiracy networks, and other groups that could have been used to organize violence or civil unrest in the period after the election. We removed thousands of these groups from our platform.

Since 2016, we've built an advanced system combining people and technology to review the billions of pieces of content that are posted to our platform every day. State-of-the-art AI systems flag content that may violate our policies, users report content to us they believe is questionable, and our own teams review content. We've also been building a parallel viral content review system to flag posts that may be going viral—no matter what type of content it is—as an additional safety net. This helps us catch content that our traditional systems may not pick up. We used this tool throughout this election, and in countries around the world, to detect and review Facebook and Instagram posts that were likely to go viral and took action if that content violated our policies.

For more on our work to remove deceptive campaigns around the world—whether foreign or domestic—please see the response to your Question 2.

Question 6. How the U.S. Government improved information sharing about threats from foreign actors seeking to interfere in our elections since 2016? Is information that is shared timely and actionable? What more can be done to improve the cooperation to stop threats from bad actors?

Answer. We work closely with law enforcement, regulators, election officials, researchers, academics, and civil society groups, among others, to strengthen our platform against election interference and the spread of misinformation. This engagement is incredibly important—we can't do this alone, and we have also worked to strengthen our relationships with government and outside experts in order to share information and bolster our security efforts.

With respect to our election protection work, we engaged with state attorneys general and other federal, state, and local law enforcement officials responsible for election protection. When they identified potential voter interference, we investigated and took action if warranted, and we have established strong channels of communication to respond to any election-related threats.

Question 7. How are you working with civil society groups like the University of Washington's Center for an Informed Public and Stanford Internet Observatory and Program?

Answer. We believe that there is a lot to learn from this election, and we're committed to making sure that we do. Earlier this year, we announced a partnership with a team of independent external academics to conduct objective and empirically grounded research on social media's impact on democracy. We want to better understand whether social media makes us more polarized as a society, or if it largely reflects the divisions that already exist; if it helps people become more informed about politics, or less; or if it affects people's attitudes towards government and democracy, including whether and how they vote. We hope that the insights these researchers develop will help advance society's understanding of the intersection of technology and democracy and help Facebook learn how we can better play our part.

Facebook is working with a group of seventeen independent researchers who are experts in the fields of elections, democracy, and social media. Social Science One facilitated the start of the project, and two of its committee chairs, Talia Stroud and Joshua A. Tucker, serve as cochairs of this project. They selected researchers who represent a variety of institutions, disciplines, areas of expertise, and methodological traditions. Facebook did not select the researchers and is taking measures to ensure that they operate independently.

Three principles guide our work and will continue to do so as we move ahead: independence, transparency, and consent.

- *Independence:* The external researchers won't be paid by Facebook, and they won't answer to Facebook either. Neither the questions they've asked nor the conclusions they draw will be restricted by Facebook. We've signed the same contracts with them that we do with other independent researchers who use our data (and those contracts are publicly posted on Social Science One's website).
- *Transparency:* The researchers have committed to publish their findings in academic journals in open access format, which means they will be freely available to the public. Facebook and the researchers will also document study plans and hypotheses in advance through a preregistration process and release those initial commitments upon publication of the studies. This means that people will be able to check that we did what we said we would—and didn't hide any of the results. In addition, to allow others to run their own analyses and further check our homework, we plan to deliver de-identified data on the studies we run. We have also invited Michael Wagner, a professor at the University of Wisconsin, to document and publicly comment on our research process as an independent observer.
- *Consent:* We are asking for the explicit, informed consent from those who opt to be part of research that analyzes individual-level data. This means research participants will consent to the use of their data and confirm that they understand how and why their data will be used.

Additionally, as part of our studies, we will also analyze aggregated user data on Facebook and Instagram to help us understand patterns. The studies—and our consent language—were reviewed and approved by an Institutional Review Board (IRB) to ensure they adhere to high ethical standards.

Question 8. How are you raising social media users' awareness about these threats? What more can be done? How do you ensure the actions you take do not cross the line into censorship of legitimate free speech?

Answer. With respect to our work to remove deceptive campaigns around the world—whether foreign or domestic—please see the response to your Question 2.

With respect to our work around misinformation more generally, people often tell us they don't want to see misinformation. People also tell us that they don't want Facebook to be the arbiter of truth or falsity. That's why we work with over 80 independent third-party fact-checkers who are certified through the non-partisan International Fact-Checking Network (IFCN) to help identify and review false news. If content is deemed by a fact-checker to be False, Altered, or Partly False, according to our public definitions, its distribution will be reduced, and it will appear lower in News Feed. We also implement an overlaid warning screen on top of factchecked content. People who try to share the content will be notified of the fact-checker's reporting and rating, and they will also be notified if content they have shared in the past has since been rated false by a fact-checker.

We also want Facebook to be a place where people can discover more news, information, and perspectives, and we are working to build products that help. Through our News Feed algorithm, we work hard both to actively reduce the distribution of clickbait, sensationalism, and misinformation and to boost news and information that keeps users informed, and we know the importance to users of staying informed about their local communities.

At Facebook, we connect people with reliable information about important issues. For example, since the pandemic started, we have worked to connect people with authoritative health sources through a number of different methods, such as redirecting people to health authorities if they searched for COVID–19 on Facebook or Instagram, and launching a COVID–19 Information Center on Facebook, which acts as a central place for people to get the latest news, information from health authorities, resources, and tips to stay healthy and safe. Between January and June, we directed over 2 billion people globally to resources and health authorities through our COVID–19 Information Center and pop-ups on Facebook and Instagram, with over 600 million people clicking through to learn more. In May, more than 25 million people in the U.S. visited the COVID–19 Information Center. More than 18 million people visited the COVID–19 Information Center in June and more than 14 million people in July.

When it came to the election, we launched a Voting Information Center to connect people with reliable information on deadlines for registering and voting and details about how to vote by mail or vote early in person, and we displayed links to the Voting Information Center when people posted about voting on Facebook. More than 140 million people have visited the Voting Information Center on Facebook and Instagram since it launched.

Additionally, we launched a Climate Science Information Center on Facebook to provide persistent access to global, regional, and local authoritative information about climate change and its effects. The Center features resources from the world's leading climate organizations and clear steps people can take to combat climate change. We're working with the Intergovernmental Panel on Climate Change (IPCC) and their global network of climate science contributors to include facts, figures, and data. Contributors include the UN Environment Programme (UNEP), The National Oceanic and Atmospheric Administration (NOAA), and the World Meteorological Organization (WMO). We'll also include posts from relevant sources to highlight climate science news.

Foreign Disinformation & Russian Election Interference. Since four years ago, our national security agencies and the private sector have made improvements to address foreign cyber and influence efforts that target our electoral process. However, there still needs to be more public transparency about foreign disinformation.

We need to close any gaps to stop any foreign disinformation about the 2020 election and disinformation in future elections. We cannot allow the Russians or other foreign actors to try to delegitimize election results or exacerbate political divisions any further.

Question 9. What more could be done to maximize transparency with the public about suspected foreign malign activity?

Answer. Please see the response to your Question 2.

Question 10. How could you share more information about foreign disinformation threats among the private sector tech community and among social media platforms and with smaller companies?

Answer. Please see the response to your Question 1.

Question 11. What should the U.S. Government be doing to promote information sharing on threats and to increase lawful data-sharing about suspected foreign malign activity?

Answer. Information sharing among the industry and the government has improved over the past few years, and we work closely with law enforcement, industry partners, and civil society. That said, the industry would benefit from a clear legal framework regarding data sharing in the context of investigating inauthentic and harmful influence operations.

We continuously look for ways to enhance our collaboration with the industry and the security research community while ensuring that we put the right checks in place to protect people's information, because we know that inauthentic behavior is not limited to a specific type of technology or service. The better we can be at working together with industry and outside security researchers, the better we'll do by our community.

Rohingya/Myanmar. In 2018, Facebook was weaponized against to whip up hate against the Muslim minority—the Rohingya. Myanmar held a general election last month. Prior to that election, there were concerns about the integrity of that election.

Question 12. What did you do and how are you continuing to make sure social media is not abused by any foreign or domestic actors to distort the electoral process in Myanmar and other countries?

Answer. We have invested heavily in people, technology, and partnerships over the past several years to examine and address the abuse of Facebook in Myanmar, and we have repeatedly taken action against violent actors and bad content on Facebook in Myanmar. We've also built a team that is dedicated to Myanmar. The ethnic violence happening in Myanmar is horrific, and we don't want our services to be used to spread hate, incite violence, or fuel tension on the ground.

Our approach to this problem, like the problem itself, is multifaceted, but our purpose is clear: to reduce the likelihood that Facebook will be used to facilitate offline harm. Our tactics include identifying and removing fake accounts; finding and removing violent actors; building better tools and technology that allow us to proactively find bad content; evolving our policies; and continuing to build partnerships and programs on the ground.

Impact of S. 4534. As you are aware, Chairman Wicker and two of our Republican colleagues have offered legislation to amend Section 230 to address, among other issues, what they call 'repeated instances of censorship targeting conservative voices."

That legislation would make significant changes to how Section 230 works, including limiting the categories of content that Section 230 immunity would cover and making the legal standard for removal of content more stringent. Critics of the Chairman's bill, S. 4534, suggest that these changes would inhibit companies' ability to remove false or harmful content from their platforms.

Question 13. I would like you to respond yes or no as to whether you believe that bills like the Chairman's would make it more difficult for Facebook to remove the following types of content—

- Bullying?
- Election disinformation?
- Misinformation or disinformation related to COVID–19?
- Foreign interference in U.S. elections?
- Efforts to engage in platform manipulation?
- Hate speech?
- Offensive content directed at vulnerable communities or other dehumanizing content?

Answer. Broadly speaking, Section 230 is a foundational law that allows us to provide our products and services to users. At a high level, Section 230 does two things. First, it encourages free expression. Without Section 230, platforms could potentially be held liable for everything people say. Without this protection, platforms would likely remove more content to avoid legal risk and would be less likely to invest in technologies that enable people to express themselves in new ways. Second, it allows platforms to remove harmful content. Without Section 230, platforms could face liability for doing even basic moderation, such as removing bullying and harassment that impact the safety and security of their communities. Repealing Section 230 entirely would likely substantially increase many companies' costs associated with legal challenges and content moderation.

Combating "Garbage" Content. Santa Clara University Law Professor Eric Goldman, a leading scholar on Section 230, has argued that the Online Freedom and

Viewpoint Diversity Act (S. 4534) wants Internet services to act as "passive" receptacles for users' content rather than content curators or screeners of "lawful but awful" third-party content.

He argues that the bill would be counterproductive because we need less of what he calls "garbage" content on the Internet, not more. Section 230 lets Internet services figure out the best ways to combat online trolls, and many services have innovated and invested more in improving their content moderation functions over the past few years.

Professor Goldman specifically points out that the bill would make it more difficult for social media companies to remove "junk science/conspiracy theories, like anti-vax content or quack COVID19 cures."

Question 14. Would S. 4534—and similar bills—hurt efforts by Facebook to combat online trolls and to fight what Professor Goldman calls "lawful but awful . . . garbage" content?

Answer. Please see the response to your Question 13.

The FCC's Capitulation to Trump's Section 230 Strategy. The Chairman of the Federal Communications Commission, Ajit Pai, announced recently that he would heed President Trump's call to start a rulemaking to "clarify" certain terms in Section 230.

And reports suggest that the President pulled the renomination of a sitting FCC Commissioner due to his concerns about that rulemaking, replacing him with a nominee that helped develop the Administration's petition that is the foundation of this rulemaking. This capitulation to President Trump by a supposedly independent regulatory agency is appalling.

It is particularly troubling that I—and other members of this committee—have been pressing Chairman Pai to push the envelope to interpret the agency's existing statutory authority to, among other things, use the E-Rate program to close the homework gap, which has only gotten more severe as a result of remote learning, and to use the agency's existing authority to close the digital divide on Tribal lands. And we expressed serious concern about Chairman Pai's move to repeal net neutrality, which the FCC majority based upon a highly conservative reading of agency's statutory authority.

In contrast, Chairman Pai is now willing to take an expansive view of the agency's authority when asked to support the President's pressure campaign against social media in an attempt not to fact check or label the President's posts.

Question 15. What are your views on Chairman Pai's announced rulemaking and the FCC's legal analysis of section 230? Would you agree that his approach on this issue is in tension with his repeal of the essential consumer protections afforded by the net neutrality rules?

Answer. Please see the response to your Question 13.

Addressing Bad Actors. I have become increasingly concerned with how easy it is for bad actors to use social media platforms to achieve their ends, and how Facebook has been too slow to stop it. For example, a video touting antimalarial drug hydroxychloroquine as a "cure" for COVID was eventually taken down this summer—but not after garnering 17 million views on Facebook.

In May, the watchdog group Tech Transparency Project concluded that white supremacist groups are "thriving" on Facebook, despite assurances that Facebook does not allow such groups on its platform.

These are obviously troubling developments, especially in light of the millions of Americans that rely on your services. You have to do better.

That said, I am not sure that modifying Section 230 is the solution for these and other very real concerns about your industry's behavior.

Question 16. From your company's perspective, would modifying Section 230 prevent bad actors from engaging in harmful conduct?

Answer. Please see the response to your Question 13.

Question 17. What do you recommend be done to address the concerns raised by the critics of Section 230?

Answer. Section 230 made it possible for every major Internet service to be built and ensured important values like free expression and openness were part of how platforms operate. Changing it is a significant decision. However, we believe Congress should update the law to make sure it's working as intended. We support the ideas around transparency and industry collaboration that are being discussed in some of the current bipartisan proposals, and we look forward to a meaningful dialogue about how we might update the law to deal with the problems we face today.

Potential Impacts of Changes to Section 230. Section 230 has been foundational to the development of the Internet of today. Most believe that absent Section 230, we would not have the massive, worldwide public forum the Internet provides.

Of course, we all understand that this forum may not be an unmitigated good, but it is equally true that Internet is a far more vibrant place than traditional media, because of the ability of users to contribute their thoughts and content.

Question 18. How do you expect that Facebook would react when faced with increased possibility of litigation over user-submitted content?

Answer. Defending lawsuits related to users' content on our platform requires a substantial amount of resources, including litigation costs and employee time, both in the United States and elsewhere.

The costs of litigation are often substantial, even when the suits are dismissed on Section 230 grounds.

Enforcement of Facebook's Content Policies. Mr. Zuckerberg, Facebook has rules prohibiting the promotion of violence and the spread of certain false claims. However, these rules mean nothing without consistent enforcement.

The *Wall Street Journal* recently put Facebook's content moderation efforts to the test. The results were alarming.

The *Journal* found that Facebook enforced its rules against misinformation and promoting violence inconsistently.

In the test, the *Journal* flagged a large number of posts that appeared to violate Facebook's own rules, but it turned out that Facebook's content review system left lots of rule-violating material online.

In many instances, Facebook did not review content flagged by the Journal within the 24-hour time period in which they promised to respond.

Question 19. Mr. Zuckerberg, will you commit to improving Facebook's enforcement of its own content moderation policies? What steps are you taking to improve your content review technology and other practices?

Answer. We have over 35,000 people working on safety and security, about 15,000 of whom review content. The majority of our content reviewers are people who work full-time for our partners and work at sites managed by these partners. We have a global network of partner companies so that we can quickly adjust the focus of our workforce as needed. This approach gives us the ability to, for example, make sure we have the right language or regional expertise. Our partners have a core competency in this type of work and are able to help us adjust as new needs arise or when a situation around the world warrants it.

We have also introduced tools that allow us to proactively detect and remove certain violating content using advances in technology, including artificial intelligence, machine learning, and computer vision. We do this by analyzing specific examples of bad content that have been reported and removed to identify patterns of behavior. Those patterns can be used to teach our software to proactively identify similar content.

These advances in technology mean that we can now remove bad content more quickly, identify and review more potentially harmful content, and increase the capacity of our review team. To ensure the accuracy of these technologies, we constantly test and analyze our systems, technology, and AI. All content goes through some degree of automated review, and we use human reviewers to check some content that has been flagged by that automated review or reported by people that use Facebook. We also use human reviewers to perform reviews of content that was not flagged or reported by people, to check the accuracy and efficiency of our automated review systems. The percentage of content that is reviewed by a human varies widely depending on the type and context of the content, and we don't target a specific percentage across all content on Facebook.

Question 20. On average, how long does Facebook take to review content flagged by users?

Answer. Most of the content we remove we find ourselves through automated systems. A significant portion of that is detected and removed immediately after it is uploaded. We work to remove this content as quickly as possible, though in some cases it may require human review to understand the context in which material was posted and to confirm if it violates our Community Standards.

Question 21. Do you agree that Facebook should remove flagged content promoting violence or misinformation within 24 hours? Will you commit to speeding up Facebook's review process?

Answer. We are proud of the work we have done to make Facebook an unwelcome place for those committed to acts of violence. In fact, our Dangerous Individuals and Organizations policy has long been the broadest and most aggressive in the indus-

try. And in August 2020, we expanded that policy to address militarized social movements and violence-inducing conspiracy networks, such as QAnon. The purpose of this policy is to prevent offline harm that may be related to content on Facebook, and so in the course of that work we contact law enforcement if we see imminent credible threats on the platform. We remove language that incites or facilitates serious violence. We also ban groups that proclaim a hateful and violent mission from having a presence on our apps, and we remove content that represents, praises, or supports them.

As for misinformation, people often tell us they don't want to see it on our platforms. That's why we work with over 80 independent third-party fact-checkers who are certified through the non-partisan International Fact-Checking Network (IFCN) to help identify and review false news. If content is deemed by a fact-checker to be False, Altered, or Partly False, according to our public definitions, its distribution will be reduced, and it will appear lower in News Feed. We also implement an overlaid warning screen on top of content marked as false. People who try to share the content will be notified of the fact-checker's reporting and rating, and they will also be notified if content they have shared in the past has since been rated false by a fact-checker.

We send content to independent third-party fact-checkers for review, but it is ultimately at their discretion to decide what to rate. The enqueued content is based on a number of signals, including machine learning-driven insights and false news reports by users, and we also allow third-party fact-checkers to enqueue content themselves.

We do not share data on how long it takes to fact-check content or how many views a post gets on average before it's fact-checked because these numbers may vary depending on the content; for example, claims related to breaking news or a complex issue may take more time to verify than content that repeats previously debunked claims. We surface signals to our fact-checking partners to help them prioritize what to rate. For example, fact-checking partners can see the estimated number of shares a post has received in the past 24 hours, and how many users have flagged it as potentially false in their News Feed. We also recognize that thorough reporting can take time. This is one of the reasons that we work with independent fact-checking partners, whose work can involve calling primary sources, analyzing videos/images, consulting public data, and more. We continue to have an open dialogue with partners about how we could further improve efficiency. We are testing ways to group content in one place to make it easier for fact-checking partners to find relevant content to review, faster.

Online Disinformation. I have serious concerns about the unchecked spread of disinformation online. From false political claims to harmful health information, each day the problem seems to get worse and worse. And I do not believe that social media companies—who make billions of dollars from ads based in part on user views of this disinformation—are giving this problem the serious attention that it deserves.

Question 22. Do you agree that Facebook can and should do more to stop the spread of harmful online disinformation?

Answer. Please see the response to your Question 21.

Question 23. Can you commit that Facebook will take more aggressive steps to stop the spread of this disinformation? What specific additional actions will you take?

Answer. Please see the response to your Question 21.

Question 24. About ten years ago, Microsoft attempted to gain a position in the ad server market with their Atlas product. They failed, and Facebook acquired the Atlas business from Microsoft in 2013. At the time, you asserted that Facebook would be a dominant player in this sector, but by 2017 Facebook had discontinued Atlas and announced that you would be exiting the ad server business. Today, Google controls about 90 percent of this business. How did Google out-compete Microsoft and Facebook? Did Google offer a superior product? Did Google have better relationships in the industry? What did Microsoft and Facebook fundamentally misunderstand about the ad server business?

Answer. Facebook invests heavily in research and development and seeks to continuously offer new products, as well as refine existing ones, in order to deliver innovative products and experiences to consumers. Facebook's goal in any acquisition is to maximize the use and benefit of the acquired company's assets in order to achieve the strategic need for which the acquisition was undertaken. Sometimes, an acquisition is not as successful as we hoped, and we make the business decision to discontinue the product or service.

Trump Administration Records. Over the course of nearly four years, President Trump and senior officials in his administration have routinely used social media to conduct government business, including announcing key policy and personnel decisions on those platforms. In addition, many believe that President Trump and his senior aides have used social media to engage in unethical and sometimes illegal conduct.

For example, Special Counsel Mueller cited several of President Trump's tweets as evidence of potentially obstructive conduct, and senior White House aides such as Kellyanne Conway and Ivanka Trump have been cited for violations of the Hatch Act and the misuse of position statute based on their use of Twitter in the conduct of their government jobs. Meanwhile, it appears that on several occasions Twitter has changed or ignored its rules and policies in ways that have allowed administration officials to continue using the platform to violate the rules for government employees and other Twitter users.

While government officials are legally obligated to preserve presidential and Federal records created or stored on social media platforms, this administration's actions cast serious doubts on whether they will comply with those obligations, and in many instances, they have already failed to do so. Facebook could play a vital role in ensuring that the historical record of the Trump administration is accessible to the American public,

Congress, and other government institutions so that people are "able to see and debate" the "words and actions" of the Trump presidency as well as future presidential administrations.

Question 25. Please describe what steps, if any, Facebook has taken to ensure that Facebook content—including posts and direct messages—published, sent, or received by Trump administration officials on Facebook accounts used for official government business are collected and preserved by your company.

Answer. We comply with our obligations under the law to preserve content posted on Facebook.

We disclose account records in accordance with our terms of service and applicable law.

Question 26. Please describe what steps, if any, Facebook has taken to ensure that the National Archives and Records Administration can obtain and preserve all Facebook content—including posts and direct messages—posted, sent, or received by Trump administration officials on Facebook accounts used for official government business.

Answer. Please see the response to your previous question.

Question 27. Please describe what steps, if any, Facebook has taken to ensure that the White House can preserve all Facebook content—including posts and direct messages—posted, sent, or received by Trump administration officials on Facebook accounts used for official government business.

Answer. Please see the response to your Question 25.

Question 28. Will you commit to ensuring that all Facebook content—including posts and direct messages—posted, sent, or received by Trump administration officials on Facebook accounts used for official government business are collected and preserved by your company?

Answer. Please see the response to your Question 25.

Question 29. How much time does an average user spend on your service if they see a news article on their timeline in that session, compared to a user who does not see a news article in their session? In what percentage of user sessions do users interact with external news content?

Answer. We know that one of the biggest issues social networks face is that, when left unchecked, people will engage disproportionately with more sensationalist and provocative content. At scale this type of content can undermine the quality of public discourse and lead to polarization. In our case, it can also degrade the quality of our services. Our research suggests that no matter where we draw the line for what is allowed, as a piece of content gets close to that line, people will engage with it more on average—even when they tell us afterwards they don't like the content. That is why we've invested heavily and have taken steps to try and minimize the amount of divisive news content people see in News Feed, including by reducing the distribution of posts containing clickbait headlines.

On Facebook, people see posts from their friends, Pages they've chosen to follow, and Groups they've joined, among others, in their News Feed. On a given day, the number of eligible posts in a user's Feed inventory can number in the thousands, so we use an algorithm to personalize how this content is organized. The goal of the News Feed algorithm is to predict what pieces of content are most relevant to the individual user, and rank (*i.e.,* order) those pieces of content accordingly every

time a user opens Facebook, to try and bring those posts that are the most relevant to a person closer to the top of their News Feed. This ranking process has four main elements: the available inventory (all of the available content from the people, Pages, and Groups a person has chosen to connect with); the signals, or data points, that can inform ranking decisions (*e.g.,* who posted a particular piece of content); the predictions we make, including how likely we think a person is to comment on a story, share with a friend, etc.; and a relevancy score for each story, which informs its position in News Feed.

We frequently make changes to the algorithm that drives News Feed ranking in an effort to improve people's experience on Facebook. For example, in 2018, we responded to feedback from our community that public content—posts from businesses, brands, and media—was crowding out the personal moments that lead us to connect more with each other. As a result, we moved from focusing only on helping people find relevant content to helping them have more meaningful social interactions. This meant that people began seeing more content from their friends, family, and Groups. We also reduce the distribution of some problematic types of content, including content that users may find spammy or low-quality, such as clickbait headlines, misinformation as confirmed by third-party fact-checkers, and links to low-quality webpages like ad farms.

Question 30. What are the clickthrough rates on your labelling on disputed or fact-checked content related to civic integrity, either when content is hidden or merely labelled? What metrics do you use to gauge the effectiveness of labelling? Please share typical numerical values of the metrics you describe.

Answer. Facebook works with third-party fact-checkers to review and rate the accuracy of content. Content across Facebook and Instagram that has been rated false or altered is prominently labeled so people can better decide for themselves what to read, trust, and share. These labels are shown on top of false and altered photos and videos, including on top of Stories content on Instagram, and link out to the assessment from the fact-checker.

We have studied the impact of labels when it comes to COVID–19 misinformation. During March and April 2020, we displayed warnings on about 50 million posts related to COVID–19 on Facebook, based on around 7,500 articles by our independent fact-checking partners. When people saw those warning labels, 95 percent of the time they did not go on to view the original content.

Question 31. Mr. Zuckerberg, I understand that Facebook is paying some publishers for their content. But, there is very little transparency about this process. Would you explain to us your methodology for paying newspapers? How are you determining who to pay in the U.S.? Will you provide clear information to the marketplace that explains your methodology? Will you list all of the publishers you pay?

Answer. Mobile technology has fundamentally changed the way people discover and consume news, and this has resulted in real challenges for publishers. We understand these challenges and have worked with publishers to adapt to digital transformation. But *how* news is distributed on Facebook warrants further discussion. News organizations voluntarily post their content on Facebook because it helps them reach new and larger audiences, and ultimately those audiences drive additional revenue for them.

To date, we have built tools to help publishers increase their subscribers by driving people from Facebook links to publisher websites. Among other benefits, Facebook provides publishers with free, organic distribution of news (and other content), which grows the audience and revenue for news publishers; customized tools and products to help news publishers monetize their content; and initiatives to assist publishers to innovate with online news content, along with bringing indirect value to publishers such as brand awareness and community building.

Publishers are also able to decide when a reader sees a paywall on content they've found via Facebook. Publishers control the relationship with their readers, with subscription payments taking place directly on their owned and operated websites. We do not take any cut of the subscription revenue because we want that money to go toward funding quality journalism. Helping publishers reach new audiences has been one of our most important goals.

Beyond the distribution and revenue tools already mentioned, we've focused on meaningful collaboration with publishers. In 2017, we launched the Facebook Journalism Project (*https://www.facebook.com/journalismproject*), an initiative focused on building innovative and sustainable solutions to support journalism. In 2019 we announced a $300 million commitment (*https://www.facebook.com/journalismproject/facebook-supports-local-news*) to news programs, partnerships, and content—with a specific focus on local news. And later that year we launched Facebook News

(*https://www.facebook.com/news*), a section of Facebook dedicated solely to authoritative and informative news content.

During the COVID–19 pandemic, we announced a $100 million investment (*https://www.facebook.com/journalismproject/coronavirus-update-news-industry-support*) to support the news industry—$25 million in emergency grant funding for local news through the Facebook Journalism Project, and $75 million in additional marketing spend to move money over to news organizations around the world.

We've also focused on supporting the global fact-checking community's work—we partnered with the Independent Fact-Checking Network to launch a $1 million grant program (*https://www.facebook.com/journalismproject/coronavirus-grants-fact-checking*) to increase their capacity during this time.

RESPONSE TO WRITTEN QUESTION SUBMITTED BY HON. AMY KLOBUCHAR TO MARK ZUCKERBERG

Political Ads. Facebook and Google have committed to voluntarily implement some measures of the Honest Ads Act, like requiring disclosures and creating an ad library for political ads, but have never truly lived up to some requirements, such as fully disclosing which categories of users ads are targeting. The full disclosures of targeting based on sensitive categories, like perceived race, ethnicity, or partisan affiliation is critical, because Russia targeted African Americans more than any other group in 2016. Intelligence officials have also repeatedly confirmed Russia is interfering in the 2020 elections and using online platforms to do so.

Question 1. Will your company voluntarily implement all the provisions of the Honest Ads Act, including fully disclosing which groups of people are being targeted by political ads in a way that does not compromise user privacy?

Answer. Facebook is committed to transparency for all ads, including ads with political content. That's why we've endorsed the Honest Ads Act and have taken many steps laid out in the bill even though it hasn't passed yet.

Facebook believes that people should be able to easily understand why they are seeing ads, who paid for them, and what other ads those advertisers are running. Our Ad Library is a unique tool to shine a light on political and social issue ads—a public archive that allows people to see all the ads politicians and campaigns are running on Facebook and Instagram and those that have run in the past. This is an important step in making political ads more transparent and advertisers more accountable: the public can see every ad served to anyone in an easily searchable database.

Earlier this year, we announced changes to provide more transparency over who is using ads to try to influence voters and to give people more control over the ads they see:

- View audience size in the Ad Library: We've added ranges for Potential Reach, which is the estimated target audience size for each political, electoral, or social issue ad, so you can see how many people an advertiser wanted to reach with every ad.
- Better Ad Library search and filtering: We've added the ability to search for ads with exact phrases, better grouping of similar ads, and adding several new filters to better analyze results—*e.g.* audience size, dates, and regions reached. This allows for more efficient and effective research for voters, academics, or journalists using these features.
- Control over Custom Audiences from a list: We rolled out a control to let people choose how an advertiser can reach them with a Custom Audience from a list. These Custom Audiences are built when an advertiser uploads a hashed list of people's information, such as e-mails or phone numbers, to help target ads. This control is available to all people on Facebook and applies to all advertisers, not just those running political or social issue ads. People have always been able to hide all ads from a specific advertiser in their Ad Preferences or directly in an ad. But now they are able to stop seeing ads based on an advertiser's Custom Audience from a list—or make themselves eligible to see ads if an advertiser used a list to exclude them.
- See fewer political ads: Seeing fewer political and social issue ads is a common request we hear from people. That's why we added a new control that will allow people to see fewer political and social issue ads on Facebook and Instagram. This feature builds on other controls in Ad Preferences we've released in the past, like allowing people to see fewer ads about certain topics or remove interests.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. RICHARD BLUMENTHAL TO MARK ZUCKERBERG

For the following questions, please provide information about your firm's content moderation decisions related to election misinformation and civic integrity covering the 2020 election period.

Question 1. Please describe what processes were used to make decisions about labeling or taking down organic and paid content related to elections or civic integrity.

Answer. During the 2020 election, Facebook was committed to doing our part to help ensure everyone had the chance to make their voice heard. That meant helping people register and vote, clearing up confusion about the election, and taking steps to reduce the chances of election related violence and unrest.

We partnered with election officials to remove false claims about polling conditions and displayed warnings on more than 150 million pieces of election-related content after review by our independent, third-party fact-checkers. We put in place strong voter suppression policies prohibiting explicit or implicit misrepresentations about how or when to vote, as well as attempts to use threats related to COVID–19 to scare people into not voting. We also removed calls for people to engage in poll watching that used militarized language or suggested that the goal was to intimidate, exert control, or display power over election officials or voters, and we filtered civic groups out of recommendations.

As the ballots were counted, we deployed additional measures that we announced in advance of the election to help people stay informed and to provide reliable information. We partnered with Reuters and the National Election Pool to provide reliable information about election results in the Voting Information Center and notified people proactively as results became available. We added labels to posts about voting by candidates from both parties to direct people to reliable information. We also attached an informational label to content that sought to delegitimize the outcome of the election or discuss the legitimacy of voting methods. We provided reliable information to combat election and voting misinformation, such as displaying "Facts About Voting" in users' News Feed and as part of the Voting Information Center, including emphasizing the longstanding trustworthiness of mail-in voting, and other assessments from non-partisan experts designed to counter false claims about the election.

When it comes to ads, we blocked new political and issue ads during the final week of the campaign, given the limited time for candidates to contest new claims; we rejected ads that made premature declarations of victory or sought to delegitimize the election; and we temporarily blocked all political and social issue ads after the polls closed to reduce opportunities for confusion and abuse.

Question 2. How many posts were reported or identified as potentially containing election misinformation or violations of civic integrity policies?

Answer. We partnered with election officials to remove false claims about polling conditions, ultimately removing 120,000 pieces of content on Facebook and Instagram for violating our voter interference policies, and we displayed warnings on more than 150 million pieces of election-related content after review by our independent, third-party fact-checkers. We also removed calls for people to engage in poll watching that used militarized language or suggested that the goal was to intimidate, exert control, or display power over election officials or voters, and we filtered civic groups out of recommendations.

Additionally, we launched a Voting Information Center to connect people with reliable information on deadlines for registering and voting and details about how to vote by mail or vote early in person, and we displayed links to the Voting Information Center when people posted about voting on Facebook. More than 140 million people visited the Voting Information Center on Facebook and Instagram since it launched.

Question 3. How many posts had enforcement action taken for containing election misinformation or violations of civic integrity policies?

Answer. Please see the responses to your Questions 1 and 2.

Question 4. Who did your firm consult to draft and implement election misinformation and civic integrity policies?

Answer. We work closely with law enforcement, regulators, election officials, researchers, academics, and civil society groups, among others, to strengthen our platform against election interference and the spread of misinformation. This engagement is incredibly important—we can't do this alone, and we have also worked to strengthen our relationships with government and outside experts in order to share information and bolster our security efforts.

With respect to our election protection work, we engaged with state attorneys general and other federal, state, and local law enforcement officials responsible for election protection. When they identified potential voter interference, we investigated and took action if warranted. And we have established strong channels of communication to respond to any election-related threats.

We also consulted with civil rights experts and community members regarding our voter suppression and intimidation policies. For example, in May 2018, we began a civil rights audit led by Laura Murphy, a highly respected civil rights and civil liberties leader. Her work has helped us build upon crucial election-related efforts, such as expanding our policy prohibiting voter suppression.

Finally, when it comes to misinformation, including election-related misinformation, we work with over 80 independent, third-party fact-checkers who are certified through the nonpartisan International Fact-Checking Network ("IFCN") to help identify and review false news. If content is deemed by a fact-checker to be false or partly false, its distribution will be reduced, and it will appear lower in News Feed. We also implement an overlaid warning screen on top of content marked as false. People who try to share the content will be notified of the fact-checker's reporting and rating, and they will also be notified if content they have shared in the past has since been rated false by a fact-checker.

Question 5. Who made final decisions about labeling or taking down a post related to election misinformation or civic integrity? Who did that person or those persons consult?

Answer. Our content reviewers moderate content based on our Community Standards. We have made our detailed reviewer guidelines public to help people understand how and why we make decisions about the content that is and is not allowed on Facebook.

When it comes to misinformation, as discussed in the answer to your Question 4, we work with independent, third-party fact-checkers to help reduce the spread of false news and other types of viral misinformation. Third-party fact-checkers are responsible for rating content, and Facebook is responsible for evaluating the consequences of those ratings. If content is deemed by a fact-checker to be false or partly false, its distribution will be reduced, and it will appear lower in News Feed. We also implement an overlaid warning screen on top of content marked as false and notify users who try to share the content (or who have shared it in the past).

Question 6. Does a different or specialize process exist for content from Presidential candidates, and if so, how does that process for review differ from the normal review?

Answer. Our Community Standards apply to all content, and we assess everyone under those Standards. Since 2016, we've also had a newsworthiness policy. First, we make a holistic determination about whether content falls within our newsworthiness policy. In the case of politicians' speech, for example, we presume a public interest value but will still evaluate it against the risk of harm. Second, the newsworthiness exception only applies to organic content; all ads, including those posted by politicians, must still comply with both our Community Standards and our Advertising Policies. Third, decisions to apply the newsworthiness policy often involve extensive internal deliberation and are made with low frequency. In 2019, for example, we made only fifteen newsworthiness exceptions for politicians globally, only one of which applied to a U.S. politician. More often, our newsworthiness policy has allowed for images that depict war or famine or attempt to raise awareness of issues like indigenous rights.

When it comes to speech from politicians, we don't believe that it's an appropriate role for us to referee political debates and prevent a politician's speech from reaching its audience and being subject to public debate and scrutiny. Speech from candidates and elected officials is some of the most scrutinized speech in our society, and we believe people should decide what is credible, not tech companies. That's why direct speech from politicians is not eligible for our independent, third-party fact-checking program. We have had this policy on the books for more than two years now, posted publicly on our site under our fact-checking program policies. This policy applies equally to all candidates for Federal public office, including presidential candidates.

Our policies don't mean that politicians can say whatever they want on Facebook. They can't spread misinformation about where, when, or how to vote, for example, or incite violence. And when a politician shares previously debunked content, including links, videos, and photos, we demote that content, display related information from fact-checkers, and reject its inclusion in advertisements. When it comes to ads, while we won't remove politicians' ads based solely on the outcome of a fact-check, we still require them to follow our Advertising Policies.

Question 7. Based on enforcement actions taken, there a discernible difference in engagement between a labeled post and unlabeled posts? Please provide any supporting information.

Answer. As discussed in further detail in the response to your Question 8, Facebook works with third-party fact-checkers to review and rate the accuracy of content. Content across Facebook and Instagram that has been rated False or Altered is prominently labeled so people can better decide for themselves what to read, trust, and share. These labels are shown on top of false and altered photos and videos, including on top of Stories content on Instagram, and they link out to the assessments from the fact-checkers.

We have studied the impact of labels when it comes to COVID–19 misinformation. During March and April 2020, we displayed warnings on about 50 million posts related to COVID–19 on Facebook, based on around 7,500 articles by our independent fact-checking partners. When people saw those warning labels, 95 percent of the time they did not go on to view the original content.

Question 8. What was the average time to add a misinformation label to a post?

Answer. People often tell us they don't want to see misinformation. That's why we work with over 80 independent, third-party fact-checkers who are certified through the non-partisan International Fact-Checking Network ("IFCN") to help identify and review false news. If content is deemed by a fact-checker to be false or partly false, its distribution will be reduced, and it will appear lower in News Feed. We also implement an overlaid warning screen on top of content marked as false. People who try to share the content will be notified of the fact-checker's reporting and rating and they will also be notified if content they have shared in the past has since been rated false by a fact-checker.

We send content to independent, third-party fact-checkers for review, but it is ultimately at their discretion to decide what to rate. The enqueued content is based on a number of signals, including machine learning-driven insights and false news reports by users, and we also allow third-party fact-checkers to enqueue content themselves.

We do not share data on how long it takes to fact-check content or how many views a post gets on average before it's fact-checked because these numbers may vary depending on the content; for example, claims related to breaking news or a complex issue may take more time to verify than content that repeats previously debunked claims. We surface signals to our fact-checking partners to help them prioritize what to rate. For example, fact-checking partners can see the estimated number of shares a post has received in the past 24 hours, and how many users have flagged it as potentially false in their News Feed. We also recognize that thorough reporting can take time—this is one of the reasons that we work with independent fact-checking partners, whose work can involve calling primary sources, analyzing videos/images, consulting public data, and more. We continue to have an open dialogue with partners about how we could further improve efficiency. We are testing ways to group content in one place to make it easier for fact-checking partners to more quickly find relevant content to review.

For the following questions, please provide information about your firm's content moderation decisions related hate speech, election interference, civic integrity, medical misinformation, or other harmful misinformation over the previous year.

Question 9. How many pieces of content were reported by users to the platform related to hate speech, election interference, civic integrity, and medical misinformation, broken down by category?

Answer. To track our progress and demonstrate our continued commitment to making Facebook safe and inclusive, we release our Community Standards Enforcement Report (available at *https://transparency.facebook.com/community-standards-enforcement*) on a quarterly basis. This report shares metrics on how Facebook is performing in preventing and removing content that violates certain Community Standards, including: adult nudity and sexual activity, bullying and harassment, child nudity and sexual exploitation of children, terrorism, organized hate, fake accounts, hate speech, regulated goods, spam, suicide and self-injury, and violent and graphic content. We also share data in this report on our process for appealing and restoring content to correct mistakes made in our enforcement decisions.

In the first three quarters of 2020, Facebook removed over 54 million pieces of content for violating our hate speech policy. Of that violating content we actioned, we identified the vast majority before users reported it—almost 95 percent in the second and third quarters of 2020. When it comes to election-related misinformation, we partnered with election officials to remove false claims about polling conditions and displayed warnings on more than 150 million pieces of content after review by our independent, third-party fact-checkers. And for COVID–19-related mis-

information, in the second quarter of 2020, we displayed warnings on approximately 98 million pieces of content on Facebook worldwide based on COVID–19-related debunking articles written by our fact-checking partners. In the U.S., we displayed misinformation warning screens associated with fact-checks related to COVID–19 on over 13 million pieces of content in the U.S. in March; over 15 million in April; over 13 million in May; over 9.7 million in June; and over 9.3 million in July.

Question 10. How many pieces of content were automatically identified or identified by employees related to hate speech, election interference, civic integrity, and medical misinformation, broken down by category?

Answer. Please see the response to your previous question.

Question 11. Of the content reported or flagged for review, how many pieces of content were reviewed by humans?

Answer. Most of the content we remove we find ourselves through automated systems. A significant portion of that is detected and removed immediately after it is uploaded. We work to remove this content as quickly as possible, though in some cases it may require human review to understand the context in which material was posted and to confirm if it violates our Community Standards.

Question 12. How many pieces of content were subject to enforcement action? Please provide a break down for each type of enforcement action taken for each category.

Answer. Please see the response to your Question 9.

Question 13. For content subject to enforcement action due to violation of hate speech rules, please identify how many pieces of content targeted each type of protected category (such as race or gender) covered by your rules. Do you track this information?

Answer. We do not allow hate speech on Facebook. We define hate speech as violent or dehumanizing speech, statements of inferiority, calls for exclusion or segregation based on protected characteristics, or slurs. These characteristics include race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disability or disease. When the intent is clear, we may allow people to share someone else's hate speech content to raise awareness or discuss whether the speech is appropriate to use, to use slurs self referentially in an effort to reclaim the term, or for other similar reasons. More information about our hate speech enforcement is available at *https://transparency.facebook.com/communitystandards-enforcement#hate-speech.*

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. EDWARD MARKEY TO MARK ZUCKERBERG

Question 1. Mr. Zuckerberg, Laura W. Murphy and the Relman Colfax firm completed a two-year civil rights audit of Facebook in July 2020. In a blog post, Facebook's Chief Operating Officer, Sheryl Sandberg, stated that Facebook will not follow every recommendation made in the audit. Please identify the specific audit recommendations that Facebook will and will not follow. Please also provide a timeline for implementation of the recommendations that Facebook will follow.

Answer. There are no quick fixes to the issues and recommendations the Auditors have surfaced. Becoming a better company requires a deep analysis of how we can strengthen and advance civil rights at every level of our company. That is what this audit has been—but it is the beginning of the journey, not the end.

Over the course of the audit process, we have made significant progress in a number of critical areas. But the auditors have been extremely candid with their feedback, urging us to go further in a range of areas. We have already started to put some of the Auditors' recommendations into place, including:

- We're beginning the process of bringing civil rights expertise in-house, starting with a commitment to hire a civil rights leader who will continue to push us on these issues internally, and embedding staff with civil rights expertise on core teams.
- We've expanded our voter suppression policies since the 2016 and 2018 elections so that we now prohibit threats that voting will result in law enforcement consequences and attempts at coordinated interference, both of which have been known to intimidate and demobilize voters.
- We included a link that directs people to our Voting Information Center on all posts about voting, including those from politicians, the goal being that we help make sure people have accurate, real-time information about voting processes in their districts.

- We attached an informational label to content that discusses the legitimacy of the election or claims that lawful methods of voting like mail-in ballots led to fraud. This label provided reliable information about the integrity of the election and voting methods.
- We extended the protections we had in place for voting to the U.S. 2020 census by adopting a robust census interference policy, which benefited from the Auditors' input and months of consultation with the U.S. Census Bureau, civil rights groups, and census experts.
- We've gone beyond existing hate speech protections to ban ads that are divisive and include fear-mongering statements.
- We have taken meaningful steps to build a more diverse and inclusive workforce, committing to bring on 30 percent more people of color, including 30 percent more Black people, in leadership positions.
- We announced a $100 million investment in Black-owned small businesses, Black creators, and nonprofits that serve the Black community in the U.S., and a commitment to spend at least $100 million with Black-owned businesses, toward a goal of $1 billion in annual spend with diverse suppliers by the end of 2021.

We continue to review seriously the recommendations made by the Auditors and invest in ongoing civil rights infrastructure and long-term change.

Question 2. Mr. Zuckerberg, children and teens are a uniquely vulnerable population online, and a comprehensive Federal privacy law should provide them with heightened data privacy protections. Do you agree that Congress should prohibit online behavioral advertising, or "targeted marketing" as defined in S. 748, directed at children under the age of 13?

Answer. We are committed to protecting the privacy and safety of minors who use our services, and we've adapted our services to do so. For example, we've adopted more limited privacy settings for minors, and restrictions on features they can use, who they can connect with, and the content they can see (including ads). Additionally, Facebook does not allow children under the age of 13 on its service and does not collect data about children under 13 that would trigger parental consent or notification.

We look forward to working with your office on this legislation in the next Congress.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. GARY PETERS TO MARK ZUCKERBERG

Question 1. In the hearing, we discussed how Facebook is working with law enforcement to disrupt real world violence stemming from activity on your platform. How many threats has Facebook proactively referred to local or state law enforcement prior to being approached for a preservation request?

Answer. We have a long history of working successfully with the DOJ, the FBI, and other government agencies to address a wide variety of threats to our platform. We reach out to law enforcement whenever we see a credible threat of imminent harm, including threats of self-harm. We have been able to provide support to authorities around the world, including in cases where law enforcement has been able to disrupt attacks and prevent harm.

We cooperate with governments in other ways, too. For example, as part of official investigations, government officials sometimes request data about people who use Facebook. We have strict processes in place to handle these government requests, and we disclose account records in accordance with our terms of service and applicable law. We also have law enforcement response teams available around the clock to respond to emergency requests.

We will take steps to preserve account records in connection with official criminal investigations for 90 days pending our receipt of formal legal process. Law enforcement may submit formal preservation requests through Facebook's Law Enforcement Online Request System (*https://www.facebook.com/records*) or by mail. We also publish regular transparency reports that provide details on global government requests and our responses at *https://transparency.facebook.com/government-data-requests.*

Question 2. In the hearing, I asked you about a recent report that an internal Facebook researcher found in 2016 that "64 percent of all extremist group joins are due to our recommendation tools." When I asked you about that research, you said you were "not familiar with that specific study." However, audio from a recent

Facebook meeting recorded you criticizing the story internally to employees. Please explain your response at the hearing. Are you now aware of that specific study?

Answer. Mr. Zuckerberg did not immediately recall the study you were referencing. We apologize for the confusion.

The study in question was not produced by the team whose primary role at the company focuses on groups that commit violence and spread disinformation, so it's not the best lens through which to understand our work in those areas.

And the story's suggestion that we buried research on this topic or didn't act on it is false. The reality is we didn't adopt some of the product suggestions cited in the story because we pursued alternatives that we believe are more effective. For example, in 2018, we responded to feedback from our community that public content—posts from businesses, brands, and media—was crowding out the personal moments that lead us to connect more with each other. As a result, we moved from focusing only on helping users find relevant content to helping them have more meaningful social interactions. This meant that users began seeing more content from their friends, family, and Groups. We also reduce the distribution of some problematic types of content, including content that users may find spammy or low-quality, such as clickbait headlines and links to low-quality webpages like ad farms.

We also fund research on misinformation and polarization to better understand the impact of our products; for example, in February we announced an additional $2 million in funding for independent research on this topic.

We are proud of the work we have done to make Facebook an unwelcome place for those committed to acts of violence. In fact, our Dangerous Individuals and Organizations policy has long been the broadest and most aggressive in the industry. And in August 2020, we expanded that policy to address militarized social movements and violence-inducing conspiracy networks, such as QAnon. The purpose of this policy is to prevent offline harm that may be related to content on Facebook, and so in the course of that work we contact law enforcement if we see imminent credible threats on the platform. Accordingly, we remove language that incites or facilitates serious violence. We also ban groups that proclaim a hateful and violent mission from having a presence on our apps, and we remove content that represents, praises, or supports them.

Moving fast to find and remove dangerous organizations, including terrorist and hate groups, takes significant investment in both people and technology. At Facebook, we have tripled the size of our teams working in safety and security since 2016 to over 35,000 people—including teams that review reports of hate speech and content that praises, supports, or represents hate groups. We also have several hundred people who exclusively or primarily focus on countering dangerous organizations as their core responsibility. This group includes former academics who are experts on counterterrorism, former prosecutors and law enforcement agents, investigators and analysts, and engineers.

Four years ago, we developed a playbook and a series of automated techniques to detect content related to terrorist organizations such as ISIS, al Qaeda, and their affiliates. We've since expanded these techniques to detect and remove content related to other terrorist and hate groups. We're now able to detect text embedded in images and videos in order to understand its full context, and we've built media-matching technology to find content that's identical or near identical to photos, videos, text, and even audio that we've already removed. When we started detecting hate organizations, we focused on groups that posed the greatest threat of violence at that time, and we've now expanded to detect more groups tied to different hate-based and violent extremist ideologies and using different languages. In addition to building new tools, we've also adapted strategies from our counterterrorism work, such as leveraging off-platform signals to identify dangerous content on Facebook and implementing procedures to audit the accuracy of our AI's decisions over time.

We understand, however, that simply working to keep violence off Facebook is not an adequate solution to the problem of online content tied to violent extremism, particularly because bad actors can leverage a variety of platforms and operate offline as well. While we work 24/7 to identify, review, and remove violent extremist content, our efforts do not stop there. We believe our partnerships with other companies, civil society, researchers, and governments are crucial to combating this threat. For example, our P2P Global Digital Challenge, which engages university students around the world in competitions to create social media campaigns and offline strategies to challenge hateful and extremist narratives, has launched over 600 counter speech campaigns from students in 75 countries, engaged over 6,500 students, and reached over 200 million people. We have also developed the Redirect Initiative to connect people searching for violent extremist material with offline organizations dedicated to helping people disconnect from extremist groups. The program is active now in four countries, including the U.S., where we have partnered

with Life After Hate, an organization founded by former violent extremists, to help people disconnect from white supremacist groups.

Question 2a. What is the current percentage of extremist group joins due to Facebook recommendation tools?

Answer. Groups that represent hate organizations, terrorist organizations, militarized social movements, and violence-inducing conspiracy networks have no place on our platform, and we remove them when our technology or content review and investigative teams identify them.

Additionally, Pages and Groups that repeatedly violate other Community Standards or repeatedly share things found false by third-party fact-checkers are not eligible to appear in recommendations surfaces. We also apply a number of restrictions on accounts that violate these same rules, including by removing them from recommendations and limiting their ability to use surfaces like Live, if the account has not yet reached the threshold of violations at which we would remove the account entirely.

Question 2b. What policy and algorithm changes has Facebook made to reduce facilitation of extremist group recruitment since that time, and how effective have those changes been? Please share any data demonstrating the impacts of such changes.

Answer. Please see the responses to your Questions 2 and 2(a).

Question 3. Following the 2016 election, Facebook informed users that they had interacted with Russian disinformation, but this was a one-time occurrence around a very specific set of content. Do you think that Facebook users have a right to know if they've been exposed to content that your own policies have deemed so dangerous that you have removed it?

Question 3a. Facebook allows notifications by text, desktop pop-up, e-mail, and through the app. Facebook also has the capability to notify users if they have seen or interacted with content that content moderators have deemed harmful disinformation or extremist—this was demonstrated after the 2016 election. Why does Facebook not do this with other content that it has removed due to violations of your community standards?

Answer. Notifying users about content that was subsequently removed could have additional harmful consequences by re-exposing those users to hate speech, terrorism, or other types of harmful content that violates our Community Standards. For example, studies have shown that re-exposure to disinformation—even if condemnatory—can sometimes reinforce the original false message. We do generally notify users about subsequently removed content they did not post but had interacted with when the content poses a serious risk that the user could cause greater harm to themselves if not notified about its subsequent removal. Therefore, for instance, we warn users who interacted with harmful misinformation about COVID–19 that was later removed, so they don't mistakenly act on that misinformation.

Facebook also notifies users when they interact with information that has been rated false by a third-party fact-checker. We work with independent, third-party fact-checkers to help reduce the spread of false news and other types of viral misinformation on our platform. If content is deemed by a fact-checker to be false or partly false, its distribution will be reduced, and it will appear lower in News Feed. We also implement an overlaid warning screen on top of content marked as false. People who try to share the content will be notified of the fact-checker's reporting and rating, and they will also be notified if content they have shared in the past has since been rated false by a fact-checker. We also take action against Pages and domains that repeatedly share or publish content that is rated false. Such Pages and domains will see their distribution reduced as the number of offenses increases, including their eligibility for recommendations and ability to advertise and monetize. Finally, Pages and domains that repeatedly publish or share false news will also lose their ability to register as a News Page on Facebook, and if a registered News Page repeatedly shares false news, its News Page registration will be revoked.

Question 4. A recent article highlighted that five states—Georgia, Oregon, Pennsylvania, Wisconsin and Michigan—have the highest risk of increased militia activity around the elections, including everything from demonstrations to violence. Has Facebook taken concrete steps to identify pages or groups that are promoting violence in these states specifically and to proactively remove that content?

Answer. We remove content calling for or advocating violence, and we ban organizations and individuals that proclaim a violent mission. Because we saw growing movements that, while not necessarily directly organizing violence, have celebrated violent acts, shown that they have weapons and suggest they will use them, or have individual followers with patterns of violent behavior, we expanded our Dangerous Individuals and Organizations policy to address militia groups as well as other orga-

nizations and movements that have demonstrated significant risks to public safety, including QAnon. In the first two months since we expanded our policy to address these groups and movements, we identified over 600 militarized social movements, removing about 2,400 Pages, 14,200 Groups, and about 1,300 Instagram accounts they maintained. In addition, we've removed about 1,700 Pages, 5,600 Groups, and about 18,700 Instagram accounts representing QAnon. For more information, please visit *https://about.fb.com/news/2020/08/addressing-movements-and-organizations-tied-to-violence/*.

Question 4a. How many people/users have to see this kind of content before Facebook decides to take it down?

Answer. We have designated more than 600 militarized social movements based solely on the behavior of the entities themselves. When we find groups, Instagram accounts, or Pages that violate our policies against militarized social movements and violence-inducing conspiracy networks, we take action regardless of the number of users who have interacted with them.

Question 4b. Why did Facebook allow more than 360,000 individuals to join the "STOP THE STEAL" group before removing it for violating your community standards?

Answer. Facebook stands for giving people a voice, and it was important to us that everyone could make their voice heard during the election. We announced a series of policies in advance to help support the integrity of the election. For example, we put in place strong voter suppression policies prohibiting explicit or implicit misrepresentations about how or when to vote, as well as attempts to use threats related to COVID–19 to scare people into not voting. We also removed calls for people to engage in poll watching that used militarized language or suggested that the goal was to intimidate, exert control, or display power over election officials or voters.

When it came to the "Stop the Steal" group, we took down the group within about 24 hours. We removed the group because it was organized around the delegitimization of the election process, and we saw worrying calls for violence from some members of the group.

Question 5. Buzzfeed recently reported that, in discussing unrest around the 2020 election, you told Facebook employees "once we're past these events, and we've resolved them peacefully, I would not expect that we continue to adopt more policies that are restricting of content." Unfortunately, the threat of domestic terrorism will not evaporate after this election cycle. Will Facebook continue to review and rigorously enforce its existing community standards to stop the calls for violence and other extremist content beyond the election season and for as long as the threats persist?

Answer. Yes. Terrorists, terrorist content, and hate speech in all forms—including white supremacy and violent extremist content—have no place on Facebook, and have always been prohibited. That will not change. If we find content that praises or supports terrorists, violent extremists, or their organizations, we remove it. Indeed, of the content that we remove on this basis, we detect the vast majority of it before anyone reports it. In the first three quarters of 2020, we took action on over 24 million pieces of terrorism content, and we identified over 99 percent of that content before users reported it to us. In the same time, we took action on over 12 million pieces of content tied to hate organizations, and we now detect over 97 percent of that content before users report it to us.

Additionally, as discussed in the response to your Question 4(a), even before the election, we strengthened our enforcement against militias, violence-inducing conspiracy networks, and other groups that could be used to organize violence. We expanded our policy because we saw growing movements that, while not necessarily directly organizing violence, have celebrated violent acts, shown that they have weapons and suggest they will use them, or have individual followers with patterns of violent behavior. We remain committed to enforcing this policy going forward.

Question 6. Facebook's community standards often draw the line at specific threats of violence for the removal of content, rather than conspiracy theories that may set the predicate for radicalization and future action. When it comes to conspiracy theories and misinformation, Facebook often chooses not to remove content, but rather to reduce the spread and to attach warnings. What testing or other analysis has Facebook done that shows your work to reduce the spread of disinformation and misinformation is effective?

Answer. As discussed in the response to your Question 4, we are committed to combating violent voices that spread misinformation and conspiracy theories. In August 2020, we expanded our Dangerous Individuals and Organizations policy to address militarized social movements and violence-inducing conspiracy networks, such as QAnon. Since then, we've identified over 600 militarized social movements, re-

moving about 2,400 Pages, 14,200 Groups, and about 1,300 Instagram accounts they maintained. We've also removed about 1,700 Pages, 5,600 Groups, and about 18,700 Instagram accounts representing QAnon.

Additionally, we've long supported programs to empower users that want to push back on radicalization. This includes the Facebook Digital Challenge, the Online Civil Courage Initiative, and the Redirect Initiative, which we began with Life After Hate and now run in 4 countries. Most recently, we began a broad campaign with the Asia Foundation to support these programs across Asia. Our Redirect Initiative model has most recently been used around QAnon. We are providing links to reliable information for people that search for QAnon-related terms, and for people who search for QAnon-linked terms like "Save Our Children," we direct them to another set of links to legitimate child safety groups.

Although it is too early to draw comprehensive conclusions about reductions in the spread of misinformation ahead of the 2020 U.S. presidential election, research from 2018 and 2019 conducted by researchers at the University of Michigan, Princeton University, University of Exeter, and Washington University at St. Louis offers encouraging findings about the scale and spread of misinformation since the 2016 U.S. elections. Namely:

- Fake news exposure fell dramatically from 2016 to 2018. Researchers have found that there was a substantial decline (75 percent) in the proportion of Americans who visited fake news websites during the 2018 midterm elections, relative to the 2016 elections.
- Also during the 2016–2018 period, Facebook's role in the distribution of misinformation was dramatically reduced. To determine Facebook's role in spreading false news, researchers looked at the three websites people visited in the 30 seconds before arriving at a fake news site. Between the fall of 2016 and the summer and fall of 2018, Facebook's role in referring visits to fake news sites dramatically dropped.

Question 7. It is clear that the existence of conspiracy theories, disinformation campaigns, and misinformation has led to violence, even if not specifically planned on your platform.

Recently, Facebook has taken action against the QAnon conspiracy for this reason. While QAnon has led to numerous instances of violence in recent months and years, Facebook only banned it recently. Why did QAnon reach that threshold now, and how will Facebook address other conspiracies?

Question 7a. Is there some set number of violent incidents that must occur before Facebook considers a group unfit for the platform?

Answer. We remove any group that has proclaimed a violent mission or engaged in documented acts of terrorism. As discussed in the responses to your Questions 4 and 6, we recently expanded our Dangerous Individuals and Organizations policy to address organizations and movements that have demonstrated significant risks to public safety but do not meet the rigorous criteria to be designated as a dangerous organization and banned from having any presence on our platform. This includes militarized social movements and violence-inducing conspiracy networks, such as QAnon. While we will allow people to post content that supports these movements and groups, so long as they do not otherwise violate our content policies, we will restrict their ability to organize on our platform.

Under this policy expansion, we impose restrictions to limit the spread of content from Facebook Pages, Groups, and Instagram accounts. We also remove Pages, Groups, and Instagram accounts where we identify indications of potential violence, including when they use veiled language and symbols particular to the movement to do so.

We will take the following actions:

- Remove From Facebook: Pages, Groups, and Instagram accounts representing these movements and organizations will be removed. We will continue studying specific terminology and symbolism used by supporters to identify the language used by these groups and movements indicating violence and take action accordingly.
- Reduce in Search: Hashtags and titles of Pages, Groups, and Instagram accounts restricted on our platform related to these movements and organizations will be limited in Search: they will not be suggested through our Search Typeahead function and will be ranked lower in Search results.
- Prohibit Use of Ads, Commerce Surfaces, and Monetization Tools: Facebook Pages related to these movements will be prohibited from running ads or selling

products using Marketplace and Shop. We also prohibit anyone from running ads praising, supporting, or representing these movements.

- Prohibit Fundraising: We will prohibit nonprofits we identify as representing or seeking to support these movements, organizations, and groups from using our fundraising tools. We will also prohibit personal fundraisers praising, supporting, or representing these organizations and movements.

Since this policy update, we've identified over 600 militarized social movements, removing about 2,400 Pages, 14,200 Groups, and about 1,300 Instagram accounts they maintained. We've also removed about 1,700 Pages, 5,600 Groups, and about 18,700 Instagram accounts representing QAnon.

When it comes to QAnon in particular, we remove any Facebook Pages, Groups, and Instagram accounts representing QAnon. Additionally, when someone searches for terms related to QAnon on Facebook and Instagram, we will redirect them to credible resources from the Global Network on Extremism and Technology (GNET), which is led by Kings College in London. To address evidence that QAnon adherents are increasingly using the issue of child safety and hashtags like #savethechildren to recruit and organize, we also direct people to credible child safety resources when they search for certain child safety hashtags. These are the latest expansions of our Redirect Initiative to help combat violent extremism, through which we will direct people to resources that can help inform them of the realities of QAnon and its ties to violence and real-world harm.

We will also continue to review content and accounts against all of our content policies in an effort to keep people safe. We will remove content from these movements that violate any of our policies, including those against fake accounts, harassment, hate speech, or inciting violence. Misinformation that does not put people at risk of imminent violence or physical harm but is rated false by third-party fact-checkers will be reduced in News Feed so fewer people see it. And any non-state actor or group that qualifies as a dangerous individual or organization will be banned from our platform. Our teams will also continue to study trends in attempts to skirt our enforcement so we can adapt. These movements and groups evolve quickly, and our teams will follow them closely and consult with outside experts so we can continue to enforce our policies against them.

Question 8. When the Network Contagion Research Institute began mapping the spread of antigovernment "boogaloo" rhetoric on Facebook in early 2020, they saw advertisements to purchase items for boogaloo's desired civil war, including a boogaloo bag and themed weapon accessories. In a recent interview, the Institute's co-founder said "We realized the algorithms of Facebook have never met an apocalyptic, militant cult set on killing cops that they didn't like, and couldn't merchandise." Since the beginning of 2020, how much revenue did Facebook generate from ads purchased by, or targeting users engaging with, militia, boogaloo, or other extremist content?

Will you provide the Committee with relevant data around user engagement with boogaloo and other extremist content?

Have violent extremist groups used paid features of Facebook's various platforms? Do they buy ads?

Answer. Facebook is committed to banning people from our platform who proclaim a violent mission. In June, Facebook designated as a dangerous organization a violent network associated with the boogaloo movement. As a result, this violent network is banned from having a presence on our platform and we remove content praising, supporting, or representing it. This network appeared to be based across various locations in the U.S., and the people within it engaged with one another on our platform. It actively promoted violence against civilians, law enforcement, and government officials and institutions. Members of this network also sought to recruit others within the broader boogaloo movement, sharing the same content online and adopting the same offline appearance as others in the movement to do so. For more information, please visit *https://about.fb.com/news/2020/06/banning-a-violent-network-in-the-us/*.

All of our normal content policies apply to advertisements and commerce pages like Marketplace and Shop. That means that dangerous organizations may not be praised, supported, or represented on those surfaces.

Question 9. Once a group is designated under your Dangerous Individuals and Organizations policy, or any other Facebook policy, does Facebook stop them from purchasing ads, receiving targeted ads, being recommended to other users, creating new events, or inviting new members to join?

Answer. A group designated under our Dangerous Individuals and Organizations policy may not use our platform for any purpose, nor may it be praised, supported, or represented on our platform. This is the most aggressive policy in the industry.

Question 10. While I appreciate that Facebook continues to evolve and learn about threats of violence on the platform, would you agree that as groups evolve and change their tactics you will always be one step behind extremist groups that seek to use social media to recruit and plan violent acts? How do you address this problem?

Answer. We face determined, well-funded adversaries who will never give up and regularly change tactics. We need to constantly adapt and improve. We do that by employing in-house experts, building scalable AI tools, and aggressively and systematically engaging outside partners, including others in industry, governments, and academic experts. We have several hundred people internally at Facebook whose primary job at Facebook deals with dangerous organizations, including many who are academic experts or former law enforcement or intelligence personnel. They track these movements as they evolve, and we adjust our enforcement as a result. We also think that building AI tools is a scalable way to identify and root out most content that violates our policies. We are making substantial investments in building and improving these tools. For example, today, more than 99 percent of the terrorism content we remove from Facebook is content we detect before anyone in our community has flagged it to us. We do this primarily through the use of automated systems like photo and video matching and text-based machine learning. We also use AI to help find child exploitation images, hate speech, discriminatory ads, and other prohibited content.

We also work with others in the industry to limit the spread of violent extremist content on the Internet. For example, in 2017, we established the Global Internet Forum to Counter Terrorism (GIFCT) with others in the industry with the objective of disrupting terrorist abuse on digital platforms. Since then, the consortium has grown and collaborates closely on critical initiatives focused on tech innovation, knowledge sharing, and research.

Question 11. When prioritizing which content to evaluate, Facebook does not always consider the amount of time that content is on the platform but rather the spread. While this may make sense for disinformation, where the threat lies in misleading the population, when dealing with content to inspire violence, who sees the content can be more important than how many. As we have seen time and again, lone actors inspired to violence can cause significant harm. How do you address this issue?

Answer. Incitement to violence has no place on our platforms, regardless of who perpetrates it. Facebook is committed to keeping those who proclaim a violent mission off of our platforms. As soon as we identify content that violates our policies, we work to remove it. The time it takes between identifying the content and removing it may simply be a function of how long it takes to review the content—a 30 minute video will take longer to review than a text post—and determining if the content in the context in which it's shared violates our policies. We want to make sure our content review teams have the time they need to review content and make an accurate decision. For instance, we may evaluate whether a post violates our hate speech policy for attacking someone based on race, religion, or gender identity, or whether the post is someone raising awareness and condemning the hate speech that was directed at them. But when we do have high confidence that something violates our policies, we deploy a range of technology and human expertise to remove the content before more people are likely to see it.

In addition to taking down violating content, we focus most of our efforts on how often content that violates our policies is actually seen by someone. While content actioned describes how many things we took down, prevalence describes how much we haven't yet identified that people may still see. We measure this by periodically sampling content viewed on Facebook and then reviewing it to see what percent violates our community standards.

Views of violating content that contains terrorism are very infrequent, and we remove much of this content before people see it. As a result, many times we do not find enough violating samples to precisely estimate prevalence. In the third quarter 2020, this was true for violations of our policies on terrorism, child nudity and sexual exploitation of children, suicide and self-injury, and regulated goods on Facebook and Instagram. In these cases, we can estimate an upper limit of how often someone would see content that violates these policies.

In the third quarter of 2020, the upper limit was 0.05 percent for violations of our policy for terrorism on Facebook. This means that out of every 10,000 views of content on Facebook, we estimate no more than five of those views contained content that violated the policy.

For more information about Facebook's efforts to detect and remove violent and extremist content from its platforms, please see the responses to your previous questions.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. KYRSTEN SINEMA TO MARK ZUCKERBERG

COVID–19 Misinformation. United States remains in the midst of a global pandemic. More than 227,000 Americans have died of COVID–19, including nearly 6,000 in my home state of Arizona. COVID has impacted the health, employment, and education of Arizonans, from large cities to tribal lands like the Navajo Nation. And at the time of this hearing, the country is facing another significant surge in cases.

The persistent spread of COVID–19 misinformation on social media remains a significant concern to health officials. Digital platforms allow for inflammatory, dangerous, and inaccurate information—or outright lies—to spread rapidly. Sometimes it seems that misinformation about the virus spreads as rapidly as the virus itself. This misinformation can endanger the lives and livelihoods of Arizonans.

Social distancing, hand washing, testing, contact tracing, and mask wearing should not be partisan issues, nor should they be the subject of online misinformation.

Question 1. What has Facebook done to limit the spread of dangerous misinformation related to COVID–19 and what more can it do?

Answer. As people around the world confront the unprecedented COVID–19 public health emergency, we want to make sure that our Community Standards protect people from harmful content and new types of potential abuse related to COVID–19. We're working to remove content that has the potential to contribute to real-world harm, including through our policies prohibiting the coordination of harm, hate speech, bullying and harassment, and misinformation that contributes to the risk of imminent violence or physical harm. Oftentimes, misinformation can cut across different types of abuse areas, for example, a racial slur could be coupled with a false claim about a group of people, and we'd remove it for violating our hate speech policy. So in addition to our misinformation policies, we have a number of other ways we might combat COVID–19 misinformation, such as:

- Under our Coordinating Harm policy, we remove content that advocates for the spread of COVID–19, as well as content that encourages or coordinates the physical destruction of infrastructure, such as 5G masts, based on the false claim that they played a role in the spread of COVID–19. This also includes removing content coordinating in-person events or gatherings when participation involves or encourages people with COVID–19 to join.
- Under our Misinformation and Harm policy, we remove misinformation that contributes to the risk of imminent violence or physical harm. We have applied this policy to harmful misinformation about COVID–19 since January. Between March and October of this year, we removed more than 12 million pieces of content on Facebook and Instagram globally for containing misinformation that may lead to imminent physical harm, such as content relating to fake preventative measures or exaggerated cures.
- Under our Hate Speech policy, we are removing content that states that people who share a protected characteristic such as race or religion have the virus, created the virus, or are spreading the virus. This does not apply to claims about people based on national origin because we want to allow discussion focused on national-level responses and effects (*e.g.,* "X number of Italians have COVID–19"). We also remove content that mocks people who share a protected characteristic such as race or religion for having COVID–19. As reported in our Community Standards Enforcement Report (CSER), content actioned under our hate speech policy increased from 9.6 million pieces of content in Q1 2020 to 22.1 million in Q3 2020. That enforcement includes COVID–19-related content. Starting in Q1, we made improvements to our proactive detection technology and expanded automation to the Spanish, Arabic, and Indonesian languages. In Q2, we followed up by expanding automation to the English, Spanish, and Burmese languages, which helped us detect and remove more content.
- Under our Bullying and Harassment policy, we remove content that targets people maliciously, including content that claims that a private individual has COVID–19, unless that person has self-declared or information about their health status is publicly available. As reported in our CSER, content actioned under our bullying and harassment policy increased from 2.4 million in Q2 2020 to 3.5 million in Q3 2020, which includes COVID–19-related content. After enforcement was impacted by temporary workforce changes due to COVID–19, we regained some review capacity in Q2 and Q3. We also increased our automation abilities and made improvements to our proactive detection technology for the English language.

For misinformation that does not pose a safety risk but undermines the authenticity and integrity of our platform, we continue to work with our global network of independent, third-party fact-checking partners. Once a post is rated false or party false by a fact-checker on Facebook or Instagram, we reduce its distribution so fewer people see it, and we show warning labels and notifications to people who still come across content that has been rated, try to share it, or already have. Based on one fact-check, we're able to kick off similarity detection methods that identify duplicates of debunked stories and apply the same strong warning labels and demotions to those duplicates. In the second quarter of 2020, we displayed warnings on about 98 million pieces of content on Facebook worldwide based on COVID–19-related debunking articles written by our fact-checking partners.

As the situation evolves, we are continuing to look at content on the platform, assess speech trends, and engage with experts, and we will provide additional policy guidance to our Community Standards when appropriate, to keep the members of our community safe during this crisis.

Spreading Accurate Information. Arizonans need accurate, scientifically based information to help get through this pandemic. Many Arizonans get their news from sources such as Facebook. As a result, your companies can play a role in helping people receive accurate information that is relevant to their communities and can aid them in their decisions that keep their families healthy and safe.

For example, earlier this month, the CDC issued a report illustrating that COVID–19 cases fell dramatically in Arizona after prevention and control measures were put into place. I shared this information on social media, and this is the type of information we should emphasize to help save lives.

Question 2. What more can Facebook do to better amplify accurate, scientifically-based health information to ensure that Arizonans understand how best to protect themselves from the pandemic?

Answer. Please see the response to your Question 1. We've also seen people turn to social media during this global health emergency, finding novel ways to stay connected and informed during these difficult times. And since the pandemic started, we have worked to connect people with reliable health sources through a number of different methods, such as redirecting people to health authorities if they search for COVID–19 on Facebook or Instagram, and launching a COVID–19 Information Center on Facebook which acts as a central place for people to get the latest news, information from health authorities, and resources and tips to stay healthy and safe. Between January and June, we directed over 2 billion people globally to resources and health authorities through our COVID–19 Information Center and pop-ups on Facebook and Instagram, with over 600 million people clicking through to learn more.

Scientific Evidence-based COVID Information. Our best sources of information related to the pandemic are doctors, researchers, and scientists. We should be relying on their expertise to help stop the spread of the virus and help our country recover from its devastating impacts.

Question 3. Who determines whether content on Facebook is scientifically supported and evidence based?

Answer. We are working with health authorities and other experts to identify claims that are false and harmful, *i.e.,* claims where, if someone believes the information, it could cause physical harm to them by increasing the likelihood of them getting or spreading the disease.

We are also working to empower our fact-checking community during COVID–19. Our fact-checking program is a key piece of our multi-pronged strategy to reduce the spread of misinformation on our platforms. This is why, since January, we have taken a number of additional steps to support our fact-checking partners' work to debunk misinformation about COVID–19.

- Expanding our fact-checking network: We continue to expand our fact-checking network around the world. Globally, we have over 80 fact-checking partners, covering over 60 languages. In the U.S., we have 10 partners.
- Grant program to support fact-checkers during COVID–19: In March, we partnered with Poynter's International Fact-Checking Network (IFCN) to launch a $1 million grant program to support fact-checkers in their work around COVID–19. In addition to providing critical funding that enables partners to maintain or increase their capacity during this time, the grants also support projects such as:
    - Translation of fact-checks from native languages to different languages;

- Multimedia production (such as videos, infographics, podcasts) about COVID–19;
- Working with health experts for evidence-based and scientific coverage;
- Audience development initiatives that use innovative formats, such as offline or interactive communication, to better reach people with reliable information; and
- Fact-checkers supporting public authorities with reliable information for better communication about COVID–19.

Since we launched this program, we have awarded grants to 21 fact-checking organizations around the world, including PolitiFact in the U.S., who received a grant for video fact-checking on coronavirus.

COVID Scams Arizonans and Americans have been inundated with fraudulent offers and scams, using social media to spread inaccurate information and perpetrate criminal scams. I've been using my own social media to help warn Arizonans about common scams related to economic assistance, false coronavirus "cures", and where they can report scams to Federal and state authorities.

Question 4. What has Facebook done to limit the spread of scams and report criminal activity and what more can be done to protect seniors, veterans, and others who have been targeted by fraudsters?

Answer. Facebook is supporting the global public health community's work to keep people safe and informed during the COVID–19 public health crisis. We're also working to address the long-term impacts by supporting industries in need and making it easier for people to find and offer help in their communities. We've been prioritizing ensuring everyone has access to accurate information, removing harmful content, supporting health and economic relief efforts, and keeping people connected.

Under our Regulated Goods policy, we've also taken steps to protect against exploitation of this crisis for financial gain by banning content that attempts to sell or trade medical masks, hand sanitizer, surface-disinfecting wipes, and COVID–19 test kits. We also prohibit influencers from promoting these sales through branded content. From March through October 2020, we removed over 14 million pieces of content globally from Facebook and Instagram related to COVID–19 and which violated our medical supply sales standards. Of these, over 370,000 were removed in the U.S.. In addition, between March and October of 2020, we removed more than 13 million pieces of content globally on Facebook and Instagram for containing misinformation that may lead to imminent physical harm, such as content relating to fake preventative measures or exaggerated cures. Of these, over 3 million were removed in the U.S.

In removing content that has the potential to contribute to real-world harm, we are also focusing on our policies related to commerce listings. We prohibit people from making health or medical claims related to COVID–19 in product listings on commerce surfaces, including those listings that guarantee a product will prevent someone from contracting COVID–19. We also prohibit the buying or selling of drugs and prescription products. When someone creates a listing on Marketplace, before it goes live, it is reviewed against our Commerce Policies using automated tools, and in some cases, further manual review. When we detect that a listing violates our policies, we reject it. We also have a dedicated channel for local governments to share listings they believe violate local laws.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. JACKY ROSEN TO MARK ZUCKERBERG

Question 1. Adversaries like Russia continue to amplify propaganda—on everything from the election to the coronavirus to anti-Semitic conspiracy theories—and they do it on your platform, weaponizing division and hate to destroy our democracy and our communities. The U.S. intelligence community warned us earlier this year that Russia is now actively inciting white supremacist violence, which the FBI and Department of Homeland Security say poses the most lethal threat to America. In recent years, we have seen white supremacy and anti-Semitism on the rise, much of it spreading online. What enables these bad actors to disseminate their hateful messaging to the American public are the algorithms on your platforms, effectively rewarding efforts by foreign powers to exploit divisions in our country?

Question 1a. Are you seeing foreign manipulation or amplification of white supremacist and anti-Semitic content, and if so, how are your algorithms stopping

this? Are your algorithms dynamic and nimble enough to combat even better and more personalized targeting that can be harder to identify?

Answer. Terrorists, terrorist content, and hate speech in all forms—including white supremacy and domestically based violent extremist content—have no place on Facebook. We prohibit content that incites violence, and we remove terrorists and posts that support terrorism whenever we become aware of them. We use a variety of tools in this fight against terrorism and violent extremism, including artificial intelligence, specialized human review, industry cooperation, and counterspeech training. Our definition of terrorism is agnostic to the ideology or political goals of a group, which means it includes everything from religious extremists and violent separatists to white supremacists and militant environmental groups. It is about whether they use violence or attempt to use violence to pursue those goals.

Anti-Semitism is abhorrent and also has no place on our platform. Facebook removes any post that celebrates, defends, or attempts to justify the Holocaust. The same goes for any content that mocks Holocaust victims, accuses victims of lying about the atrocities, or advocates for violence against Jewish people in any way. We also updated our hate speech policy earlier this year to prohibit any content that denies or distorts the Holocaust.

If we find instances of coordinated inauthentic behavior conducted on behalf of a foreign actor, we apply the broadest enforcement measures, including the removal of every on-platform property connected to the operation itself and the people and organizations behind it.

We have also invested significantly in combating inauthentic behavior, whether it takes the form of individual fake accounts or broader coordinated networks. Over the past several years, our team has grown to over 200 people with expertise ranging from open-source research to threat investigations, cybersecurity, law enforcement and national security, investigative journalism, engineering, product development, data science, and academic studies in disinformation.

Question 1b. Have you increased or modified your efforts to quell Russian disinformation in the wake of recently revealed efforts by Russia and Iran weaponize stolen voter data to exploit divisions in our nation? How have you or will you adjust your algorithms to reduce the influence of such content—knowing that these countries' newly obtained data will allow for even better targeting, making their deception harder to identify?

Answer. When we find instances of coordinated inauthentic behavior conducted on behalf of a government entity or by a foreign actor, in which the use of fake accounts is central to the operation, we apply the broadest enforcement measures, including the removal of every on-platform property connected to the operation itself and the people and organizations behind it. We regularly share our findings about the networks we find and remove for coordinated inauthentic behavior.

Our teams continue to focus on finding and removing deceptive campaigns around the world, whether they are foreign or domestic. In October, we removed 14 networks of accounts, Pages, and Groups. Eight of them—from Georgia, Myanmar, Ukraine, and Azerbaijan—targeted domestic audiences in their own countries, and six networks—from Iran, Egypt, the U.S., and Mexico—focused on people outside of their countries. And this past March, we removed a network of 49 Facebook accounts, 69 Pages, and 85 Instagram accounts linked to activity we had previously removed and attributed to the Russian Internet Research Agency (IRA). We have shared information about our findings with law enforcement, policymakers, and industry partners. And we publish regular reports on the coordinated inauthentic behavior we detect and remove from our platforms. Our October 2020 report can be found at *https://about.fb.com/news/2020/11/october-2020-cib-report/*.

We are making progress rooting out this abuse, but as we've said before, it's an ongoing effort.

We're committed to continually improving to stay ahead. That means building better technology, hiring more people, and working closely with law enforcement, security experts, and other companies.

Question 1c. Are you consulting outside groups to validate moderator guidelines on hate speech, including what constitutes anti-Semitic content? Are you collecting data on hate speech content? If so, what are you doing with that data to combat hate speech on your platforms?

Answer. In developing and iterating on our policies, including our policy specific to hate speech, we consult with outside academics and experts from across the political spectrum and around the world.

We define hate speech as a direct attack on people based on what we call protected characteristics—race, ethnicity, national origin, religious affiliation, sexual orientation, sex, gender, gender identity, and serious disability or disease. We also

provide some protections for immigration status. We define an attack as violent or dehumanizing speech, statements of inferiority, and calls for exclusion or segregation. You can see more about these policies here: *https://www.facebook.com/communitystandards/objectionable_content/hate_speech.*

To track our progress and demonstrate our continued commitment to making Facebook safe and inclusive, we regularly release our Community Standards Enforcement Report (available at *https://transparency.facebook.com/community-standards-enforcement*). This report shares metrics on how Facebook is performing in removing content that violates our Community Standards. We release a "prevalence" metric that estimates how much violating content in particular categories has been posted on the platform. We have recently added this prevalence metric for hate speech content. We also share data on our process for appealing and restoring content to correct mistakes in our enforcement decisions.

Question 2. Recently, there have been high profile cybersecurity breaches involving private companies, government agencies, and even school districts—including in my home state of Nevada. A few months ago, a hacker subjected Clark County School District—Nevada's largest school district and our country's fifth largest, serving more than 320,000 students—to a ransomware attack. In the tech industry, there was a notable breach of Twitter in July, when hackers were able to access an internal IT administrator tool used to manage accounts. Dozens of verified accounts with high follower counts—including those of President Obama, Bill Gates, and Jeff Bezos—were used to send out a tweet promoting a Bitcoin scam. What we learned from this breach is stunning . . . the perpetrators were inside the Twitter network in one form or another.

Question 2a. How often do your staff attend cybersecurity training? Do you hire outside cybersecurity firms to look at your systems, offering a fresh look and catching overlooked flaws?

Answer. Protecting the security of information on Facebook is at the core of how we operate. Security is built into every Facebook product, and we have dedicated teams focused on each aspect of data security. From encryption protocols for data privacy to machine learning for threat detection, Facebook's network is protected by a combination of advanced automated systems and teams with expertise across a wide range of security fields. Our security protections are regularly evaluated and tested by our own internal security experts and tools. We supplement this in some cases with independent contracted security evaluations, and more broadly with external security experts through our industry-leading Facebook Bug Bounty program, described in more depth below.

Protecting a global community of billions of users involves a wide range of teams and functions, and our expectation is that those teams will continue to grow across the board. For example, we have information security, threat intelligence, and related engineering teams that are dedicated to traditional cybersecurity, including protecting people's accounts and information. We are continuing to expand these teams, along with other groups at Facebook working on security. Since 2011, we have also run an industry-leading and widely recognized bug bounty program where we encourage security researchers to responsibly disclose potential issues so we can fix the bugs. Our bug bounty program has been instrumental in helping us quickly detect new bugs, spot trends, and engage the best security talent outside of Facebook to help us keep the platform safe. Over the last several years, we have continued to innovate in this area by expanding the bug bounty program to include an industry-first data abuse bounty program, where researchers can report misuse of Facebook data, even where it may be happening off of our platform. As an additional check, we also have a so-called "red team" of internal security experts who plan and execute staged "attacks" on our systems. We then take the red team's findings and use them to build out protections to further strengthen our systems' security.

With respect to training, new and existing Facebook employees are required to complete a computer-based training focusing on confidentiality and security. Topics covered include Facebook's key privacy principles, Facebook's policies, privacy laws and regulations, vendor security audits, privacy and security by design, the importance of ensuring user data is kept secure from unauthorized access, and general security awareness leading practices. The learning and development team performs weekly monitoring to ensure employees receive and take their required trainings.

Facebook's Security Team conducts an annual, month-long security awareness campaign called "Hack-tober." The month includes hacks, where the Security Team targets Facebook employees and where the employees target the Security Team, security scavenger hunts looking for bugs in code, presentations from internal and external speakers, and an internal security "capture the flag." Facebook also encour-

ages Security Team members to attend security conferences hosted outside the Company to increase awareness of environmental, regulatory, and technological changes that may impact system security and confidentiality.

Question 3. The COVID–19 pandemic has shined a light on our Nation's digital divide and on the technological inequalities facing millions of American students, including those in Nevada. Lack of access to broadband disproportionately affects low-income communities, rural populations, and tribal nations—all of which are present in my state. In addition to broadband access, many students still do not have regular access to a computer or other connected device, making online learning incredibly difficult, and sometimes impossible.

Facebook stepped up during the pandemic to help close the digital divide, including by offering many educational resources to help teachers and parents during the pandemic.

Question 3a. As classes continue to meet online, or in a hybrid model, what more can Facebook do to help students and teachers?

Question 3b. How does Facebook plan to remain engaged in K–12 education after we get through the pandemic? In particular, what role can you play in closing not only the urban/rural divide, but also the racial divide in access to technologies and the Internet?

Answer. The COVID–19 pandemic has underscored the importance of Internet connectivity. While many people have shifted their lives online, there are still more than 3.5 billion people, including more than 18 million Americans, who lack reliable Internet access. To help, we have partnered with the Information Technology Disaster Resource Center (ITDRC) and NetHope to provide Internet connectivity to communities most impacted by COVID–19. The goal of these partnerships is to better understand the unique barriers these communities face in getting online and to create the programs and infrastructure needed to increase the availability and affordability of high-quality Internet access.

- We're providing a $2 million grant to support ITDRC's projectConnect initiative, which will help rural and underserved communities in the U.S. gain access to the internet. We're also sharing insights from Facebook Disease Prevention Maps to help ITDRC better understand options for Internet coverage in specific regions and more quickly determine the type of support needed to address connectivity challenges.
- We're providing a $260,000 grant to support NetHope's COVID–19 response. In addition, through sharing our Disease Prevention Maps, we'll help NetHope identify the world's most vulnerable and affected communities, including migrants and refugees, in order to provide them with protective health equipment and Internet connectivity kits.

Question 4. One of my top priorities in Congress is supporting the STEM workforce and breaking down barriers to entering and succeeding in STEM fields. This includes ensuring we have a diverse STEM workforce that includes people of color and women. In the past several years, tech companies have begun releasing diversity reports and promising to do better at hiring Black and Latino workers, including women. In overall employment, Facebook is doing much better today in building a diverse workforce. However, in 2020, just 1.7 percent of Facebook's tech employees were Black, and only 4.3 percent were Latino, up slightly from 2019, but not substantially higher than six years ago in 2014, despite the fact that the Latino population in the U.S. has surged during that time, including in Nevada.

I know that tech companies in Nevada understand that by increasing the number of women and people of color in tech careers, we diversify the qualified labor pool that the U.S. relies on for innovation. This will help us maintain our global competitiveness and expand our economy, and I hope your companies redouble your efforts to this effect.

Question 4a. Can you discuss the full set of 2020 data on women and the people of color who work at your companies, and would you please discuss what you are doing to increase these numbers in 2021?

Answer. Diversity is extremely important to Facebook, and we recognize that we still have work to do. We value diversity because we understand that it leads to better decisions, better products, and better culture. It is also more reflective of our community on Facebook.

Over the last seven years, Facebook has worked hard to make good on our commitment to diversity and inclusion. Our company has grown a lot. So has our approach. We are more focused than ever on creating a diverse workforce and supporting our people. They are the ones building better products and serving the communities on our platforms.

Today, there are more people of diverse backgrounds and experiences, more people of color, more women in both technical and business roles, and more underrepresented people in leadership at Facebook. Most notably, we have achieved higher representation of women in leadership by focusing on hiring and growing female leaders within the company. Over the last several years, the majority of new female leaders were internally promoted. And importantly, even as we have grown, we have worked very hard on making Facebook a more welcoming, respectful workplace.

Every year, Facebook publishes diversity data in a diversity report. Since 2014, when our strategic efforts began, we've made progress increasing the number of people from traditionally underrepresented groups employed at Facebook, but we recognize that we need to do more. In 2020, 37 percent of our workforce were women, up from 31 percent in 2014, and over 34 percent of our leadership are also women, up from 23 percent in 2014. We've almost doubled the percentage of Black employees—from 2 percent in 2014 to almost 4 percent in 2020, and we've increased the percentage of Hispanic employees from 4 percent in 2014 to over 6 percent in 2020. For more information, see *https://diversity.fb.com/read-report/*.

Looking forward, we are dedicated to prioritizing diverse hiring and are committed to our goal of having a company where, in the next five years, at least 50 percent of our workforce is comprised of women, people of color, and other underrepresented groups, and to increase people of color in leadership to 30 percent—including a 30 percent increase in Black leaders. When it comes to hiring, we have a diverse slate approach modeled after the Rooney Rule. This ensures that recruiters present qualified candidates from underrepresented groups to hiring managers looking to fill open roles, and it sets the expectation that hiring managers will consider candidates from underrepresented backgrounds when interviewing for an open position. We've seen steady increases in hiring rates for underrepresented people since we started testing this approach in 2015. We're also focused on increasing the diversity and inclusion capabilities of managers and leaders to build inclusive teams, departments, and organizations so that our products and community will benefit from the diverse perspectives of our people. We know that we still have a lot of work to do. We aren't where we need to be on diversity, but we are committed to improving, and we will work hard to get to where we know we need to be.

Question 4b. What are you doing more broadly to support STEM education programs and initiatives for women and people of color, including young girls of color?

Answer. In order to ensure that the next generation of tech innovators better reflects who we all are, it is critical that children from underrepresented communities be exposed to technology and computer science at the pre-secondary education level and remain engaged in those fields through high school and beyond.

To that end, in 2012, we launched the Facebook Academy initiative, a six-week summer internship program for local teens near our headquarters in Menlo Park, California. Through that program, we have enrolled 100 high school students from our local communities.

In 2015, we launched TechPrep, a resource hub created specifically for learners from underrepresented groups and their parents and guardians. It not only exposes students to computer science, but also introduces them to hundreds of different resources that fit their needs, based on age and skill level. TechPrep is available in both English and Spanish and enables students and their supporters to find local classes, workshops, and learning programs just by entering a zip code.

We have created CodeFWD by Facebook, a free online education program that helps educators inspire underrepresented and female 4th-to 8th-grade students to pursue computer programming. Teachers who participate in the program are eligible to receive a free coding robot and a classroom kit to further the learning process. We have participants from 43 states, including the Harlem Children's Zone, the Chicago Youth Center, Boys & Girls Clubs, and Latinitas, a charter school in Texas.

In 2016, we announced a $15 million commitment to Code.org. This commitment has helped Code.org drive the development of curricula, public school teacher training, and student skills-building, particularly among traditionally underrepresented populations in engineering and computer science.

Beyond the specific programming described above, we are continually investing in opportunities to bring computer science and STEM programming to middle-and high-school students. At the college and university level, we know that if we're going to hire people from a broader range of backgrounds, it's not enough to simply show up for recruiting events. We need to create practical training opportunities for these students to build on their academic experiences.

Facebook University, our longest-running program in this area, is an eight-week paid internship program that enables students from underrepresented communities to get to know Facebook's people, products, and services by working across engineering, analytics, product design, operations, and global marketing solutions roles.

Facebook University has graduated hundreds of students since its inception more than six years ago.

We are also investing in partnerships with organizations that contribute to developing the long-term pool of talent such as Girls Who Code, Year Up, Ron Brown Scholars, T Howard Foundation, Posse Foundation, MLT, The Consortium, and Jopwell.

We recently signed a partnership with CodePath.org, a non-profit whose goal is to "eliminate educational inequity in technical education starting with college computer science (CS) education." This partnership will help CodePath reach 2,000 more computer science students at over 20 universities to increase students' preparation for the rigor of tech interviews at companies across the U.S. These include community colleges, HSIs, HBCUs, and other institutions that have traditionally attracted students of color.

We have announced a new pilot program to bring Oculus Go units and virtual reality training to a number of HBCUs across the country, starting with Florida A&M. This will put technology and storytelling capabilities into the hands of students who will work alongside a team of professionals to create virtual campus tours for prospective students, for some of whom the cost of making a pre-enrollment visit is prohibitively expensive. This will not only help recruiting efforts but will also expose students at HBCUs to emerging technology.

We have partnered with the UNCF to design courses for their HBCU CS Summer Academy. We will also continue to co-host the HBCU CS Faculty Institute, in partnership with UNCF's Career Pathways Initiative, as we have done since 2016. This program offers faculty important professional development opportunities.

In our Boston, New York, and Washington, D.C. offices, we have created Above and Beyond Computer Science, a volunteer-led program of Facebook engineers that helps prepare local college students for the technical interview process by reviewing computer science concepts and applied problem solving. Seventy percent of the students who have participated identify as from a population underrepresented in tech. Our focus is now on expanding the size of this initiative, including creating a remote, web-based pilot program.

As part of our Engineer in Residence Program, Facebook software engineers teach indemand computer science coursework at historically Black-and Hispanic-serving institutions such as Morgan State University, Cal State Monterey Bay, and the New Jersey Institute of Technology, whose student populations are highly diverse. In addition to designing and teaching undergraduate computer science coursework customized for each university's unique context, Facebook Engineers in Residence also fulfill the responsibilities of an adjunct faculty member: hosting office hours, grading, managing teaching assistants, facilitating mock interviews, and providing networking and mentoring opportunities for students.

For three years running, Facebook has also been the title sponsor of the ASBC HBCU College Festival, the Nation's largest such festival, organized by the Alfred Street Baptist Church and the ASBC Foundation. During the 2018 festival alone, 2,117 instant offers for admission to HBCUs were made, and $4.8 million in scholarships were awarded.

Question 5. To continue being the most innovative country in the world, we need to maintain a workforce that can innovate. By 2026, the Department of Labor projects there will be 3.5 million computing-related jobs, yet our current education pipeline will only fill 19 percent of those openings. While other countries have prioritized STEM education as a national security issue, collaborating with non-profits and industry, the United States has mostly pursued an approach that does not meaningfully include such partnerships. The results of such a strategy are clear. A recent study found that less than half of K–12 students are getting any cyber related education, despite a growing demand for cyber professionals, both in national security fields and in the private sector.

Question 5a. What role can Facebook play in helping the United States boost its competitiveness in STEM fields, so that our economy can better compete with others around the globe?

Answer. Please see the response to your Question 4(b).

FACEBOOK'S CIVIL RIGHTS AUDIT—FINAL REPORT

JULY 8, 2020

TABLE OF CONTENTS

About
Acknowledgements
Introduction by Laura W. Murphy
Chapter One: Civil Rights Accountability Structure
Chapter Two: Elections & Census 2020
Chapter Three: Content Moderation & Enforcement
Chapter Four: Diversity & Inclusion
Chapter Five: Advertising Practices
Chapter Six: Algorithmic Bias
Chapter Seven: Privacy

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 U.S. 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 cu-

mulative, 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 U.S. 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 percent, including 30 percent 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 U.S. 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 nonprofits 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 U.S. 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 ex-

pert, 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 U.S. 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.

CHAPTER ONE: CIVIL RIGHTS ACCOUNTABILITY STRUCTURE

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 Auditors recognize 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 U.S. 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 U.S. 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 U.S. 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 U.S. 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 U.S. Elections and Census, supervised by a single manager in the U.S.

(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 out-

fitting 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 U.S. 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 U.S. 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 oppor-

tunities 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 U.S. (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 U.S. Census Bureau and other nonprofits, 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 percent 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 U.S. 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 U.S. presi-

dential 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 in-

accurate 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, e-mails, 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 U.S. 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 percent 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 U.S. 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 U.S. 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 short-hand 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 percent 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 percent 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 U.S. 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 the Board). 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 circumstances (such as describing violations 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."
- 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 U.S. 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 en-

forcement 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 re-

moval 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 percent) and Asian (43 percent), with less than 13 percent 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 in-

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

crease 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 percent 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 percent more people of color, including 30 percent 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 U.S. 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.

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;

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

- 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 percent and 5.2 percent of employees across all positions, respectively, 1.5 percent and 3.5 percent of employees in technical roles, 8.2 percent and 8.8 percent of employees in business and sales roles, and 3.1 percent and 3.5 percent of employees in senior leadership roles. While Asian employees represent 43 percent of the workforce (and 52 percent of employees in technical roles), they represent only 24.9 percent 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 percent to 8 percent.

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

In 2020, Facebook will report that 8 percent of its U.S. workforce self-identified as LGBTQA+ (based on a 54 percent response rate), noting a 1 percent 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 U.S. has remained relatively steady at 2 percent between 2018 and 2020. As for people with disabilities, Facebook will report that 3.9 percent of its U.S. 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 percent 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 percent 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 percent, including increasing the representation of Black employees in such roles by 30 percent. 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 percent 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

(i) 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.)

(ii) 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 diver-

sity 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 percent 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.

CHAPTER FIVE: ADVERTISING PRACTICES

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 U.S. 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 U.S. 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 U.S. 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.

CHAPTER SIX: ALGORITHMIC BIAS

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 re-

peat 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.

CHAPTER SEVEN: PRIVACY

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 U.S. 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 e-mail 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 percent of the data requests Facebook receives, and Facebook provides at least some requested data in response to such emergency requests approximately 74 percent 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 rec-

ognition 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.

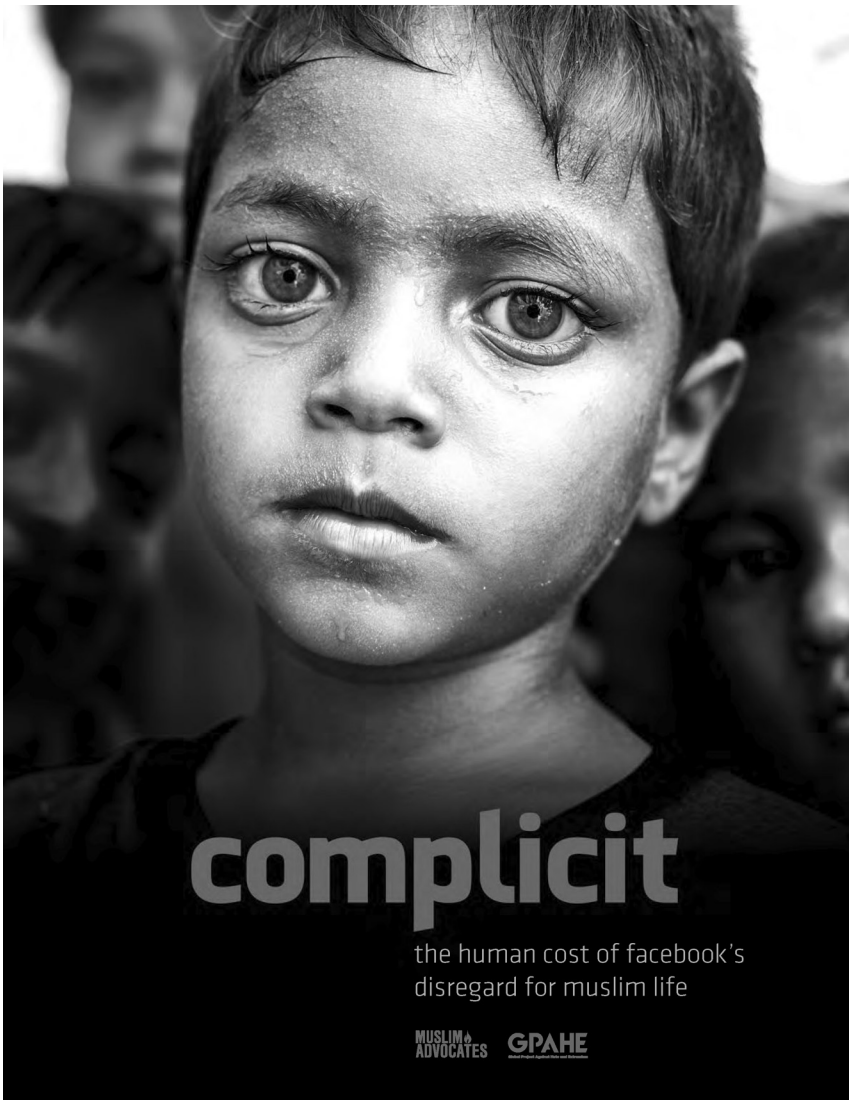
Facebook is also taking steps to support manual contact tracing efforts—that is, efforts which promote awareness and understanding of off-platform tracing initiatives that do not involve the sharing of Facebook data. For example, through its COVID Information Center and advertising, Facebook is helping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


complicit
the human cost of facebook's
disregard for muslim life
MUSLIM ADVOCATES
GPAHE

Published October 21, 2020

A joint publication of Muslim Advocates and the Global Project Against Hate and Extremism (GPAHE).

Muslim Advocates is a national civil rights organization working in the courts, in the halls of power and in communities to halt bigotry in its tracks. We ensure that American Muslims have a seat at the table with expert representation so that all Americans may live free from hate and discrimination. More information on Muslim Advocates is available at www.muslimadvocates.org.

The Global Project Against Hate and Extremism (GPAHE) uses research, advocacy, and education to counter the rising tide of extremism across borders by: documenting and exposing how extremist and hateful movements use social media and the internet to organize, propagandize, and recruit followers and sympathizers; educating anti-hate actors, policymakers, and the public about extremist movements, their leadership, and their activities; and examining how extremist movements influence each other across borders, specifically analyzing the impact of American extremists on other countries. More information on GPAHE is available at www.globalextremism.org.

© All rights reserved.

table of contents

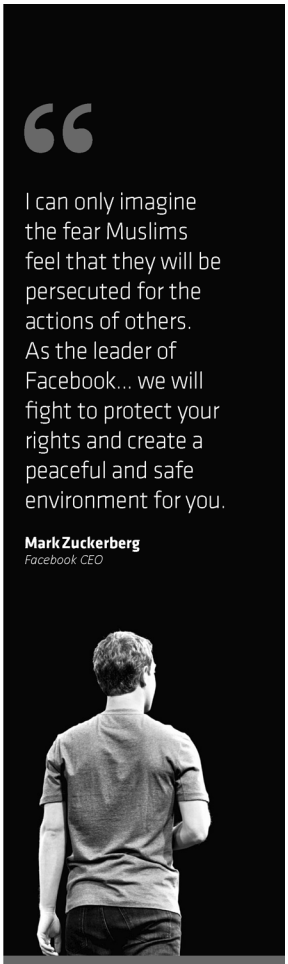
1 **Executive Summary**
5 **Recommendations**
6 **Introduction**
13 **Country Summaries**
14 China
16 Germany
18 Hungary
19 India
25 Myanmar
29 New Zealand
31 Sri Lanka
33 Sweden
34 United States
38 **Conclusion**
39 **Appendix**

Executive Summary

In 2015, Facebook CEO Mark Zuckerberg vowed to make his social media platform a welcoming home for Muslims after members of the community faced a brutal backlash in the wake of horrific, violent attacks in Paris. One 17-year-old in France reported, "There was a flood of violent language on Facebook to kill Muslims."[1] Zuckerberg posted on Facebook, "After the Paris attacks and hate this week, I can only imagine the fear Muslims feel that they will be persecuted for the actions of others. As the leader of Facebook... we will fight to protect your rights and create a peaceful and safe environment for you."[2] Citing the lessons of his Jewish background, Zuckerberg wrote that he wanted to add his voice in support of Muslims "in our community and around the world."

It's hard now to imagine more hollow words being spoken. Since Zuckerberg made that commitment, Facebook has been used to orchestrate the Rohingya genocide in Myanmar,[3] mass murders of Muslims in India,[4] and riots and murders in Sri Lanka that targeted Muslims for death.[5] Anti-Muslim hate groups and hate speech run rampant on Facebook with anti-Muslim posts, ads, private groups, and other content.[6] Armed, anti-Muslim protests in the U.S. have been coordinated from Facebook event pages.[7] The Christchurch, New Zealand, mosque massacres were live-streamed on the site and the videos shared by untold numbers worldwide.[8]

1 Adam Nossiter and Liz Alderman, "After Paris Attacks, a Darker Mood Toward Islam Emerges in France," *The New York Times*, Nov. 16, 2015.
2 Mark Zuckerberg, "I want to add my voice in support of Muslims in our community and around the world," Facebook, Dec. 9, 2015.
3 Paul Mozur, "A Genocide Incited on Facebook, With Posts from Myanmar's Military," *The New York Times*, Oct. 15, 2018.
4 Naveed Siddiqui, "3 killed, 100 others arrested in India after violence erupts over anti-Islam Facebook post," *Dawn*, Aug. 13, 2020.
5 Tasmin Nazeer, "Facebook's Apology for its Role in Sri Lanka's Anti-Muslim Riots Should Spark Change," *The Diplomat*, May 15, 2020.
6 Mehdi Hasan, "Dear Mark Zuckerberg: Facebook Is an Engine of Anti-Muslim Hate the World Over. Don't You Care?" *The Intercept*, Dec. 7, 2019.
7 Claire Allbright, "A Russian Facebook page organized a protest in Texas. A different Russian page launched the counterprotest," *The Texas Tribune*, Nov. 1, 2017.
8 Emanuel Stoakes, "New Zealand mosque attack victims confront gunman in courtroom," *The Washington Post*, Aug. 24, 2020; Andrew Liptak, "Facebook Says It Removed 1.5 million Videos of the New Zealand Mass Shooting," *The Verge*, March 17, 2019.



Facebook CEO Mark Zuckerberg is in charge of a platform that has become the global engine of anti-Muslim bigotry.
Photo: Alessio Jacona



Hate groups like Secure America Now use Facebook to demonize Muslims as invaders who are undermining democratic values, institutions, even monuments. Facebook Video

This report details Facebook's damaging impact on Muslim communities in nine countries and demonstrates how Facebook has willfully ignored the dangers posed by anti-Muslim content to the welfare of Muslims across the globe. Facebook's own civil rights audit, completed in 2020[9] after a nearly two-year process that came only after sustained pressure from human rights groups, singled out anti-Muslim hate on the platform as a longstanding problem.[10] The auditors wrote, "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," created an atmosphere where "Muslims feel under siege on Facebook."

The company falsely equated the protection of their users' safety, civil rights, and human rights with **satisfying partisan political interests**.

Despite this damning finding, Facebook has yet to address the anti-Muslim problems found by its own audit or those identified by advocates around the world in the preceding years. The company's actions, or lack thereof, indicate that its decision to engage in the audit was not necessarily out of concern for lives lost because of its inaction, but rather for political expediency. Incredibly, the same day that the civil rights audit was announced in May 2018, the company also announced that one of the United States Senate's most notorious anti-Muslim and anti-immigrant members, former Senator Jon Kyl, would conduct his own audit into anti-conservative bias on the platform. Kyl has a history of demonizing Muslims and working with anti-Muslim hate groups.[11] This is illustrative of Facebook's penchant to fail to simply do what's right and what they've promised; instead, the company falsely equated the protection of their users' safety, civil rights, and human rights with satisfying partisan political interests. Investigating human and civil rights abuses and global anti-Muslim hate content that leads to the loss of life should not have been presented as needing to be balanced with an investigation of alleged anti-conservative bias.

9 Facebook, *Facebook Civil Rights Audit–Final Report,* July 8, 2020.
10 Muslim Advocates, "Civil Rights Groups Urge Facebook to Address Longstanding Issues with Hate Speech and Bigotry," Press Release, Oct. 31, 2017.
11 Bridge Initiative Team, Georgetown University, "Factsheet: Jon Kyl," Sept 27, 2018; Brett Barrouquere, "To Replace John McCain Arizona Governor Picks Politician with a History of Anti-Muslim Rhetoric, Jon Kyl," Southern Poverty Law Center's Hatewatch, Sept. 4, 2018.

For years, civil rights organizations and policymakers have raised concerns both privately and publicly, begging the company to take action. It has not. Out of desperation to be heard, many organized boycotts,[12] called for the board and company leadership to be replaced,[13] and asked governments to step in.[14] Still, Facebook's strategy was to wait and do almost nothing.

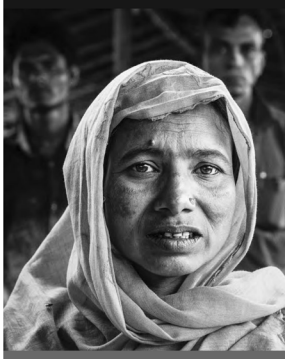
Facebook's actions are not dissimilar to IBM's support of Germany's Nazi Party, where it allowed its technology to be used by Hitler's regime to enact its genocidal policies in exchange for profits.[15] Facebook has done the same in places like India and Myanmar, enabling mass violence by anti-Muslim regimes and individual actors in an attempt to grow or monopolize Internet markets.

Outside of the U.S., where vulnerable communities have fewer options for recourse, the picture is even bleaker. In a most horrifying case, Facebook was cited by the U.N. for playing a "determining" role in the genocide perpetrated against the Muslim Rohingya community in Myanmar.[16] And in India, the Delhi State Assembly's Peace and Harmony Committee just recently found that Facebook was complicit in the Delhi riots of February 2020, and should be investigated for every riot since 2014.[17] Evidence shows that Facebook has at times seemingly collaborated with anti-Muslim regimes, such as the current ruling party in India, to protect hate speech by its leadership in contravention of its own anti-hate policies.[18]

Facebook's leadership has said repeatedly that the company's policies against hate apply to everyone regardless of who they are or where they are, and yet the company continues to allow anti-Muslim material to stay on the platform,[19] using a variety of excuses

12 Queenie Wong, "Facebook Ad Boycott: Why Brands 'Hit Pause on Hate,'" *CNET*, July 30, 2020.
13 Muslim Advocates, "32 Civil Rights Groups Lose Faith in Facebook, Call for Significant Reforms to its Board," Press Release, Dec. 17, 2018.
14 Stephanie Bodoni and Natalia Drozdiak, "EU Says It Can Force Facebook to Pull Posts With Hate Speech Anywhere in the World," *Fortune*, Oct. 3, 2019.
15 Edwin Black, *IBM and the Holocaust: The Strategic Alliance between Nazi Germany and America's Most Powerful Corporation* (New York: Crown Books, 2001).
16 Eli Meixler, "U.N. Fact Finders Say Facebook Played a 'Determining' Role in Violence Against the Rohingya," *Time*, March 13, 2018.
17 Aditya Chunduru, "Key Takeaways from Peace & Harmony Committee Meeting," *Medianama*, Sept. 1, 2020.
18 *Al Jazeera*, "Facebook ignored hate speech by India's BJP politicians: Report," *Al Jazeera*, Aug. 15, 2020.
19 Deepa Seetharaman, "Facebook Employees Pushed to Remove Trump's Posts as Hate Speech," *The Wall Street Journal*, Oct. 21, 2016.

For years, civil rights organizations and policymakers have raised concerns both privately and publicly, **begging the company to take action. It has not.**



Fatema, 35, who is seven months pregnant, fled the genocide against Rohingya Muslims in Myanmar that was fueled by online hate. Photo: Anna Dubuis/DFID

MUSLIM ADVOCATES GPAHE

including "newsworthiness."[20] For example, Facebook has said that exceptions to its hate speech policy are sometimes made if content is "newsworthy, significant or important to the public interest."[21]

At the same time, the company contends that hate speech does not help its bottom line.[22] If so, why has Facebook repeatedly chosen to leave up dangerous hate content, often generated by public figures with large audiences, such as President Donald Trump?[23] Another case in point is India, where Facebook now has its largest user base[24] and is investing in expansion and growth to dominate that very large market.[25] The company is working closely with the current Indian government,[26] in particular Prime Minister Narendra Modi, who was himself allegedly involved with stoking anti-Muslim violence when he was governor of the Indian state of Gujarat that led to the deaths of 1,000 Muslims.[27] There appears to be a clear financial incentive to pander to Modi, who has one of the largest number of followers of any political leader on Facebook,[28] and his political party. It is, of course, hate speech by such figures that is the most dangerous because of their reach and influence.[29]

Facebook is seeding and cultivating anti-Muslim bigotry amongst its users, leading to real world violence against the 270 million Muslims living in the nine countries covered in this report.[30] Facebook is indisputably the world's engine for anti-Muslim violence. The time for discussion of this issue is over. Facebook must act now and end anti-Muslim hate on its platform, no matter who or what entity, is proliferating it.

20 David Gilbert, "Facebook is Giving a Free Pass on Hate Speech," *Vice News*, Sept. 25, 2019.
21 Kate Cox, "Facebook Confirms Its 'Standards' Don't Apply to Politicians," *Ars Technica*, Sept. 25, 2019.
22 Nick Clegg, "Facebook Does Not Benefit from Hate," Facebook, July 1, 2020.
23 Ian Bremmer, "Zuckerberg won't censor Trump's anti-Muslim 'hate post' on Facebook: Report," *Hindustan Times*, Oct. 22, 2016.
24 Jeff Joseph Paul Kadicheenie, "How India Became Facebook's Biggest Market," *The Hoot*, Aug. 16, 2017.
25 Ravi Agrawal, "Why Facebook Is Betting Big on India," *Foreign Policy*, Apr. 23, 2020.
26 Billy Perrigo, "Facebook's Ties to India's Ruling Party Complicate Its Fight Against Hate Speech," *Time*, Aug. 27, 2020.
27 Sanjoy Majumder, "Narendra Modi 'allowed' Gujarat 2002 anti-Muslim riots," *BBC News*, Apr. 22, 2011.
28 Surabhi Agarwal, "With 41.7 million followers, Modi becomes most followed world leader on Facebook," *The Economic Times*, May 27, 2017.
29 Dangerous Speech Project, "Dangerous Speech: A Practical Guide," Aug. 4, 2020.
30 This is a rough estimate based on Facebook penetration in a country and the share of the population in that country that is Muslim. There is no data available on how many Muslim users of Facebook exist in the countries where the company operates, so these numbers are tentative. Population numbers derived primarily from World Bank and UN data.

The company continues to **allow anti-Muslim material** to stay on the platform, using a variety of excuses including "newsworthiness".[31]



Syrian refugees arrive in Germany amidst growing anti-Muslim and anti-refugee sentiment that is amplified through social media.
Foto: Jazzmany

Recommendations

To bring an end to anti-Muslim hate on Facebook, we recommend the company immediately:

- Rigorously, and without regard to political or economic implications, enforce Facebook's community standards and Dangerous Individuals and Organizations designations worldwide to address anti-Muslim hate.
- Ban the use of event pages for the purpose of harassment, organizing, and violence targeting the Muslim community.
- Create a senior staff working group responsible for the reduction of hate speech on the platform, and require regular updates through Facebook's Transparency Report on the company's progress in removing offending content, including anti-Muslim hate content.

Introduction

Facebook CEO Mark Zuckerberg's 2015 promise to make his platform a welcoming home for Muslims has proven to be an illusion. In country after country, the Facebook platform has been used as a driver for anti-Muslim violence,[31] the result of the company repeatedly and willfully ignoring the dangers of anti-Muslim content. None of this has happened without Facebook's knowledge[32] or without public outcry from its users, civil and human rights groups,[33] and lawmakers.[34] Facebook is simply looking the other way while its platform becomes the global engine of anti-Muslim bigotry.

Many people may not be aware that Facebook's actions have led to incredible levels of anti-Muslim violence in multiple countries, and that even when warned that material posted to the platform was likely to end in violence, the company repeatedly chose not to act. Unfortunately, decisions to ignore its own community standards that should protect Muslim communities have been influenced instead by anti-Muslim ruling governments,[35] militias,[36] and political influencers, the very people who can reach and influence the largest audiences.

In 2020, dehumanizing content about Muslims remains widespread on Facebook. A recent analysis showed that the United States and Australia "lead in the number of active Facebook pages and groups dedicated to referring this dehumanizing [anti-Muslim] content."[37]



Refugees are forced to leave Hungary amidst growing anti-Muslim hate that proliferates online. Photo: Bence Jardany

31 Mehdi Hasan, "Dear Mark Zuckerberg."
32 Ariana Tobin, "Civil Rights Groups Have Been Warning Facebook About Hate Speech In Secret Groups For Years," *ProPublica*, July 2, 2019.
33 Rishi Iyengar and Donnie O'Sullivan, "Civil rights groups invited to Zuckerberg's home slam Facebook's 'lackluster response,'" *CNN*, Dec. 4, 2019.
34 Taylor Hatmaker, "With Advertiser Boycott Growing, Lawmakers Press Facebook on White Supremacy," *Techcrunch*, June 30, 2020.
35 Hannah Ellis-Petersen and Shaikh Azizur Rahman, "Facebook Faces Grilling by MPs in India Over Anti-Muslim Hate Speech," *The Guardian*, Sept. 1, 2020.
36 Tyler Sonnemaker, "Facebook 'did nothing' about violent militia and hate groups for 5 years despite being warned at least 10 times, advocacy group says," *Business Insider*, Sept. 18, 2020.
37 Rita Jabri-Markwell, "The online dehumanisation of Muslims made the Christchurch massacre possible," ABC, Aug. 31, 2020.

This content is significantly driven by the white supremacist hate speech and hate groups that thrive on Facebook and routinely demonize Muslims. In May 2020, the Tech Transparency Project (TTP) found more than 100 U.S. white supremacist groups were active on Facebook, on their own group pages as well as on auto-generated content.[38] Among the many vulnerable communities attacked by these groups, Muslims were demonized by circulating footage and making comments about the Christchurch mosque shootings. In the wake of TTP's report, Facebook did alter some of the auto-generated content and groups, but many remained on the platform.[39]

Anti-Muslim hate on Facebook is well-documented

The problem with anti-Muslim hate on Facebook has been documented for some years. Two years ago, computer scientist Megan Squire did an analysis of right-wing extremist groups on Facebook and found an incredible overlap with anti-Muslim hate.[40] Squire found that anti-Muslim attitudes are not only flourishing on the platform but also acting as a "common denominator" for a range of other extremist ideologies, including anti-immigrant groups, pro-Confederate groups, militant anti-government conspiracy theorists, and white nationalists. Among thousands of Facebook users who were members of multiple extremist Facebook groups, Squire found that 61 percent of "multi-issue" users who were in anti-immigrant groups had also joined anti-Muslim groups; the same was true for 44 percent of anti-government groups, 37 percent of white nationalist groups, and 35 percent of neo-Confederate groups. "Anti-Muslim groups are way worse, in every way, than what I would have guessed coming in," Squire said at the time, "Some of the anti-Muslim groups are central players in the hate network as a whole. And the anti-Muslim groups show more membership crossover with other ideologies than I expected."[41]



Protestors take to the street after President Trump implements the Muslim ban. Facebook refused to remove his posts about the ban even though they violated the company policies. Photo: Stephen Melkisethian

38 Tech Transparency Project, "White Supremacist Groups Are Thriving on Facebook," May 21, 2020.
39 Tech Transparency Project, "White Supremacist Groups."
40 Ishmael N. Daro, "Here's How Anti-Muslim Groups On Facebook Overlap With A Range Of Far-Right Extremism," *BuzzFeed*, Aug. 4, 2018.
41 Ishmael N. Daro, "Here's How."

Civil rights organizations have repeatedly warned Facebook that anti-Muslim posts, ads, private groups, and other content are rampant on a global scale.[42] As early as 2015, Muslim Advocates informed Facebook that its event pages were being used to organize activities by anti-Muslim militias and hate groups, including armed anti-Muslim protests in the U.S.[43] The Southern Poverty Law Center also reached out to Facebook privately starting in 2014 to warn the company about hate groups on its platform, including dozens of anti-Muslim hate groups.[44]

Civil rights audit

In perhaps the only significant step made by Facebook at the behest of civil rights organizations, in 2018 the company agreed to an independent civil rights audit, which was finalized in July 2020.[45] Though primarily focused on domestic civil rights issues, Facebook's auditors singled out anti-Muslim hate on the platform as a longstanding problem. They wrote, "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" created an atmosphere where "Muslims feel under siege on Facebook."[46] Facebook has yet to address any of the longstanding anti-Muslim problems on its platform which have been raised for years by civil and human rights groups and confirmed and validated by their own audit. The company's actions indicate that its decision to engage in the audit at all was not out of concern for lives lost because of its inaction, but for political expediency.

> "The organization of events designed to intimidate members of the Muslim community created an atmosphere where **Muslims feel under siege on Facebook**.
>
> **Facebook's Civil Rights Audit**
> *Final Report*

42 Jason Schwartz, "Civil rights groups slam Facebook," *Politico*, Oct. 31, 2017.
43 Tyler Sonnemaker, "Facebook 'did nothing.'"
44 Information from Heidi Beirich, who was director of the Southern Poverty Law Center's Intelligence Project from 2010-2019 and engaged in several discussions with Facebook staff over hate groups on the platform.
45 John Fingas, "Facebook will conduct civil rights audit following bias allegations," *Engadget*, May 2, 2018.
46 Facebook, *Facebook Civil Rights Audit*.



In 2019, an Australian white supremacist massacred 51 Muslims in Christchurch, New Zealand, while streaming it on Facebook Live. Photo: Guillaume Garin

The problem is international in scope

Outside of the U.S., the picture is even bleaker. Tragically, there are many other cases where the hatred and poison against Muslims originating on Facebook has turned lethal, leading to riots, deaths, and mass killings. Evidence shows that Facebook has at times seemingly collaborated with anti-Muslim regimes, such as the current ruling party of India, to protect hate speech by its leadership in contravention of its own anti-hate policies.[47]

Facebook was cited by the U.N. for playing a "determining" role in the genocide perpetrated against the Muslim Rohingya community in Myanmar that began in 2016.[48] In September 2018, Facebook COO Sheryl Sandberg admitted in a Senate committee hearing that the company had failed the Rohingya,[49] calling the situation "devastating"[50] and even admitting that Facebook may have a legal obligation to remove accounts that lead to mass violence. Yet, anti-Muslim material remained rife on the platform. One year later, in March 2019, the Christchurch, New Zealand, mosque massacres were broadcast on Facebook Live and shared countless times worldwide. In September 2020, the Delhi State Assembly's Peace and Harmony Committee found that Facebook was complicit in the Delhi riots of February 2020, and should be investigated for every riot since 2014.[51]

47 Billy Perrigo, "Facebook's Ties."
48 Eli Meixler, "U.N. Fact Finders."
49 *BBC Trending*, "The country where Facebook posts whipped up hate," *BBC*, Sept. 12, 2018.
50 Evelyn Douek, "Facebook's Role in the Genocide in Myanmar: New Reporting Complicates the Narrative," *Lawfare*, Oct. 22, 2018.
51 Aditya Chunduru, "Key Takeaways."

In the wake of recent disclosures that Facebook allowed anti-Muslim hate speech to run rampant in India, its own employees have spoken out about this long-term and repeated failure to deal with anti-Muslim hate content. In an August 2020 open letter addressed to the company's leadership, 11 employees demanded that the platform denounce "anti-Muslim bigotry" and ensure Facebook's policies are applied equally across the platform.[52] "It is hard not to feel frustrated and saddened by the incidents reported. We know we're not alone in this. Employees across the company are expressing similar sentiment," the letter read. "The Muslim community at Facebook would like to hear from Facebook leadership on our asks."



The pain from the Christchurch massacre reverberated around the world, leading to memorials like this in the United States. Photo: Charles Edward Miller

Facebook grows the white supremacist and anti-Muslim movement

By ignoring this issue, Facebook for years allowed the most dangerous white supremacist propaganda to fester and grow, recruiting untold numbers of people into the ranks of a movement that has inspired violence and genocide targeting Muslims, immigrants, and Jews. The Christchurch shooter had been radicalized into Identitarianism,[53] spread on multiple Facebook accounts,[54] and was an adherent of the racist Great Replacement theory - an international white supremacist conspiracy movement that promotes the idea that white people are slowly experiencing a genocide in their own home countries due to a plot by elites to displace white people with rising numbers of non-white immigrants.[55] The Great Replacement theory argues that immigrants, especially Muslims, are destroying Western countries and turning them into foreign places. These thinkers often argue disparagingly that Europe is becoming Eurabia.[56]

52 Regina Mihindukulasuriya, "Denounce 'anti-Muslim bigotry' — Facebook staff in US, other countries write to company," *The Print*, Aug. 20, 2020.
53 Hope Not Hate, "What is Identitarianism?" Undated.
54 Mark Townsend, "Infiltrator exposes Generation Identity UK's march towards extreme far right," *The Guardian*, Aug. 24, 2019.
55 Lauretta Charlton, "What is the Great Replacement?" *The New York Times*, Aug. 6, 2019.
56 Rita Jabri-Markwell, "The Online Dehumanization."



Extremist groups like the Proud Boys used Facebook to organize rallies and spread anti-Muslim hate. Photo: Robert P. Alvarez

The New Zealand shooter was clear about his motive. In his manifesto, Brenton Tarrant wrote that he wanted to stop the Great Replacement, and he targeted Muslims, including their children, for extermination.[57] As Peter Lentini has written, "Tarrant's solution to the crisis [posed by a Muslim "invasion"] – indeed one on which he felt compelled to enact – was to annihilate his enemies (read Muslim immigrants). This included targeting non-combatants. In one point, he indicates that [immigrants] constitute a much greater threat to the future of Western societies than terrorists and combatants. Thus, he argues that it is also necessary to kill children to ensure that the enemy line will not continue."[58]

Generation Identity (GI), the sprawling, international organization that pushes the Great Replacement idea, was rampant on Facebook for years.[59] In June 2018, Facebook finally took action against GI, but only after several members of the Austrian chapter were investigated for potentially running a criminal organization (the investigation ended without charges). Facebook deplatformed the entire network citing violations of policy.[60] But this was too little, too late. Before being blocked, more than 120,000 people followed Generation Identity on Facebook.[61] Facebook is complicit in the worldwide transmission of Identitarian thinking, which is anti-Muslim at its heart.

Since October 2018, there have been at least six mass attacks motivated by Great Replacement ideology.[62] In addition to Christchurch, attacks were staged at two American synagogues, an El Paso Walmart, a synagogue in Halle, Germany, and two shisha bars in Hanau, Germany. Two of these mass shootings specifically targeted Muslims, in Hanau and Christchurch, but all were directed against immigrants or Jews who were seen as abetting non-white immigration. Anders Breivik, who in 2012 murdered dozens of children in Norway because he believed they would grow up to become adult supporters of Muslim immigrants to the country, actively promoted his ideas on Facebook before the attack.[63]

57 Andreas Önnerfors, "'The Great Replacement' - Decoding the Christchurch Terrorist Manifesto," Centre for the Analysis of the Radical Right, March 18, 2019.
58 Rita Jabri-Markwell, "The Online Dehumanization."
59 Heidi Beirich and Wendy Via, *International White Nationalist Movement Spreading on Twitter and YouTube*, Global Project Against Hate and Extremism, July 2020.
60 *Al Jazeera*, "Facebook pulls the plug on far-right Generation Identity," *Al Jazeera*, June 22, 2018.
61 Alexander Durie, "At the Heart of Hate: Inside France's Identitarian Movement," *Are We Europe*, 2018.
62 Heidi Beirich and Wendy Via, *International White Nationalist*.
63 Camila Ragfors, "Norway attacks: Breivik was my friend on Facebook. I've seen what fed his hatred," *The Guardian*, July 24, 2011.

Muslim users of the platform

By refusing to take anti-Muslim material on its platform seriously, Facebook is in effect poisoning its Muslim users, and its users in general. The nine countries covered in this report have a total of 566 million Facebook users, which make up about 20 percent of their customers worldwide.[64] Furthermore, the top three markets for Facebook users are India, the United States, and Indonesia, which is a Muslim-majority country.[65] Muslims in India are an astonishing 200 million[66] and in Indonesia, the country with the largest Muslim population in the world, are 225 million.[67] Of Facebook's top 20 markets by country, six are Muslim-majority countries: Indonesia, Egypt, Bangladesh, Pakistan, Turkey, and Nigeria.[68] Two of Facebook's top 20 markets—India and Myanmar—have been the site of massacres of Muslims orchestrated on Facebook.

Facebook's ongoing refusal to enforce its policies and protect Muslims is occurring as Facebook usage rapidly grows in Muslim-majority markets. In 2019, according to data compiled by the University of Oregon,[69] more than seven out of 10 citizens in six Muslim-majority countries use Facebook and WhatsApp: Egypt, Jordan, Lebanon, Qatar, Saudi Arabia (KSA), Tunisia, and the United Arab Emirates. And the use of these networks far outpaced the use of other social channels. As one example, Facebook had 187 million active users in the Middle East (comprised of mostly Muslim-majority countries) in 2019. The University of Oregon's data also showed that half of young people say they get their daily news on Facebook instead of newspapers, TV, or even online news portals.

All of these people, Muslim and non-Muslim, are being assaulted by the anti-Muslim, bigoted content that Facebook refuses to address. Facebook must act now and end anti-Muslim hate on the platform, no matter who, or what entity, is proliferating the hate. The only reason to allow dangerous hate content from anyone to stay up, and especially from political figures with reach and influence, is because it's profitable.[70]



A girl wails as the Modi government strips Kashmiri Muslims of their rights and Facebook takes away their ability to use WhatsApp to communicate. Photo: Firdous Qadri

64 This is a rough estimate based on Facebook penetration in a country and the share of the population in that country that is Muslim. There is no data available on how many Muslim users of Facebook exist in the countries where the company operates.
65 Statista, "Leading countries based on Facebook audience size as of July 2020."
66 Kunal Purohit, "The Islamophobic roots of population control efforts in India," *Al Jazeera*, Aug. 9, 2019.
67 Masrur Jamaluddin, Sandi Sidhu and Helen Regan, "Indonesia has the world's biggest Muslim population. It just banned holiday travel over Ramadan," *CNN*, Apr. 24, 2020.
68 Statista, "Leading countries."
69 Damian Radcliffe and Hadil Abuhmaid, *Social Media in the Middle East: 2019 In Review*, The University of Oregon School of Journalism and Communication, Jan. 2020.
70 Ernesto Verdeja and Bettina Spencer, "The Short Fuse: Autocrats, Hate Speech and Political Violence," *Just Security*, July 22, 2020.

Country Summaries: Anti-Muslim Organizing and Violence on Facebook

Unfortunately, there is little public awareness regarding Facebook's role as a chief driver of anti-Muslim hate and violence throughout the world. Even more disturbing, the company chose repeatedly over years not to act in the face of overwhelming evidence that material posted to Facebook was likely to end in violence. The following summary describes Facebook's role in the deaths of thousands of Muslims in multiple countries. These examples not only illustrate Facebook's complicity but also demonstrate how profoundly hate content can impact people's lives and cause violence offline.

What's documented here is illustrative of the problem with anti-Muslim hate on the platform, but likely represents just the tip of the iceberg, and is by no means exhaustive, as information about Facebook's role in dozens of countries and/or anti-Muslim incidents remains patchy, or unexamined, or nonexistent. Given the widespread nature of anti-Muslim hate on the platform, it is very likely that millions of Muslims around the world have been impacted negatively, jeopardizing their safety and their freedom, by Facebook's failure to act on anti-Muslim hate speech.

Facebook, and all of us, cannot forget that, every day, Muslims across the world are the targets of bigotry and hate crimes while simply going about their lives. And that anti-Muslim rhetoric and organizing that thrives online inspires and enables this abuse. These country reports are tragic, cautionary tales of how easily everyday hate speech turns into violence and murder. And how easily Facebook's platform can be used to inspire and organize mass violence and genocide.



China

Facebook does not even operate in China, but the Chinese government still uses the platform to amplify anti-Muslim and specifically anti-Uighur sentiment.[71] More than a million Muslim Uighurs have been imprisoned and brutalized by the Chinese government in concentration camps across the Xinjiang province.[72] The Uighurs and other Muslim ethnic minorities have been separated from their families, forced to labor for their captors, beaten, tortured, and raped.[73]

In 2019, there were reports that "Chinese state-owned media is running ads on Facebook seemingly designed to cast doubt on human rights violations" against the Uighurs.[74] There were three ads — two active and one inactive — within Facebook's ad library describing alleged successes of happy detainees in the camps and falsely claiming that the detention centers do not interfere with religious beliefs and practices. Two of the ads were targeted to American and other countries' audiences. The paid ads aimed to convince Westerners that the camps in Xinjiang are not sites of human rights abuses, contrary to the findings of several governments, human rights organizations, experts on China, the U.N., and other international bodies.[75]



More than a million Muslim Uighurs have been detained and brutalized by the Chinese government in concentration camps across the Xinjiang province. Photo: Azamat Imanaliev

71 Alexandra Ma, "Mysterious automated calls, vanished relatives, and sinister Facebook comments: How China intimidates Uighurs who don't even live in the country," *Business Insider*, Dec. 9, 2019.
72 Jen Kirby, "Concentration camps and forced labor: China's repression of the Uighurs, explained," *Vox*, Sept. 25, 2020.
73 Ellen Ioanes, "Rape, medical experiments, and forced abortions: One woman describes horrors of Xinjiang concentration camps," *Business Insider*, Oct. 22, 2019.
74 Ryan Mac, "These New Facebook Ads From Chinese State Media Want You To Believe Xinjiang's Muslim Internment Camps Are Just Great," *BuzzFeed*, Aug. 20, 2019.
75 Sigal Samuel, "China paid Facebook and Twitter to help spread anti-Muslim propaganda," *Vox*, Aug. 22, 2019.



The Guardian published drone footage of what is believed to be showing police leading hundreds of blindfolded and shackled Muslim Uighurs during a transfer of inmates in Xinjiang. Photo: Screenshot from The Guardian

On Facebook, the state-controlled tabloid *Global Times* posted a sponsored video titled, "Xinjiang center trainees graduate with hope for future."[76] It purports to show former detainees baking bread, as an example of the "vocational skills" Uighurs supposedly learn in the camps. Additional ads were posted to Twitter.

After the ads were reported to Twitter, the company removed them immediately.[77] Given the same information, Facebook, in stark contrast, decided to keep accepting such ads and said they would take a "close look at ads that have been raised to us to determine if they violate our policies."[78] Instead of simply refusing paid ads from Chinese state-controlled media, Facebook chose to passively rely on outside experts to flag problematic posts, which it may or may not then remove, at a pace that may or may not be quick enough to avert harm. In effect, Facebook was enabling China to use the platform to cover up widespread human rights abuses and violence in Xinjiang.

Facebook was enabling China to use the platform to **cover up widespread human rights abuses and violence** in Xinjiang.

76 Ryan Mac, "These New Facebook Ads."
77 Ryan Gallagher, "Twitter Helped Chinese Government Promote Disinformation on Repression of Uighurs," *The Intercept*, Aug. 19, 2019.
78 Sigal Samuel, "China paid."

Germany

Sparked by online rumors that a man was killed defending a woman from rape by a Muslim refugee, riots targeting Muslim refugees and immigrants broke out in August of 2018, starting in the state of Saxony.[79] The Facebook account of the municipal political party,[80] Pro-Chemnitz,[81] pushed the misinformation and organized the protest that ended in mob violence. In calling for the protest, it claimed the victim in the rumored stabbing was "a brave helper who lost his life trying to protect a woman."[82]

Chancellor Angela Merkel suggested the violence was a threat to Germany's post-war constitution, saying, "We have video footage of the fact that there was [hunting people down], there were riots, there was hatred on the streets, and that has nothing to do with our constitutional state."[83]



A woman holds a sign that says "Islam does not belong in Germany." Photo: Foto Berlin

79 *Deutsche Welle*, "German state official: Fake news fueled Chemnitz riots," *Deutsche Welle*, Aug. 29, 2018.
80 *Deutsche Welle*, "German state official."
81 Katy Dartford and Ayman Oghanna, "Chemnitz: How the East German city became known as a hotbed for extremism," *Euronews*, March 23, 2020.
82 J. Lester Feder and Pascal Anselmi, "The Real Story Behind The Anti-Immigrant Riots Rocking Germany *BuzzFeed*, Aug. 31, 2018.
83 *Der Spiegel*, "Es darf auf keinem Platz und keiner Straße zu solchen Ausschreitungen kommen," *Der Spiegel*, Aug. 28, 2018.

In another incident, a teenaged Syrian refugee, Anas Modamani, took a selfie with Chancellor Angela Merkel that he posted on Facebook.[84] After terrorist attacks in Brussels and Berlin in 2016, Modamani's selfie began appearing on Facebook, this time doctored to falsely label him as one of the perpetrators of the attacks.[85] Afraid of being recognized, Modamani was afraid to leave his home. Chan-jo Jun, a lawyer, brought a landmark lawsuit against Facebook on behalf of Modamani for being smeared online.[86] His litigation was unsuccessful, but it helped lead[87] Germany to pass one of the world's most aggressive laws targeting online hate speech,[88] the so-called NetzDG law.[89]

But the damage was already done. The rumors helped support the growth of the anti-Muslim Alternative for Germany party, which would eventually be elected to seats in several German state parliaments.[90] The AfD now has over 500,000 followers on Facebook.[91] Research in 2018 by academics at the University of Warwick specifically showed that thousands of hate-filled Facebook posts were linked to an increase in racially-motivated attacks on refugees in Germany, who are predominantly Muslim.[92] The research specifically cited material from the AfD as fueling this awful trend. Now, anti-Muslim and anti-refugee sentiment is widespread among sectors of the German population. In 2019, there were more than 800 attacks on Muslims in Germany.[93]

Anti-Muslim and anti-refugee sentiment is widespread among sectors of the German population. In 2019, there were **more than 800 attacks on Muslims in Germany**.

84 Melissa Eddy, "How a Refugee's Selfie with Merkel Led to a Facebook Lawsuit," *The New York Times*, Feb. 6, 2017.
85 *BBC*, "Syrian migrant launches Facebook 'fake news' legal fight," *BBC*, Jan. 12, 2017.
86 Linda Kinstler, "Can Germany Fix Facebook?" *The Atlantic*, Nov. 2, 2017.
87 Patrick Evans, "Will Germany's new law kill free speech online?" *BBC*, Sept. 18, 2017.
88 *Deutsche Welle*, "Germany implements new internet hate speech crackdown," *Deutsche Welle*, Jan. 1, 2016.
89 Geoffrey Smith, "Germany's New Law Is a Milestone for Social Media Regulation in Europe," *Fortune*, June 30, 2017.
90 Rafaela Dancygier, "The anti-Muslim AfD just scored big in Germany's election. What does this mean for German Muslims?" *The Washington Post*, Sept. 25, 2017.
91 AfD Facebook page.
92 Samuel Stolton, "German Facebook use linked to refugee violence as AfD boost online presence," *Euractiv*, Aug. 23, 2018.
93 *Deutsche Welle*, "Over 800 attacks on Muslims reported in Germany in 2019," *Deutsche Welle*, March 28, 2020.

Hungary

Hungary has one of the most anti-Muslim governments in Europe. During the European refugee crisis, which began in 2015, Prime Minisiter Orbán and his government refused to allow refugees fleeing the Middle East into his country and specifically said in 2018 that the term "refugees" is a misnomer and that those coming to Europe were "Muslim invaders."[94] For Prime Minister Viktor Orbán, Muslims cannot be a part of Europe[95] and Muslims should be kept out of Europe to "keep Europe Christian."[96] Orbán's Facebook page,[97] where he pushes his views, has more than a million followers.

There is little information on Facebook usage to spread hate in Hungary, but one incident is instructive. In March 2018, Facebook reversed a decision to remove an anti-immigrant video targeting Muslims posted by János Lázár,[98] then chief of staff to Hungarian Prime Minister Viktor Orbán (Lazar was called a racist by the U.N. human rights chief in 2018[99]). The video featured Lazar saying, "If we let them in and they are going to live in our towns, the result will be crime, poverty, dirt, and impossible conditions in our cities." Lazar accused Facebook of censorship after the social network removed his post. Facebook, in reposting the racist video, said that it was making an exception to its ban on hate speech: "Exceptions are sometimes made if content is newsworthy, significant or important to the public interest," Facebook said.

The anti-refugee and anti-Muslim sentiment is so severe in Hungary that even those who work at organizations devoted to helping these populations have been giving up.[100] Hungary's Muslim population has faced beatings, vigilante attacks, and abuse.[101]

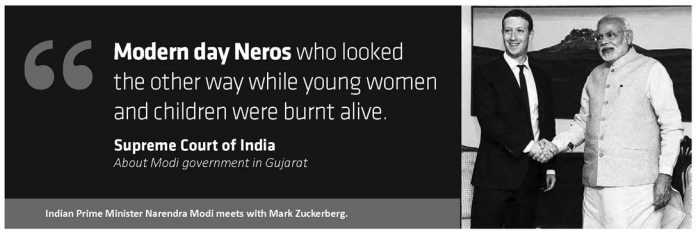
> If we let them in and they are going to live in our towns, the result will be crime, poverty, dirt, and impossible conditions in our cities.
>
> **János Lázár**
> *Former chief of staff to Hungarian Prime Minister Viktor Orbán*

94 Emily Schultheis, "Viktor Orbán: Hungary doesn't want 'Muslim invaders,'" *Politico*, Jan. 8, 2018.
95 HVG, "Orbán: "Az iszlám soha nem volt Európa része," HVG, Oct. 16, 2015.
96 Robert Mackey, "Hungarian Leader Rebuked for Saying Muslim Migrants Must Be Blocked 'to Keep Europe Christian,'" *The New York Times*, Sept. 3, 2015.
97 Viktor Orbán Facebook page.
98 *Reuters* Staff, "UPDATE 1-Facebook removes, then restores anti-immigrant video in Hungary," *Reuters*, March 7, 2018.
99 *VOA News*, "Facebook Removes, Then Restores Anti-Immigrant Video in Hungary," *VOA News*, March 7, 2018.
100 Joe Wallen, "'Hungary is the worst': Refugees become punch bag under PM Viktor Orbán," *The Independent*, July 18, 2018.
101 Paul Peachy, "Hungary's Muslims fear fallout from anti-Islam rhetoric," *The National*, March 10, 2018.

India

The situation is so dire regarding the anti-Muslim bias of Facebook India that the company's senior executives were summoned before a parliamentary committee for a closed door hearing on September 2, 2020. The committee hearing followed allegations that the company's top policy official in India, Ankhi Das,[102] had prevented the removal of hate speech and anti-Muslim posts by ruling party Bharatiya Janata (BJP) politicians[103] in order to protect and promote the Hindu nationalist party and its Prime Minister Narendra Modi, who have advanced anti-Muslim policies.[104]

Ties between the company and the BJP are deep. Both Zuckerberg[105] and Sandberg[106] have met personally with Modi, who is the most popular world leader on Facebook.[107] Before becoming prime minister, Zuckerberg even introduced his parents to Modi,[108] a strange choice considering Modi's horrific track record of stoking violence against Muslims. In February 2002, while head of the Gujarat government, Modi allegedly encouraged massive anti-Muslim riots.[109] As the state was overcome with violence and over a thousand Muslims were murdered, leaders of the BJP and its even more nationalist ally, the Vishwa Hindu Parishad, gave speeches provoking Hindus to teach Muslims a lesson.[110] Modi himself gave an incendiary speech, mocking riot victims and calling relief camps for Muslims "child-producing factories."[111] The intensity and brutality of the violence unleashed against Muslims in 2002 led the Supreme Court of India to describe the Modi government in Gujarat as, "Modern day Neros who looked the other way while young women and children were burnt alive."[112]



102 Sanskriti Tiwari, "Who is Ankhi Das and What's the Controversy Surrounding Her?" *SheThePeople.tv*, Aug. 31, 2020.
103 Newley Purnell and Rajesh Roy, "Facebook Faces Hate-Speech Questioning by Indian Lawmakers After Journal Article," *The Wall Street Journal*, Aug. 18, 2020.
104 Samanth Subramanian, "How Hindu Supremacists are Tearing India Apart," *The Guardian*, Feb. 2, 2020.
105 Mehdi Hasan, "Dear Mark Zuckerberg."
106 New Delhi Bureau, "Facebook COO Sheryl Sandberg meets PM Modi," *The Hindu Business Line*, July 3, 2014.
107 Saumya Tewari, "Narendra Modi the most popular world leader on Facebook: Study," *Live Mint*, April 23, 2020.
108 Mehdi Hasan, "Dear Mark Zuckerberg."
109 Sanjoy Majumder, "Narendra Modi 'allowed' Gujarat 2002 anti-Muslim riots," *BBC*, April 22, 2011.
110 Rana Ayyub, "Narendra Modi Looks the Other Way as New Delhi Burns," *Time*, Feb. 28, 2020.
111 Narendra Modi, "Should We Run Relief Camps? Open Child Producing Centres?,'" *Outlook Magazine*, Sept. 30, 2002.
112 Rana Ayyub, "Narendra Modi Looks."



Prime Minister Modi is celebrated by a Hindu nationalist party which promotes anti-Muslim hate online.

The problems go right to the top in Facebook's India operations. Das is herself implicated in spreading anti-Muslim bigotry.[113] She shared a post on her own Facebook page which referred to Muslims as a "degenerate community" for whom "nothing except purity of religion and implementation of Shariah matter," which Das said "spoke to me... as it should to [the] rest of India"[114] (After press reports that led to mass outrage, Das weakly apologized referencing her commitment to feminism, using an anti-Muslim trope regarding Muslims and sexism.[115]) The controversy was spotlighted in an August report that said Facebook India employees repeatedly attempted to remove content and label BJP leaders as dangerous individuals, and Das overruled their recommendations in order to protect her personal political interests.[116] One such post that Das overruled was by Anantkumar Hegde, a BJP member of parliament, who alleged that Muslims were spreading coronavirus in the country as part of an anti-Muslim conspiracy called "Corona Jihad."[117]

Das' abuse of her authority in support of her political allies has led to the loss of lives. In February 2020, former BJP lawmaker Kapil Mishra gave a speech threatening activists protesting the BJP's anti-Muslim citizenship law in Delhi, the country's capital, which makes citizenship difficult for Muslims.[118] This was a post flagged to Das but allowed to stay up.[119] Mishra said his supporters would use force to stop the protests.[120] Hours after the video was posted on Facebook, rioting broke out in the city that led to dozens of deaths,[121] most of them Muslims. Some of this violence was arranged on Facebook's subsidiary, WhatsApp, which has repeatedly been the source of mob violence targeting Muslims in India.[122]

113 *Al Jazeera*, "Facebook executive who shared anti-Muslim post apologises: Report," *Al Jazeera*, Aug. 27, 2020.
114 Al Jazeera, "Facebook executive."
115 Pranav Dixit, "A Facebook Executive Who Shared An Anti-Muslim Post Has Apologized To Employees," *BuzzFeed*, Aug. 24, 2020.
116 Jeff Horwitz and Newley Purnell, "Facebook Executive Supported India's Modi, Disparaged Opposition in Internal Messages," *The Wall Street Journal*, Aug. 30, 2020.
117 *Al Jazeera*, "Facebook ignored hate speech by India's BJP politicians: Report," *Al Jazeera*, Aug. 15, 2020.
118 Soumya Shankar, "India's Citizenship Law, In Tandem with National Registry, Could Make BJP's Discriminatory Targeting of Muslims Easier," *The Intercept*, Jan. 30, 2020.
119 Al Jazeera, "Facebook Ignored."
120 Newley Purnell and Jeff Horwitz, "Facebook's Hate-Speech Rules Collide With Indian Politics," *The Wall Street Journal*, Aug. 14, 2020.
121 Al Jazeera, "Facebook ignored."
122 OpIndia Staff, "Delhi Riots: WhatsApp chats acquired by OpIndia show how some anti-CAA protestors were aware that violence was planned, discussed 'de-escalation,'" *OpIndia*, Sept. 25, 2020.

Das' inability to carry out her responsibilities in an objective manner manifests itself in many different situations. When T. Raja Singh, another member of the BJP, called for the slaughter of Rohingya Muslim refugees, threatened to demolish mosques, and labeled Indian Muslim citizens as traitors, Facebook's online security staff determined his account should be banned for not only violating its community standards, but also for falling under the category of "Dangerous Individuals and Organizations."[123] Das stepped in to protect Singh from punitive action, because "punishing violations by politicians from Mr. Modi's party would damage the company's business prospects in the country," according to Facebook employees.[124] Outrage in response to these disclosures forced Facebook to finally ban Singh from the platform in early September 2020.[125] Lynchings spurred by bigoted and propagandistic WhatsApp posts led the company to consider banning mass messaging on the system.[126]

India is Facebook's largest and most lucrative market with nearly 350 million users and another 400 million on WhatsApp.[127] The BJP, which has more than 16 million followers on its page, is Facebook India's biggest advertising spender in recent months.[128] Facebook has multiple commercial ties with the Indian government, including partnerships with the Ministry of Tribal Affairs, the Ministry of Women, and the Board of Education.[129]

> "Punishing violations by politicians from Mr. Modi's party would damage the company's business prospects in the country.
>
> **Facebook employees**

123 C.J. Werleman, "How Facebook threatens vulnerable Muslim communities," *Middle East Eye*, Aug. 30, 2020.
124 Jeff Horwitz and Newley Purnell, "Facebook Staff Demand Policy Changes on India Hate Speech," *The Wall Street Journal*, Aug. 21, 2020.
125 *The New Arab*, "Facebook bans India ruling party politician over anti-Muslim hate speech," *The New Arab*, Sept. 3, 2020.
126 Shannon Liao, "WhatsApp tests limiting message forwarding after violent lynchings in India," *The Verge*, July 20, 2018.
127 Helen Ellis-Pederson, "Facebook Faces Grilling."
128 *Outlook* Web Bureau, "BJP again tops the political ad spend on Facebook In India," *Outlook*, Aug. 27, 2020.
129 Helen Ellis-Pederson, "Facebook Faces Grilling."

There are many more connections between anti-Muslim content on Facebook and violence in India. In May 2020, a BJP member of parliament in West Bengal, Arjun Singh, posted an image on Facebook that he wrongly claimed was a depiction of a Hindu who had been brutalized by Muslim mobs.[130] It was captioned: "How long will the blood of Hindus flow on in Bengal…we will not stay quiet if they [Muslims] attack ordinary people." Four hours later, an angry mob of about 100 Hindus descended on a town in West Bengal and a Muslim shrine was vandalized.[131] Facebook failed to remove the posts until after the company experienced backlash as a result of the violent attacks, which local Muslims alleged had been incited by Singh's post. Overall, dozens of Muslims have been lynched since 2012 by vigilantes, with many of the incidents triggered by fake news regarding cow slaughter or smuggling shared on WhatsApp.[132]

New Delhi riots

Facebook appeared to play a pivotal role in the February 2020 New Delhi riots in which more than 50 people died and thousands of homes and several mosques were destroyed. While both Hindus and Muslims were affected in the riots, Muslims were targeted in far greater numbers by mobs of young men, many of whom had traveled into the city to harass Muslims after seeing fake news shared widely on Facebook that Muslim religious leaders were calling for Hindus to be kicked out of Delhi.[133] One post by a BJP member who is also a member of the right-wing militant Hindu organization Bajrang Dal, prompted hundreds to comment that they and their Hindu "brothers" would join the fight to defend Delhi from the Muslims.[134] And two days before the anti-Muslim riots began in Delhi, a member of Modi's cabinet said Muslims should have been sent out of India to Pakistan in 1947 during the partition of India.[135] Ultimately, the Delhi State Assembly's Peace and Harmony Committee said it had prima facie found Facebook guilty of aggravating the Delhi riots, and posited that it should be investigated for every riot since 2014.[136]

During the riots, Facebook was also used by members of the mobs to glorify their violence. In early February, a Bajrang Dal activist posted a video claiming to have "killed a Mulle [derogatory term for Muslims]" and the next day wrote in a public Facebook post that he had just sent a "jihadi to heaven."[137] It took about three days for his Facebook account to be deactivated. In the wake of the violence, hundreds of Muslim families fled New Delhi.[138]

130 Helen Ellis-Pederson, "Facebook Faces Grilling."
131 Helen Ellis-Pederson, "Facebook Faces Grilling."
132 *Al Jazeera*, "Indian MPs Grill Facebook Over Hate Speech, Allegations of Bias," *Al Jazeera*, Sept. 3, 2020.
133 Human Rights Watch, "Shoot the Traitors" Discrimination Against Muslims under India's New Citizenship Policy," Human Rights Watch, April 9, 2020.
134 Helen Ellis-Pederson, "Facebook Faces Grilling."
135 Rana Ayyub, "Narendra Modi Looks."
136 Aditya Chunduru, "Key Takeaways."
137 Helen Ellis-Pederson, "Facebook Faces Grilling."
138 Rana Ayyub, "Narendra Modi Looks."

Human rights advocates cry out

Human rights organizations, including the Indian American Muslim Council, South Asians Building Accountability & Healing, and the Coalition to Stop Genocide in India (made up of dozens of organizations in the U.S. and other countries) assert that Facebook simply refuses to remove anti-Muslim hate content in India and have requested an investigation by the United States Congress.[139] In September of 2020, a letter signed by 41 civil rights organizations from around the world called on Facebook to put an end to anti-Muslim hate on its platform and immediately suspend Das, among other requests, to protect the safety and security of Muslims.[140]

Facebook's anti-Muslim actions in India have been repeatedly called out by civil society actors. In October 2019, a report by the nonprofit organization Avaaz accused Facebook[141] of having become a "megaphone for hate" against Muslims in the northeastern Indian state of Assam — where nearly two million people, many of them Muslims,[142] have been stripped of their citizenship by the BJP government.[143] Another report, by the South Asian human rights group Equality Labs, found "Islamophobic [anti-Muslim] content was the biggest source of hate speech on Facebook in India, accounting for 37 percent of the content," and that 93 percent of the hate speech they reported to Facebook was not removed.[144] They also reported on how Facebook is being used to spread hate speech and misinformation accusing Muslims of deliberately infecting non-Muslims and Hindus with COVID-19, again contributing to potential violence against Muslims.[145] Meanwhile, as the Modi government was stripping Muslims of their rights, Facebook was taking WhatsApp accounts away from Muslims in Kashmir.[146]

The government had suspended Internet in the region to prevent communication and Facebook's policy automatically discontinues WhatsApp participation after 120 days. As a result, the government prevented Muslims in the region from organizing, and Facebook contributed by further reducing communication opportunities.

> "Facebook became a "megaphone for hate" against Muslims in the northeastern Indian state of Assam.
>
> **Avaaz report**

139 *The Milli Gazette*, "Facebook scandal should be probed by US Congress say human rights organizations," *The Milli Gazette*, Aug. 18, 2020.
140 Global Project Against Hate and Extremism, et al, "Major Human Rights and Internet Watchdog Organizations Sign On to Demands for #AuditFBIndia," Sept. 9, 2020.
141 Roli Srivastava, "Facebook a 'megaphone for hate' against Indian minorities," *Reuters*, Oct. 30, 2019.
142 Sigal Samuel, "India's massive, scary new detention camps, explained," *Vox*, Sept. 17, 2019.
143 Makana Eyre, "Why India Just Stripped 1.9 Million People of Citizenship," *The Nation*, Sept. 10, 2019.
144 Equality Labs, *Facebook India*, 2019.
145 Equality Labs, *CoronaJihad*, 2020.
146 Pranav Dixit, "Kashmiris Are Disappearing From WhatsApp," *BuzzFeed*, Dec. 4, 2019.

In late August, a group of 54 former Indian bureaucrats wrote to Zuckerberg asking the company to perform an audit of how Facebook's hate speech policy is applied and to do so without Das' involvement.[147] The letter pointed to the financial aspects of Facebook's situation in India causing a conflict of interest. "That this (not censoring hate speech by members of the BJP) seems to have been done to protect Facebook's commercial interests is even more reprehensible... We note that such behavior on Facebook's part has become a subject of debate in other countries as well. Commercial interests at the cost of human lives? If these are the crass calculations Facebook indulges in, it is no surprise that the calculus of hate is spreading like a virus in many parts of the world," the letter read.

> Commercial interests at the cost of human lives? If these are the crass calculations Facebook indulges in, it is no surprise that the **calculus of hate is spreading like a virus in many parts of the world**.
>
> **Former Indian bureaucrats**
> *Letter to Facebook*

Many, however, question the utility of continuing to urge Facebook to address hate on the platform driven by the BJP and other Hindu nationalist organizations in India. Malay Tewari, a Kolkata-based activist, argued Facebook "rarely" responded to his complaints about BJP-linked posts and "quite strangely, Facebook posts which expose the propaganda or hate campaign of the BJP, which do not violate community standards, are often removed."[148] Indian journalist Rana Ayyub agreed saying, "For years now, verified Facebook pages of BJP leaders such as Kapil Mishra have routinely published hate speeches against Muslims and dissenting voices.[149] The hate then translates into deadly violence, such as the February anti-Muslim attacks in Delhi that left many people dead in some of the worst communal violence India's capital has seen in decades... It's clear that Facebook has no intention of holding hatemongers accountable and that the safety of users is not a priority."

In late August, it was reported that Facebook, in an effort to evaluate its role in spreading hate speech and incitements to violence, had commissioned an independent report by the U.S. law firm Foley Hoag LLP on the platform's impact on human rights in India.[150] Work on the India audit, previously unknown, began before political bias in Facebook's India operations was documented by journalists in August and September of 2020. Additionally, the new Facebook Oversight Board has indicated that it may step into the India situation. The board has said that it has the authority to decide "how Facebook treats posts from public figures that may violate community standards," including against hate speech and that it "won't shy away from the tough cases and holding Facebook accountable."[151] One can only hope that the Oversight Board will honor this commitment.

147 *Hindustan Times*, "54 ex-bureaucrats write to Facebook CEO, seek audit of hate-speech policy," *Hindustan Times*, Aug. 25, 2020.
148 Helen Ellis-Pederson, "Facebook Faces Grilling."
149 Rana Ayyub, "For Facebook, weakening India's democracy is just the cost of business," *The Washington Post*, Aug. 20, 2020.
150 Billy Perrigo, "Facebook's Ties."
151 Anumeha Chaturvedi, "Hate speech row: Facebook oversight board will investigate political 'bias' cases," *The Economic Times*, Aug. 20, 2020.

Myanmar

Beginning in 2012, activists, businessmen, and tech experts in Myanmar began warning Facebook that members of the military and ultra-nationalist Buddhists were directing hate speech and violence against the Rohingya,[152] long a targeted and vulnerable Muslim community in that country.[153] In a country of more than 53 million people, only four percent are Muslim Rohingya.[154]

By 2013, Facebook was ubiquitous in the country, serving basically as the Internet. The state-run newspaper said in 2013 that in Myanmar, "a person without a Facebook identity is like a person without a home address" (by 2020, nearly 90 percent of the population uses Facebook).[155]

Aela Callan, a foreign correspondent on a fellowship from Stanford University, met Facebook's then vice president of global communications to discuss hate speech and fake user pages that were pervasive in Myanmar in 2013.[156] Callan visited Facebook's California headquarters again in March 2014 with a staffer from a Myanmar tech organization to raise these issues with the company. Callan wanted to show Facebook "how serious it [hate speech and disinformation] was."[157] Her pleas were rebuffed. "It was seen as a connectivity opportunity rather than a big pressing problem," Callan said. "I think they were more excited about the connectivity opportunity because so many people were using it, rather than the core issues." Callan said hate speech seemed like a "low priority" at the time.

Vicious, false rumors that a Muslim Mandalay teashop owner raped a Buddhist employee started spreading across Facebook in 2014.[158] Soon, armed men were marauding through the streets of the capital on motorbikes and by foot wielding machetes and sticks. Rioters torched cars and ransacked shops. During the multi-day melee, two men—one Muslim and one Buddhist—were killed and 20 others were injured.[159] Similar violence would follow, spurred by hate-driven rumors on Facebook.

[For Facebook] hate speech seemed like a "low priority" at the time (2014).

Aela Callan
Foreign correspondent



The Rohingya, a targeted Muslim community in Myanmar, have been violently attacked following hate-driven rumors on Facebook. Photo: Evangelos Petratos

152 Billy Perrigo, "Facebook's Ties."
153 *BBC*, "Myanmar Rohingya: What you need to know about the crisis," *BBC*, Jan. 23, 2020.
154 Religious LIberty Project, "Myanmar," Harvard Divinity School, Undated.
155 Timothy McLaughlin, "How Facebook's Rise Fueled Chaos and Confusion in Myanmar," *Wired*, June 7, 2018.
156 TImothy McLaughlin, "How Facebook's Rise."
157 TImothy McLaughlin, "How Facebook's Rise."
158 TImothy McLaughlin, "How Facebook's Rise."
159 United Nations High Commissioner for Refugees, "2014 Country Reports on Human Rights Practices - Burma," *RefWorld*, 2015.



Rohingya Muslim refugees flee Myanmar after militants use Facebook to spread propaganda and perpetrate violence against them.
Photo: Sk Hasan Ali

That violence was serious, but the use of Facebook to coordinate the Rohingya genocide was yet to come. Facebook ignored repeated warnings that the military and Buddhist militants were using the platform to spread anti-Rohingya hate and propaganda. By the end of 2017, about 700,000 Rohingya had fled the country after Myanmar's military launched operations against what it called "insurgents" in the state of Rakhine.[160] Though the Rohingya have been in Myanmar for generations, the government denied citizenship to most Rohingya, arguing that they are illegal immigrants from neighboring Bangladesh.[161] A U.N. fact-finding mission in 2018 reported that, "People died from gunshot wounds, often due to indiscriminate shooting at fleeing villagers. Some were burned alive in their homes - often the elderly, disabled and young children. Others were hacked to death."[162]

Facebook does too little, much too late

It took until August 2018 – a year after 25,000 Rohingya were killed and 700,000 fled Myanmar– for Facebook to ban Tatmadaw (military) leaders from its platform[163] (only to return in June of 2020).[164] Facing growing public pressure, the company also commissioned and published an independent Human Rights Impact Assessment on the role its services were playing in the country[165] and committed to hiring 100 native Burmese speakers as content moderators.[166] Prior to deplatforming Tatmadow, Facebook banned four armed ethnic groups in the country, but not the Tatmadaw, presumably because they are state actors.[167] Repeated requests for comment from Facebook on why some ethnic groups were banned while a state actor engaged in ethnic violence was not, went unanswered.[168]

160 Council on Foreign Relations, "Rohingya Crisis in Myanmar," Global Conflict Tracker, Sept. 24, 2020.
161 Shoon Naing, "Myanmar forces Rohingya to accept cards that preclude citizenship: group," *Reuters*, Sept. 3, 2019.
162 United Nations Human Rights Council, "Fact-finding Mission on Myanmar: concrete and overwhelming information points to international crimes," Office of the High Commissioner for Human Rights, March 12, 2018.
163 Hannah Ellis-Pederson, "Facebook removes accounts associated with Myanmar military," *The Guardian*, Aug. 27, 2018.
164 Swe Lei Mon, "Tatmadaw Returns to Facebook After Two Year Absence," *Myanmar Times*, June 9, 2020.
165 Alex Warofka, "An Independent Assessment of the Human Rights Impact of Facebook in Myanmar," Facebook, Nov. 5, 2018.
166 Poppy McPherson, "Facebook says it was 'too slow' to fight hate speech in Myanmar," *Reuters*, Aug. 16, 2018.
167 Julia Carrie Wong, "Overreacting to failure': Facebook's new Myanmar strategy baffles local activists," *The Guardian*, Feb. 7, 2019.
168 United Nations Human Rights Council, "Fact-finding Mission on Myanmar."

This, even though the 2018 U.N. fact-finding report had noted that "actions of the Tatmadaw in both Kachin and Shan States since 2011 amount to war crimes and crimes against humanity."[169] Facebook has yet to impose across-the-board bans on military-run accounts of the kind applied to the four minority rebel groups. Instead, it seems to be taking a deeply statist approach toward these groups, thereby helping an army that stands accused of genocide.[170]

A U.N. fact-finding mission was shocked by Facebook's lack of action and found the company played a "determining role" in stirring up hatred against Rohingya Muslims in Myanmar.[171] The chairman of the U.N. mission, Marzuki Darusman, said that social media had "substantively contributed to the level of acrimony amongst the wider public against Rohingya Muslims… Hate speech is certainly, of course, a part of that."[172] Yanghee Lee, Special Rapporteur for human rights violations in Myanmar said, "I'm afraid that Facebook has now turned into a beast, and not what it originally intended."[173]

Facebook leadership has admitted its complicity in these events. Zuckerberg said in 2018 that Facebook needed to improve in Myanmar, though that admission came much too late for critics who said he failed to adequately take responsibility for what had been a long-term issue.[174] The company announced in July of that year that it would expand its efforts to remove material worldwide that could incite violence.[175] In a surprising concession before the U.S. Senate Intelligence Committee in September 2018, Sandberg called the events in Myanmar "devastating" and acknowledged the company had to do more, highlighting that Facebook had put increased resources behind being able to review content in Burmese.[176] She also accepted that Facebook had a moral and legal obligation to take down accounts that incentivize violence in countries like Myanmar (even so, two years later, Facebook continues to leave these kinds of accounts up, including in countries like India where violence against Muslims has been serious).

169 United Nations Office of the High Commissioner for Human Rights, "Myanmar: UN Fact-Finding Mission releases its full account of massive violations by military in Rakhine, Kachin and Shan States," United Nations Human Rights Office, Sept. 18, 2018.
170 Angshuman Choudhury, "How Facebook Is Complicit in Myanmar's Attacks on Minorities," *The Diplomat*, Aug. 25, 2020.
171 Alex Warofka, "An Independent Assessment."
172 *BBC*, "UN: Facebook has turned into a beast in Myanmar," *BBC*, March 13, 2018.
173 *BBC*, "UN: Facebook has turned."
174 Jen Kirby, "Mark Zuckerberg on Facebook's role in ethnic cleansing in Myanmar: 'It's a real issue,'" *Vox*, April 2, 2018.
175 Sheera Frankel, "Facebook to Remove Misinformation that Leads to Violence," *The New York Times*, July 18, 2018.
176 Evelyn Douek, "Senate Hearing on Social Media and Foreign Influence Operations: Progress, But There's A Long Way to Go," *Lawfare*, Sept. 6, 2020.

Failure to cooperate with the International Criminal Court

Given this catastrophic failure, one would think Facebook leadership would want to do absolutely everything in their power to support the current genocide case against Myanmar in the International Criminal Court (ICC). However, that hasn't been the case. In August 2020, Facebook balked at providing the ICC posts made by Myanmar's military and other leadership in a case of genocide against the regime being pursued by The Gambia and the Organization of Islamic Countries. The head of a U.N. investigative body on Myanmar said Facebook had not shared evidence of "serious international crimes," despite vowing to work with investigators looking into abuses in the country.[177] Nicholas Koumjian, head of the Independent Investigative Mechanism for Myanmar (IIMM), stated the social media giant was holding material "highly relevant and probative of serious international crimes" but had not shared any during year-long talks.[178] After an uproar in the press in August 2020, Facebook finally turned over some documents to the court that "partially complied" with the request.[179] By not cooperating with the Gambian legal team and creating a roadblock in the ongoing trial, Facebook is not being a "force for good in Myanmar" as it has repeatedly promised. It is failing to aid an important international effort to establish accountability in the country.[180]

In anticipation of the upcoming November Myanmar elections, Facebook released information on the small changes it has made, committing to remove "verifiable misinformation and unverifiable rumors" that are assessed as having the potential to suppress the vote or damage the "integrity" of the electoral process between September and November 22, 2020.[181] It also introduced a new feature that limits to five the number of times a message can be forwarded on WhatsApp.[182] Additionally, Facebook claims to now have three fact-checking partners in Myanmar and is working with two partners to verify the official Facebook pages of political parties, all of which will purportedly help detect hate speech that could lead to violence.[183]

177 *Reuters*, "UN investigator says Facebook hasn't shared 'evidence' of Myanmar crimes," *The New York Post*, Aug. 11, 2020.
178 Poppy McPherson, "U.N. investigator says Facebook has not shared 'evidence' of Myanmar crime," *Reuters*, Aug. 11, 2020.
179 Aarti Bali, "Facebook shares data on Myanmar with United Nations investigators," *WeForNews*, Aug. 26, 2020.
180 Matthew Smith, "Facebook Wanted to Be a Force for Good in Myanmar. Now It Is Rejecting a Request to Help With a Genocide Investigation," *Time*, Aug. 18, 2020.
181 *Al Jazeera*, "Facebook says stepping up efforts to tackle Myanmar hate speech," *Al Jazeera*, Sept. 1, 2020.
182 Manish Singh, "WhatsApp Introduces New Limit on Message Forwards to Fight Spread of Misinformation," *Techcrunch*, April 7, 2020.
183 Natasha Lomas, "Facebook Touts Beefed Up Hate Speech Detection Ahead of Myanmar Election," *Techcrunch*, Sept. 1, 2020.

New Zealand

The last words of Haji-Daoud Nabi before he and his 50 fellow worshipers were gunned down in Christchurch, New Zealand was the welcoming phrase, "Hello, brother."[184] On March 15, 2019, in a performance for Facebook Live, Brenton Tarrant slaughtered worshipers at two different mosques in one of the deadliest white supremacist attacks in recent history. The 17-minute long Facebook live-stream broadcast went viral; in the 24 hours following the attack, attempts were made to re-upload the footage 1.5 million times while Facebook scrambled to stop its spread.[185] The video of the attack was cross-posted across various social networks, and links to the live-stream and Tarrant's manifesto were posted on the unregulated message board 8chan.[186]

The New Zealand shooter was clear about his anti-Muslim motive. In his manifesto, Tarrant, who was in touch with members of the anti-Muslim, white supremacist and transnational Generation Identity movement, wrote that he wanted to stop the Great Replacement, and he targeted Muslims, including their children, for extermination. As Peter Lentini has written, "Tarrant's solution to the crisis [posed by a Muslim "invasion"] – indeed one on which he felt compelled to enact – was to annihilate his enemies (read Muslim immigrants). This included targeting non-combatants. In one point, he indicates that [immigrants] constitute a much greater threat to the future of Western societies than terrorists and combatants. Thus, he argues that it is also necessary to kill children to ensure that the enemy line will not continue."[187]



Over 50 Muslims of all ages, including young children, were massacred in Christchurch by a white nationalist who subscribed to the racist Great Replacement theory that has flourished on Facebook. Photo: Felton Davis

184 Mary Catherine Ford, "For Muslims, the mosque is the center of family life — and now a reminder of brutal massacre," *Chicago Tribune*, March 19, 2019.
185 Jon Porter, "Facebook says the Christchurch attack live stream was viewed by fewer than 200 people," *The Verge*, March 19, 2019.
186 Craig Timberg, Drew Harwell, Hamza Shaban, Andrew Ba Tran and Brian Fung, "The New Zealand Shooting shows how YouTube and Facebook spread hate and violent images–yet again," *The Washington Post*, March 15, 2019."
187 Rita Jabri-Markwell, "The Online Dehumanisation."

Facebook's non-response

The abuse of Facebook Live was foreseeable. Facebook had already been criticized for the use of Facebook Live in broadcasting suicides and violent attacks, and yet the company took little action.[188] In 2017, Facebook came under harsh criticism after a raft of suicides were live-streamed, including one of a 12-year-old girl in Georgia that was left live on the site for two weeks before being removed.[189] The company responded then that it was adding more resources to monitoring, but offered the excuse that the sheer volume of content broadcast live on the platform made it impossible to monitor it all.[190] Shockingly, after the Christchurch massacre was broadcast on its own platform, it took the company's senior leadership more than 10 days to speak publicly about this tragedy, and that only happened after a tremendous amount of public criticism.[191] Facebook said they would explore restrictions on live-streaming from the platform. In response to a tragedy at this scale, one would expect significant changes to Facebook Live and restrictions which would prevent abuses like this from ever happening again. But when change came, it was minor, and announced six weeks after the attacks. In late May of 2019, Facebook announced only a small and relatively insignificant change that said that anyone breaking certain rules in broadcasting content on Facebook Live would be temporarily barred from using the service, with the possibility of a 30-day ban on a first offense.[192] (Previously, it did not typically bar users until they had broken these rules multiple times.) Multiple offenders, or people who posted particularly egregious content, *could* be permanently barred from Facebook.

Inspiration for the attack leads back to Facebook

Tarrant may have first been exposed to the ideology that propelled his attack because it was circulated widely on Facebook for many, many years. The shooter had been radicalized into the white supremacist Identitarian movement and was an adherent of the racist Great Replacement theory - an international white supremacist conspiracy that promotes the idea that white people are slowly experiencing a genocide in their own home countries due to a plot by elites to displace them with rising numbers of non-white immigrants.[193] This racist thinking is most prominently recognized in the form of the sprawling, multinational organization, Generation Identity,[194] to which Tarrant gave a donation.[195] For years, GI material flourished on Facebook and garnered many thousands of followers until the movement was deplatformed in 2018.[196] Even after the deplatforming, the Institute for Strategic Dialogue identified some 11,000 members of Facebook groups devoted to the Great Replacement idea.[197]

188 Nicolas Vega, "Facebook: We Can't."
189 Mix, "Facebook took 2 weeks to remove video of 12-year-old girl that livestreamed her suicide," *The Next Web*, Jan. 16, 2017.
190 Nicolas Vega, "Facebook: We can't stop all live-stream suicides," *The New York Post*, Oct. 25, 2017.
191 Julie Carrie Wong, "Facebook finally responds to New Zealand on Christchurch attack," *The Guardian*, March 29, 2019.
192 Cade Metz and Adam Satariano, "Facebook Restricts Live Streaming After New Zealand Shooting," *The New York Times*, May 14, 2019.
193 Jacob Davey and Julia Ebner, *'The Great Replacement:' The Violent Consequences of Mainstreamed Extremism*, Institute for Strategic Dialogue, 2019.
194 Heidi Beirich and Wendy Via, "International White Nationalist Movement."
195 Jason Wilson, "Christchurch shooter's links to Austrian far right 'more extensive than thought,'" *The Guardian*, May 15, 2019.
196 *Al Jazeera*, "Facebook pulls the plug on far-right Generation Identity," *Al Jazeera*, June 22, 2018.
197 Jacob Davey and Julia Ebner, "'*The Great Replacement.*'"

Sri Lanka

Hate speech and rumors targeting Muslims contributed to a 2018 outbreak of anti-Muslim violence in Sri Lanka.[198] One viral video on Facebook falsely depicted a Muslim restaurateur seemingly admitting to mixing "sterilization pills" into the food of Sinhala-Buddhist men;[199] other heinous material, such as a post advocating to "kill all Muslims, do not spare even an infant, they are dogs,"[200] contributed to an outbreak of anti-Muslim violence in Sri Lanka in 2018.[201] At least three people were killed and 20 injured in the 2018 unrest, during which mosques and Muslim businesses were burned, mainly in the central part of the Buddhist-majority nation. Facebook left most of the incendiary material up for days, even after it had been reported.[202] Ultimately, the Sri Lankan government had to shut Facebook down to stem the violence. Sri Lankan officials ultimately determined that mobs used Facebook to coordinate attacks, and that the platform had "only two resource persons" to review content in Sinhala, the language of Sri Lanka's ethnic majority whose members were behind the violence.[203]

A 2018 audit, contracted by Facebook from the human rights organization Article One found that hate speech and rumors spread on Facebook "may have led to 'offline' violence."[204] Facebook, more so than any other social media platform, was used by Buddhist nationalists to spread propaganda against Sri Lanka's Muslims, which make up 10 percent of the country's population.[205] Article One suggested that before the unrest, Facebook had failed to take down such hate content, which "resulted in hate speech and other forms of harassment remaining and even spreading" on the platform.

> "Kill all Muslims, do not spare even an infant, they are dogs.
>
> **Hate Speech**
> *Facebook post*

198 Amalini De Sayrah, "Facebook helped foment anti-Muslim violence in Sri Lanka. What now?" *The Guardian*, May 5, 2018.
199 *Euraisan Times* Desk, "After Rohingya Genocide; Facebook Now Apologizes To Sri Lanka Over Anti-Muslim Riots," *Eurasian Times*, May 15, 2020.
200 *Today*, "Facebook admits 'serious mistake' after Edwin Tong questions failure to remove hate post in Sri Lanka," *Today*, Nov. 28, 2018.
201 *Al Jazeera*, "Sri Lanka: Facebook apologises for role in 2018 anti-Muslim riots," *Al Jazeera*, May 13, 2020.
202 Amalini De Sayrah, "Facebook helped foment."
203 *Al Jazeera*, "Sri Lanka."
204 *Al Jazeera*, "Sri Lanka."
205 Tasnim Nazeer, "Facebook's Apology."

"We deplore this misuse of our platform," the company said in a response to the Sri Lanka report.[206] "We recognize, and apologize for, the very real human rights impacts that resulted." Facebook also highlighted actions it had taken to address the problems, including hiring content moderators with local language skills, implementing technology that automatically detects signs of hate speech and keeps abusive content from spreading, and trying to deepen relationships with local civil society groups. However, again, it was too little, too late. Significant damage, strained relations between communities, and a rise in anti-Muslim sentiment had already fueled tensions.



> We recognize, and apologize for, the very real human rights impacts that resulted.
>
> **Facebook**

Muslims in Sri Lanka were the targets of hate speech and rumors circulating on Facebook that led to an outbreak of anti-Muslim violence in 2018. Photo: Kevin Hellon

206 Joshua Brustein, "Facebook Apologizes for Role in Sri Lankan Violence," *Bloomberg*, May 12, 2020.

Sweden

In late August 2020, far-right Danish politician Rasmus Paludan, head of the Stram Kurs (Hard Line) anti-immigrant party, tried to cross over the bridge from Denmark into Malmo, Sweden, for a Quran burning event with his anti-Muslim allies.[207] He was turned away by Swedish authorities on the bridge and barred from Sweden for two years. Paludan responded with an angry, anti-Muslim message on Facebook: "Sent back and banned from Sweden for two years. However, rapists and murderers are always welcome!"[208]

Despite Paludan's absence, his supporters moved forward with their protest and burned a Quran near a Malmo mosque. Riots broke out resulting in significant property destruction, all of which was propelled by social media videos of the desecration that were circulated widely online.[209] Paludan, who has an active Facebook page full of anti-Muslim videos, was sentenced to a month in jail in Denmark for a string of offences, including racism, in early 2020. His conviction included charges regarding posting anti-Muslim videos on social media channels. Paludan's Facebook page remained active[210] as of September 2020.[211]



Members of the Stram Kurs (Hard Line) anti-immigrant party burn a copy of the Quran and circulate the video online leading to mass riots.
Photo: Anonymous

207 *Al Jazeera*, "Sweden: Far-right activists burn copy of Quran, plan more stunts," *Al Jazeera*, Sept. 10, 2020.
208 *Deutsche Welle*, "Malmo: Violence erupts after anti-Muslim activity," *Deutsche Welle*, Aug. 29, 2020.
209 *Deccan Chronicle*, "Riot in Sweden after Danish leader banned from Quran-burning rally," *Deccan Chronicle*, Aug. 29, 2020.
210 https://www.facebook.com/rasmuspaludan.
211 *BBC*, "Protest against Quran-burning turns violent in Sweden," *BBC*, Aug. 29, 2020.

United States

Facebook is an American company, and as such, its actions and policies reflect on the U.S. and its influence in other countries. And its decisions to allow American-produced, anti-Muslim hate content to flourish on its platform no doubt serves as an example of what the company finds acceptable worldwide.

Facebook knows that it allows anti-Muslim content

The problem with anti-Muslim hate on Facebook has been widely documented for years. An analysis of Facebook data in early 2020 showed that the United States and Australia "lead in the number of active Facebook pages and groups dedicated to referring this dehumanizing [anti-Muslim] content."[212] In May 2020, the Tech Transparency Project found more than 100 American white supremacist groups, many of them explicitly anti-Muslim, active on the platform both on their own group pages as well as on auto-generated content. In the wake of TTP's report, Facebook did nominally alter some of the auto-generated content, but the hate groups largely remained.[213]

An analysis by computer scientist Megan Squire of far-right groups on Facebook found a significant crossover with anti-Muslim hate.[214] Squire found that anti-Muslim attitudes are not only flourishing on the platform, but also acting as a "common denominator" for a range of other extremist ideologies, including xenophobic anti-immigrant groups, pro-Confederate groups, militant anti-government conspiracy theorists, and white nationalists. Squire said of her research, "Some of the anti-Muslim groups are central players in the hate network as a whole. And the anti-Muslim groups show more membership crossover with other ideologies than I expected."[215]

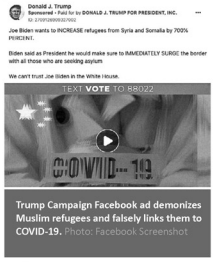
A 2018 study by the Southern Poverty Law Center (SPLC) identified 33 anti-Muslim Facebook groups that used violent imagery, including weapons, in their main photos.[216] According to the SPLC, twenty of these groups promoted the stereotype that all Muslims are violent. One group, "Islam Unveiled," had an image of an ISIS fighter executing prisoners lying prone in a shallow grave. The picture was accompanied by a quote credited to "Muhammad, Prophet of Islam," which says, "Killing Unbelievers is a SMALL MATTER to us." The SPLC also found a group, "PRO-ISLAMOPHOBIA SAVES LIVES!!" whose cover photo was an image of a charred, blackened body which should have been prohibited according to Facebook's claim that posts depicting "charred or burning people" are prohibited.[217] After an academic researcher flagged the group, an automated response was generated and provided to SPLC: "We looked over the group you reported and though it doesn't go against one of our specific Community Standards, we understand that the group or something shared in it may still be offensive to you."[218]

212 Rita Jabri-Markwell, "The Online Dehumanization."
213 Tech Transparency Project, *White Supremacist Groups*.
214 Ishmael N. Daro, "Here's How."
215 Ishmael N. Daro, "Here's How."
216 Swathi Shanmugasundaram, "The persistence of anti-Muslim hate on Facebook," Southern Poverty Law Center's Hatewatch, May 5, 2018.
217 Michal Lev-Ram, "Facebook Tries to Bring More Transparency to Opaque Set of Guidelines," *Fortune*, April 24, 2018.
218 Swathi Shanmugasundaram, "The persistence."

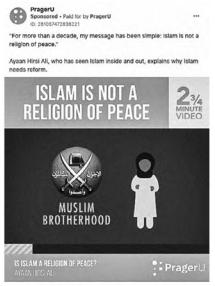
There are other ways anti-Muslim material is spread on Facebook. In 2019, a major study found dozens of current and former American law enforcement officers as members of Facebook groups dedicated to anti-Muslim bigotry.[219] Many were private groups, where this hatred was allowed to flourish outside of any oversight. With names such as "Veterans Against islamic Filth," "PURGE WORLDWIDE (The Cure for the Islamic disease in your country)" and "Americans Against Mosques," these groups serve as private forums to share bigoted messages about Muslims, and they have proven attractive to police officers.

Facebook's problem with anti-Muslim content was also documented by its own civil rights audit, which came after years of pressure from human rights advocates to address hate content online.[220] Released in July 2020, the auditors singled out the problem of anti-Muslim hate as needing to be addressed, their conclusions were partly based on information provided by human rights groups who had been lobbying the company for years to take the problem seriously. Facebook has, thus far, not addressed any of the anti-Muslim problems found by its own audit. The company's actions indicate that its decision to engage in the audit at all was not out of concern for lives lost because of its inaction, but for political expediency.

Perhaps the most notable case of allowing anti-Muslim hate speech on Facebook, despite it violating the site's rules, involves President Donald Trump. In 2016, Zuckerberg decided not to remove a post by Trump calling for a ban of all Muslims entering the U.S.[221] Zuckerberg acknowledged in a meeting with his staff that Trump's call for a ban did qualify as hate speech, but said the implications of removing them were too drastic.[222] Once again, Facebook refused to apply its hate speech policies against a politically powerful individual.



Trump Campaign Facebook ad demonizes Muslim refugees and falsely links them to COVID-19. Photo: Facebook Screenshot



Facebook ad from right-wing propaganda group PragerU promotes the idea that Islam is inherently violent. Photo: Facebook Screenshot

219 Will Carless and Michael Corey, "American cops have openly engaged in Islamophobia on Facebook, with no penalties," *Reveal News*, June 27, 2019.
220 Facebook, *Facebook's Civil Rights Audit*.
221 Julia Carrie Wong, "Zuckerberg proves he is Facebook's editor by allowing Trump's hate speech," *The Guardian*, Oct. 21, 2016.
222 Julia Carrie Wong, "Zuckerberg proves."

Double standard for targeting Muslim public officials

While President Trump's Facebook anti-Muslim post is protected, Muslim public figures have been threatened and attacked on the platform. On numerous occasions, Muslim public officials in Congress and around the country have been targeted with hateful content – even death threats – on Facebook and Instagram.[223] Additionally, two Muslim congresswomen, Ilhan Omar and Rashida Tlaib, were targeted by an international fake news operation that spread anti-Muslim propaganda on Facebook.[224] Facebook also allowed the Trump campaign to run multiple, false ads[225] against these congresswomen.[226] No action was taken by Facebook against the ads.

Perhaps most shockingly, Facebook allowed a man charged with threatening to kill Congresswoman Ilhan Omar to post violent and racist content for years, and took no action to remove his posts when he was arrested in 2019.[227] Upstate New Yorker Patrick Carlineo posted several entries to his Facebook page alluding to violence against Muslims and U.S. officials, including former president Barack Obama. Carlineo frequently posted anti-Muslim material, using racist slurs and saying he wished he could confront a group of Muslim politicians with "a bucket of pig blood." Carlineo's profile wasn't removed by Facebook until a reporter contacted them two weeks after he was arrested for threatening to kill Omar (he pleaded guilty and was sentenced to one year in prison).[228]

Abuse of the event organizing pages

Facebook's event pages have been particularly problematic. For years, white nationalists, militias, and anti-Muslim hate groups have been using Facebook event pages to organize armed, hate rallies targeting mosques and Muslim community centers across the country.[229] Shockingly, the company permitted white nationalist militias to directly intimidate worshippers and threaten mosques. In 2016, two Russian Facebook pages organized dueling rallies in front of the Islamic Da'wah Center of Houston.[230] Heart of Texas, a Russian-controlled Facebook group that promoted Texas secession, played into the stereotype of the state as a land of guns and barbecue and amassed hundreds of thousands of followers. One of their ads on Facebook announced a noon rally on May 21, 2016 to "Stop Islamification of Texas." A separate Russian-sponsored group, United Muslims of America (stealing the identity of a legitimate California-based Muslim organization), advertised a "Save Islamic Knowledge" rally for the same place and time. The armed protest was peaceful but terrorized those inside the religious center.[231]

223 Mellissa Nan Burke, "Tlaib not cowed by 'hateful' threats, behavior," *The Detroit News*, Jan. 27, 2019.
224 David Smith, Michael McGowan, Christopher Knaus and Nick Evershed, "Revealed: Ilhan Omar and Rashida Tlaib targeted in far-right fake news operation," *The Guardian*, Dec. 5, 2019.
225 Muslim Advocates, *Donald Trump Ad*.
226 Andrea Germanos, "Facebook and Twitter urged to suspend Donald Trump after attack on Ilhan Omar," *Salon*, April 15, 2019.
227 Jon Swaine, "Facebook allowed violent posts by man charged with Ilhan Omar death threat," *The Guardian*, April 16, 2019.
228 Jon Swaine, "Facebook allowed."
229 Muslim Advocates, "Facebook Ignored Warnings About Event Page Abuses for Years," Press Release, Sept. 16, 2020.
230 Claire Allbright, "A Russian Facebook page."
231 Claire Ballor, "Shariah law protesters, some toting rifles, gather in front of North Texas Islamic center," *The Dallas Morning News*, June 10, 2017.



In August of 2020, a group of anti-Muslim activists used Facebook to live-stream a hate rally outside a mosque in Milwaukee, Wisconsin. Photo: Facebook Livestream

Facebook's civil rights auditors highlighted the company's failure to enforce policies prohibiting a call to arms during an anti-Muslim protest organized on their events pages during August of 2019.[232] The auditors described an event page which was used to intimidate attendees of the Islamic Society of North America's annual convention in Houston, Texas. Despite the fact that this was the second year in a row where this same hate group threatened the conference, it took Facebook more than 24 hours to remove the event page. Facebook did later acknowledge that the Houston incident represented an enforcement misstep, and the auditors used this example to conclude that Facebook's "events policy provides another illustration of the need for focused study and analysis on particular manifestations of hate."[233]

Live-streaming hate

In August of 2020, a group of anti-Muslim activists used Facebook to live-stream a hate rally outside a mosque in Milwaukee, Wisconsin.[234] Holding a sign that read "Halt Islam," anti-Muslim street preacher Ruben Israel yelled hateful, threatening slurs from a megaphone outside the Islamic Society of Milwaukee, the largest mosque in the city. During the protest, Israel used Facebook to broadcast multiple false, offensive slurs and conspiracies about Muslims, shouting about "wicked, perverted Islam" and asking Muslims in the mosques if they have "anything ticking" on them, and whether they had a pilot's license. He also told a Muslim couple near the mosque, "don't tell me you're here getting government assistance while you hate our country." Despite these clear violations of Facebook's hate speech and live-streaming policies, it took outside groups to alert the company before the content was removed.

232 Facebook, *Facebook Civil Rights Audit*.
233 Facebook, *Facebook Civil Rights Audit*.
234 Muslim Advocates, "Facebook Allows Livestream of Anti-Muslim Mosque Protest in Milwaukee," Press Release, Aug. 24, 2020.

Conclusion

In its August 2020 update of its hate speech policies, which banned conspiracy theories and stereotypes targeting vulnerable communities,[235] Facebook yet again left Muslims off the list.[236] This was just another example in a long line of disappointing decisions made by the company, particularly given the conclusions of Facebook's own audit released the month prior that highlighted the significant problem of anti-Muslim hate on the platform. From the documentation in this report of international anti-Muslim violence stoked on Facebook, it seems clear that Facebook has no regard for the lives of Muslims. The August 2020 policy changes banning blackface and anti-Semitic conspiracy theories and stereotypes are clearly welcome and important expansions of the company's hate speech policies.[237] It is baffling, however, that the company chose not to include implicit stereotypes and tropes about Muslims, such as the false ideas that they are inherently violent, foreign, or criminal. These stereotypes are all dangerous and have resulted in discrimination, hate crimes, and mass murder of Muslims around the world. Yet, given the many opportunities and repeated warnings over many years to protect Muslims from hate content on its platform, the company has refused to act. Ultimately, one thing is clear: for Facebook, jeopardizing the safety and security of Muslims is just the cost of doing business.

Facebook has **no regard** for the lives of Muslims.

For Facebook, jeopardizing the safety and security of Muslims is just the cost of doing business.

235 Jacob Kastrenakes, "Facebook bans blackface and anti-Semitic stereotypes in hate speech update," *The Verge*, Aug. 11, 2020.
236 Muslim Advocates, "New Facebook Hate Speech Policy Update Leaves Out Muslims," Press Release, Aug. 13, 2020.
237 *BBC*, "Facebook adds 'blackface' photos to banned posts," *BBC*, Aug. 12, 2020.

APPENDIX

Timeline of Failure:

Facebook ignores five years of warnings of event pages being used to promote violence

The following is a five-year timeline of Muslim Advocates' repeated attempts to warn Facebook about event page abuses and of the company's repeated failure to heed those warnings:

2015
May 29

Notorious, anti-Muslim bigot and militia supporter Jon Ritzheimer organized a hate rally targeting the Islamic Community Center of Phoenix during Friday prayers. The rally was organized on a Facebook event page which had 600 RSVPs and stated, "People are also encouraged to utilize (sic) second amendment right at this event just (sic) our first amendment comes under the much anticipated attack." Muslim Advocates spoke to Facebook about this rally and event page and expressed our concern about event pages being used by hate groups. ***Nothing was done.***

2015
October

Jon Ritzheimer again used Facebook event pages to organize protests targeting more than 20 mosques around the country. Muslim Advocates complained about these protests and again demanded that Facebook address the targeting of Muslims on the platform's event pages. Even though Muslim Advocates was clear that lives were in danger, ***Facebook did nothing.***

2016
February

Muslim Advocates had an in-person meeting with Monika Bickert, head of policy development, at Facebook's headquarters, where she was shown photos of armed protestors in front of mosques organized on Facebook event pages. Muslim Advocates asked her to change event page policies because lives were in danger. ***Facebook failed to take any action.***

2017
June

ACT for America, the largest anti-Muslim group in the U.S., used Facebook event pages to organize protests in front of Muslim spaces in cities all over the country. Many of these event pages included a call to arms. The June date involved protests in 60 cities. Muslim Advocates again pleaded with Facebook to take action and expressed concern about the safety of the Muslim community. During 2017, Muslim Advocates began demanding that Facebook create a non-discrimination policy banning white nationalists and hate groups from using events pages from targeting communities based on identity. Again, ***Facebook did nothing.***

MUSLIM ADVOCATES GPAHE

2017
September

Documents released by Congress revealed that Russian-linked groups used Facebook event pages to promote anti-Muslim animus and direct protesters to a Houston mosque. After this revelation was made public, Muslim Advocates again went to Facebook and warned them about how event pages were being used to promote anti-Muslim hate and endanger Muslims.

2017
Oct 11

Muslim Advocates participated in a small group meeting with Facebook Chief Operating Officer Sheryl Sandberg, where she again was told about the abuse of the events pages and how they were being used by white nationalists to target mosques and Muslims. Muslim Advocates shared with her of protestors brandishing guns at these rallies and asked her to take action. ***Nothing was done.***

2019
Mar 7

Muslim Advocates and coalition partners met with Facebook officials and had an extensive discussion about hate groups and events pages, again requesting polices for the pages be reviewed and updated in order to stop white supremacist groups from using the pages to targets Muslims and other vulnerable communities.

2019
August

A militia group, the Texas Patriot Network, used a Facebook event page to plan an armed protest at the largest Muslim convention in the country. Muslim Advocates warned Facebook about the page and had to wait more than 24 hours before it was taken down. Even worse, the same group planned a similar armed protest at the same event the year prior where militia members were found to have discussed possible targets and called them "dead meat." Despite Mark Zuckerberg's claim that "we do not allow hate groups on Facebook," this group still has an active Facebook presence.

2019
Nov 4

Muslim Advocates Executive Director Farhana Khera attended a small group dinner with Facebook CEO Mark Zuckerberg, Sheryl Sandberg and other members of Facebook leadership. She highlighted the problem with the event pages again and asked them to stop white nationalists and hate groups from abusing their event pages. ***Nothing happened.***

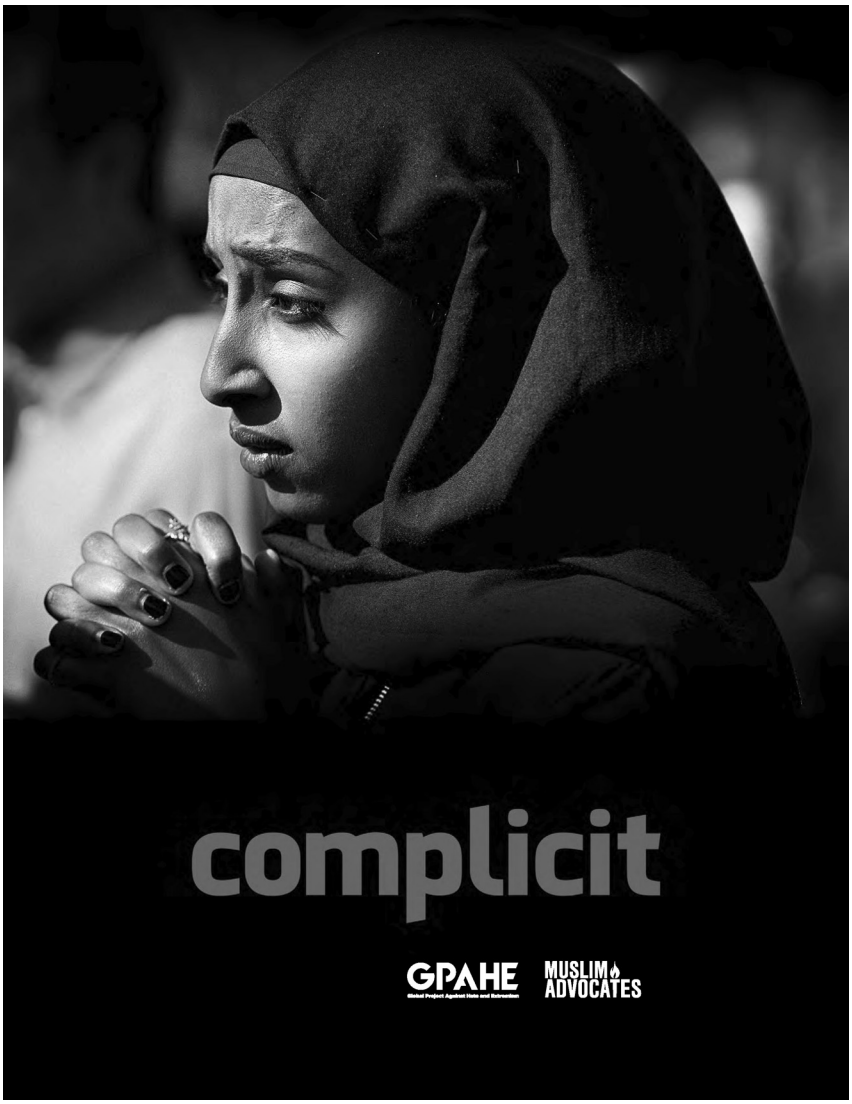


2020
Jul 10

During a call with Sheryl Sandberg about the final report of the civil rights audit, Muslim Advocates yet again asked for Facebook to end the targeting of vulnerable communities on the event pages. Muslim Advocates warned that people were going to get hurt but no action was taken.

2020
Aug 26

An armed 17-year-old who drove up from Illinois shot and killed two people protesting the police killing of James Blake in Kenosha, Wisconsin. It was later reported that a Facebook event page from a local Kenosha militia that included a call to arms was flagged at least 455 times but not taken down by the company.


complicit
GPAHE
Global Project Against Hate and Extremism
MUSLIM ADVOCATES