S. HRG. 116–609

MASS VIOLENCE, EXTREMISM, AND DIGITAL RESPONSIBILITY

HEARING

BEFORE THE

COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION UNITED STATES SENATE

ONE HUNDRED SIXTEENTH CONGRESS

FIRST SESSION

SEPTEMBER 18, 2019

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

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

U.S. GOVERNMENT PUBLISHING OFFICE
52–751 PDF WASHINGTON : 2023

SENATE COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION

ONE HUNDRED SIXTEENTH CONGRESS

FIRST 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 September 18, 2019 1

Statement of Senator Wicker 1

Statement of Senator Cantwell 3

Statement of Senator Fischer 36

Statement of Senator Blumenthal 38

Statement of Senator Thune 39

Statement of Senator Blackburn 41

Statement of Senator Scott 43

Statement of Senator Duckworth 45

Statement of Senator Young 48

Statement of Senator Rosen 51

Statement of Senator Lee 52

Statement of Senator Baldwin 54

Statement of Senator Sullivan 56

Statement of Senator Cruz 58

WITNESSES

Monika Bickert, Vice President for Global Policy Management and Counterterrorism, Facebook 4

Prepared statement 6

Nick Pickles, Director, Public Policy Strategy, Twitter, Inc. 8

Prepared statement 10

Derek Slater, Global Director, Information Policy, Google LLC 15

Prepared statement 17

George Selim, Senior Vice President, National Programs, ADL (Anti-Defamation League) 19

Prepared statement 21

APPENDIX

Letter dated September 16, 2019 to Chairman Roger Wicker and Ranking Member Maria Cantwell from Gretchen Peters, Executive Director, and Dr. Amr Al-Azm, Co-founder, Director of ATHAR Project, Alliance to Counter Crime Online 63

Letter dated September 17, 2019 to Mark Zuckerberg, Chief Executive Officer, and Sheryl Sandberg, Chief Operating Officer, Facebook; Jack Dorsey, Chief Executive Officer, Twitter; Sundar Pichai, Chief Executive Officer, Google; and Susan Wojcicki, Chief Executive Officer, YouTube from The Leadership Conference on Civil and Human Rights, Color Of Change, Lawyers' Committee for Civil Rights Under Law, Muslim Advocates, and NAACP Legal Defense and Educational Fund, Inc. 64

Response to written questions submitted to Monika Bickert by:

Hon. Shelley Moore Capito 66

Hon. Amy Klobuchar 67

Hon. Richard Blumenthal 68

Hon. Tom Udall 70

Hon. Jacky Rosen 75

Response to written questions submitted to Nick Pickles by:

Hon. Shelley Moore Capito 76

Hon. Richard Blumenthal 77

Hon. Tom Udall 80

Hon. Jacky Rosen 83

Response to written questions submitted to Derek Slater by:

Hon. Shelley Moore Capito 85

Page
Response to written questions submitted to Derek Slater by—Continued
Hon. Richard Blumenthal 86
Hon. Tom Udall 89
Hon. Jacky Rosen 91
Response to written questions submitted to George Selim by:
Hon. Richard Blumenthal 93
Hon. Tom Udall 95
Hon. Jacky Rosen 95

MASS VIOLENCE, EXTREMISM, AND DIGITAL RESPONSIBILITY

WEDNESDAY, SEPTEMBER 18, 2019

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

The Committee met, pursuant to notice, at 10 a.m., in room SH–216, Hart Senate Office Building, Hon. Roger Wicker, Chairman of the Committee, presiding.

Present: Senators Wicker [presiding], Thune, Cruz, Fischer, Sullivan, Blackburn, Young, Capito, Lee, Scott, Cantwell, Blumenthal, Udall, Duckworth, and Rosen.

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

The CHAIRMAN. Today the Committee gathers to discuss what the technology industry is doing to remove violent and extremist content from their platforms. This is a matter of serious importance to the safety and well-being of our Nation's communities. I sincerely hope we can engage in a collaborative discussion about what more can be done, within the jurisdiction of this committee, to keep our communities safe from those wishing to do us harm. Today, we welcome representatives from the world's largest social media companies and online platforms.

We will hear from Ms. Monika Bickert, Head of the Global Policy Management for Facebook, and Mr. Nick Pickles, Public Policy Director at Twitter, Mr. Derek Slater, Global Director of Information Policy at Google, and Mr. George Selim, Senior Vice President of Programs for the Anti-Defamation League. Over the past two decades, the United States has led the world in the development of social media and other services that allow people to connect with one another.

Open platform providers like Google, Twitter, and Facebook and products like Instagram and YouTube have dramatically changed the way we communicate and have been used positively in providing spaces for like-minded groups to come together and in shedding light on despotic regimes and abuses of power throughout the world. No matter how great the benefits to society these platforms provide, it is important to consider how they can be used for evil at home and abroad. On August 3, 2019, 20 people were killed, and more than two dozen were injured in a mass shooting at an El Paso shopping center. Police have said that they are reasonably confident that the suspect posted a manifesto to a website called

8chan, 27 minutes prior to the shooting. 8chan moderators removed the original post, though users continued sharing copies.

Following the shooting, President Trump called on social media companies to work in partnership with local, State, and Federal agencies to develop tools that can detect mass shooters before they strike—I certainly hope we talk about that challenge today. Sadly, the El Paso shooting is not the only recent example of mass violence with an online dimension.

On March 15, 2019, 51 people were killed and 49 were injured in shootings at two mosques in Christchurch, New Zealand. The perpetrator filmed the attacks using a body camera and livestreamed the footage to his Facebook followers, who began to reupload the footage to Facebook and other sites. Access to the footage quickly spread and Facebook stated that it removed 1.5 million videos of the massacre within 24 hours of the attack. 1.2 million views of the videos were blocked before they could be uploaded. Like the El Paso shooter, the Christchurch shooter also uploaded a manifesto to 8chan. The 2016 shooting at the Pulse Nightclub in Orlando, Florida, killed 49 and injured 53 more. The Orlando shooter was reportedly radicalized by ISIS and other jihadist propaganda through online sources. Days after the attack, the FBI Director stated that investigators were highly confident that the shooter was self-radicalized through the internet.

According to an official involved in the investigation, analysis of the shooter's electronic devices revealed that he had consumed "a hell of a lot of jihadist propaganda," including ISIS beheading videos. Shooting survivors and family members of victims brought a Federal lawsuit against those three social media platforms under the Anti-Terrorism Act. The Sixth Circuit dismissed the lawsuit on the grounds that this was not an act of international terrorism. With over 3.2 billion Internet users, this committee recognizes the challenge facing social media companies and online platforms and their ability to act and remove content threatening violence from their sites.

There are questions about tracking of a users' online activity: does this invade an individual's privacy, thwart due process, or violate constitutional rights? The automatic removal of threatening content may also impact an online platform's ability to detect possible warning signs. Indeed, the First Amendment offers strong protections against restricting certain speech. This undeniably adds to the complexity of our task. I hope these witnesses will speak to these challenges and how their companies are navigating these challenges. In today's internet-connected society, misinformation, fake news, deep fakes, and viral online conspiracy theories have become the norm. This hearing is an opportunity for witnesses to discuss how their platforms go about identifying content and material that threatens violence and poses a real and potentially immediate danger to the public.

I hope our witnesses will also discuss how their content moderation processes work. This includes addressing how human review or technological tools are employed to remove or otherwise limit violent content before it is posted, copied, and disseminated across the internet. Communication with law enforcement officials at the Federal, State, and local levels is critical to protecting our neighborhoods and communities. We would like to know how companies are coordinating with law enforcement when violent or extremist content is identified.

And finally, I hope witnesses will discuss how Congress can assist in ongoing efforts to remove content promoting violence from online platforms and whether best practices or industry codes of conduct in this area would help increase safety, both online and offline. So, I look forward to hearing testimonies from our witnesses, and hope we engage in a constructive discussion about potential solutions to a pressing issue. And I am delighted at this point to recognize my friend and Ranking Member, Senator Cantwell.

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

Senator CANTWELL. Thank you, Mr. Chairman, and thank you for holding this important hearing and for our witnesses being here this morning. Across the country, we are seeing and experiencing a surge of hate and as a result we need to think much harder about the tools and resources we have to combat this problem both online and offline. While the First Amendment to the Constitution protects free speech, speech that incites eminent violence is not protected, and Congress should review and strengthen laws that prohibit threats of violence, harassment, stalking, and intimidation to make sure that we stop the online behavior that does incite violence.

In testimony before the Senate Judiciary Committee in July, Federal Bureau of Investigation FBI Director Chris Wray said that the white supremacist violence is on the rise. He said the FBI takes this threat “extremely seriously,” and has made over 100 arrests so far this year. We are seeing in my state over the last several years, we have suffered a shooting at the Jewish community center in Seattle, a shooting of a Sikh in Kent, Washington, a bombing attempt at the Martin Luther King Day parade in Spokane, and over the last year, we have seen a rise in the desecration of both synagogues and mosques. The rise in hate across the country has also led to multiple mass shootings, including the Tree of Life congregation in Pittsburgh, the Pulse nightclub in Orlando, and most recently, the Walmart in El Paso.

Social media is used to amplify that hate and the shooter at one high school in the Parkland posting said the image of himself with guns and knives on Instagram wrote social media posts prior to the attack on his fellow students. In El Paso, the killer published a white supremacist anti-immigration manifesto on a 8chan message board, and my colleague just mentioned this streaming of live content related to the Christchurch shooting and the horrific incidents that happened there. In Miramar, the military engaged in a systematic campaign on Facebook, using fake names and sham accounts to promote violence against Muslim Rohingya. These human lives were all cut short by deep hatred and extremism that we have seen has become more common.

This is a particular problem on the dark web, where we see certain websites like 8chan and a host of 24/7, 365 hate rallies. Adding technology tools to mainstream websites to stop the spread of these dark websites are a start, but there needs to be more to be a comprehensive and coordinated effort to ensure that people are not directed into these cesspools. I believe calling on the Department of Justice to make sure that we are working across the board on an international basis with companies as well to fight this issue is an important thing to be done. We do not want to push people off of social media platforms only to then being on the dark web, where we are finding less of them. We need to do more at the Department of Justice to shut down these dark websites, and social media companies need to work with us to make sure that we are doing this. I do want to mention, just last week, as there is much discussion here in Washington about initiatives.

The State of Washington has passed three gun initiatives by the vote of the people, closing background loopholes, and also relating to private sales and extreme person laws, all voted on by a majority of people in our state and successfully passed. So I do appreciate, just last week representatives from various companies of all sizes in the tech industry sending the Senate a letter, asking for passage of bills requiring extensive background checks.

So very much appreciate that and your support of extreme person laws to keep guns out of the hands of people who a court has determined are dangerous in the possession of that. So this morning, we look forward to asking you about ways in which we can better fight these issues. I do want us to think about ways in which we can all work together to address these issues. I feel that working together, these are successful tools that we can deploy in trying to fight extremism that exists online. Thank you, Mr. Chairman, for the hearing.

The CHAIRMAN. Thank you, very, very much. And now we will hear oral testimony from our four witnesses. And we ask you—your entire statements will be submitted for the record, without objection. We ask you to limit your comments at this point to five minutes. Ms. Bickert, you are recognized. Thank you for being here.

STATEMENT OF MONIKA BICKERT, VICE PRESIDENT FOR GLOBAL POLICY MANAGEMENT AND COUNTERTERRORISM, FACEBOOK

Ms. BICKERT. Thank you, Chairman Wicker, Ranking Member Cantwell, and distinguished members of the Committee. Thank you for the opportunity to be here today, and to answer your questions and explain our efforts in these areas. My name is Monika Bickert and I am Facebook's Vice President for Global Policy Management and Counterterrorism. I am responsible for our rules around content on Facebook and our company's response to terrorist with the intent to use our services. On behalf of everyone at Facebook, I would like to begin by expressing my sympathy and solidarity with the victims, families, communities, and everybody affected by the recent terrible attacks across the country.

In the face of such heinous acts, we remain committed to assisting law enforcement and standing with the community against hate and violence. We are thankful to be able to provide a way for those affected by this horrific violence to communicate with loved ones, organize events for people to gather and grieve, raise money to help support communities, and begin to heal.

Our mission is to give people the power to connect with one another and to build community. But we know that people need to be safe in order to build that community. And that is why we have rules in place against harmful conduct including hate speech and inciting violence. Our goal is to ensure that Facebook is both a place where people can express themselves, but where they are also safe.

While we are not aware of any connection between the recent attacks and our platform, we certainly recognize that we all have a role to play in keeping our community safe. That is why we remove content that encourages real-world harm, this includes contents that is involving violence or incitement, promoting or publicizing crime, coordinating harmful activities, or encouraging suicide or self-injury. We do not allow any individuals or organizations who proclaim a violent mission, advocate for violence, or are engaged in violence to have any presence on Facebook, even if they are talking about something unrelated, this includes organizations and individuals involved in or advocating for terror activity, domestic and international, organized hate and that includes white supremacy, white separatism, or white nationalism, or other violence.

We also do not allow any content posted by anyone that praises or supports these individuals, organizations, or their actions. When we find content that violates our standards, we remove it promptly, we also disable accounts when we see severe or repeated violations, and we work with law enforcement directly when we believe there is a risk of physical harm or a direct threat to public safety. While there is always room for improvement, we already remove millions of pieces of content every year for violating our policies and much of that is before anybody has reported it to us. Our efforts to improve our enforcement of these policies are focused in three areas.

First, building new technical solutions that allow us to proactively identify content that violates our policies. Second, investing in people who can help us implement these policies. At Facebook, we now have more than 30,000 people across the company who are working on safety and security efforts, this includes more than 350 people whose primary focus is counterhate and counterterrorism.

And third, building partnerships with other companies, civil society, researchers, and Governments so that together we can come up with shared solutions. We are proud of the work we have done thus far to make Facebook a hostile place for those engaged in or advocating for acts of violence, but the work will never be complete.

We know that bad actors will continue to attempt to skirt detection with more sophisticated efforts, and we are dedicated to continuing to advance our work and show our progress. We look forward to working with the Committee, regulators, others in the tech industry, and civil society to continue this progress.

Again, I appreciate the opportunity to be here today, and I look forward to your questions. Thank you.

[The prepared statement of Ms. Bickert follows:]

PREPARED STATEMENT OF MONIKA BICKERT, VICE PRESIDENT FOR GLOBAL POLICY MANAGEMENT AND COUNTERTERRORISM, FACEBOOK

I. Introduction

Chairman Wicker, Ranking Member Cantwell, and distinguished members of the Committee, thank you for the opportunity to appear before you today. My name is Monika Bickert, and I am the Vice President of Global Policy Management and Counterterrorism at Facebook. In that role, I lead our efforts related to Product Policy and Counterterrorism. Prior to assuming my current role, I served as lead security counsel for Facebook, working on issues ranging from children's safety to interactions with law enforcement. And before that, I was a criminal prosecutor with the Department of Justice for 11 years in Chicago and Washington, DC, where I prosecuted Federal crimes including public corruption and gang violence.

On behalf of everyone at Facebook, I would like to express our sympathy and solidarity with the victims, families, communities, and everyone else affected by the recent terrible attacks across the country. In the face of such heinous acts, we remain committed to cooperating with law enforcement and standing with our community against hate and violence. We are thankful to be able to provide a way for those affected by the horrific recent attacks to communicate with loved ones, to organize events for people to gather and grieve, and to raise money to help support these communities as they begin to heal.

Facebook's mission is to give people the power to build community and bring the world closer together. We are proud that more than two billion people around the world come to Facebook every month to connect and share with one another. But people need to feel safe in order to build this community. That is why Facebook prohibits harmful conduct on its platform, including hate speech and inciting violence. Our goal is to ensure that Facebook is a place where both expression and personal safety are protected and respected.

We are not aware of any connection between our platform and the recent attacks, but we recognize that we all have a role to play in keeping our communities safe. At Facebook, we have strong policies and invest significant resources to protect our users on and offline.

II. Facebook's Policies Against Hate and Violence

Facebook is committed to protecting our community by removing any content from our services that encourages real-world harm. Because harmful content can take many forms, we have several policies in place to address these issues, all of which are published in our Community Standards, which define the content that is and is not allowed on our platform.

When we find content that violates our standards, we remove it. We invest in technology, processes, and people to help us identify violations and act quickly to mitigate any impact. There is always room for improvement, but we remove millions of pieces of content every year, much of it before any user reports it. We outline below several of the important steps that we take to prevent violence and keep our users safe.

Prohibition Against Violence and Incitement: We care deeply about our users and we want them to be safe. Therefore, it is critical to our mission to help prevent potential offline harm that may be related to content on Facebook. We remove content, disable accounts, and work with law enforcement when we believe there is a risk of physical harm or direct threats to public safety.

Prohibition of Dangerous Individuals and Organizations: In an effort to prevent and disrupt real-world harm, we do not allow any individuals or organizations that proclaim a violent mission, advocate violence, or are engaged in violence to have a presence on Facebook for any purpose, even if it appears benign. This includes organizations or individuals involved in the following:

- Terrorist activity, both domestic and international;
- Organized hate, including white supremacy and white nationalism;
- Human trafficking; and
- Organized violence or criminal activity.

We do not allow propaganda or symbols that represent any of these organizations or individuals to be shared on our platform unless they are being used to condemn or inform—for example, by media organizations. We do not allow content that praises any of these organizations or individuals or any acts committed by them. And we do not allow coordination of support for any of these organizations or individuals or any acts committed by them.

No Promoting or Publicizing Crime: We prohibit people from promoting or publicizing violent crime, theft, and/or fraud because we do not want to condone this activity and because there is a risk of copycat behavior. We also do not allow people to depict criminal activity or admit to crimes they or their associates have committed.

Policies Against Coordinating Harm: In an effort to prevent and disrupt real-world harm, we prohibit people from facilitating or coordinating future activity, criminal or otherwise, that is intended or likely to cause harm to people, businesses, or animals. People can draw attention to harmful activity that they may witness or experience as long as they do not advocate for or coordinate harm.

Combatting Suicide and Self-Injury: We also use and continue to develop tools and resources to proactively identify and help people who may be at risk of suicide or self-injury. We leverage pattern recognition technology to detect posts or live videos where someone might be expressing an intent to harm themselves. We also use artificial intelligence (AI) to prioritize the order in which our team reviews reported content relating to suicide or self-injury. This ensures we can get the right resources to people in distress and, where appropriate, we can more quickly alert first responders. And we remove content that encourages suicide or self-injury, including certain graphic imagery and real-time depictions that experts tell us might lead others to engage in similar behavior. We also work with organizations around the world to provide assistance and resources to people in distress.

Cooperation with Law Enforcement: Law enforcement plays a critical role in keeping people safe, and we have a long history of working successfully with law enforcement to address a wide variety of threats. As a former Federal prosecutor, I know that this cooperation is vital. When we do receive reports or otherwise find content that violates our policies, we remove it. And we proactively reach out to law enforcement if we see a credible threat of imminent harm.

III. Facebook's Efforts to Combat Violence and Hate

Our efforts to combat violent and hateful content are focused in three areas: developing new technical capabilities for our products, investing in people, and building partnerships.

Product Enhancements: Facebook has invested significantly in technology to help meet the challenge of proactively identifying violent content, including through the use of AI and other automation. These technologies have become increasingly central to keeping hateful and violent content off of Facebook.

We use a wide range of technical tools to identify violent and hateful content. This includes hashes—or digital fingerprints—that allow us to find secondary versions of known bad content; text parsing; digital "fan-outs" to identify profiles, groups, and pages related to those we have identified as problematic; and more holistic machine learning that can assess all aspects of a post and score whether it is likely to violate our Community Standards.

We also know that bad actors adapt as technology evolves, and that is why we constantly update our technical solutions to deal with more types of content in more languages, and to react to the new ways our adversaries try to exploit our products. For example, in response to the tragic events in Christchurch, we made changes to Facebook Live to restrict users if they have violated certain rules—including our Dangerous Organizations and Individuals policy. We now apply a "one-strike" policy to Live: anyone who violates our most serious policies will be restricted from using Live for set periods of time—for example, 30 days—starting on their first offense. We have also updated our proactive detection systems and reduced the average time it takes for our AI to find a violation on Facebook Live to 12 seconds—a 90 percent reduction in our average detection time from a few months ago. Being able to detect violations sooner means that in emergencies where every minute counts, we can assist faster.

Investments in People: We know that we cannot rely on AI alone to identify potentially violent content. Context often matters. To understand more nuanced cases, we need human expertise.

One of our greatest human resources is our community of users. Our users help us by reporting accounts or content that may violate our policies—including the small fraction that may be related to acts of violence. To review those reports, and to prioritize the safety of our users and our platform more generally, we have more than 30,000 people working on safety and security across the company and around the world. That is three times as many people as we had dedicated to such efforts in 2017. Our safety and security professionals review reported content in more than 50 languages, 24 hours a day.

We also have a team of more than 350 people at Facebook whose primary job is dealing with terrorists and other Dangerous Individuals and Organizations. This team includes language and cultural specialists, former law enforcement and intelligence professionals, and academics that have studied these groups and individuals for years. Many of them came to Facebook specifically because they are committed to the mission of keeping people safe.

This team was previously focused on counterterrorism, and we used our most sophisticated tools to predominantly combat ISIS, al-Qaeda, and their affiliates, which were recognized then as posing the greatest threats to our global community. Now, they lead our efforts against all people and organizations that proclaim or are engaged in violence. We are taking the initial progress we made in combatting content affiliated with ISIS, al-Qaeda, and their affiliates, and we are further building out techniques to identify and combat the full breadth of violence and extremism covered under our Dangerous Organizations policy.

Partnerships: We are proud of the work we have done to make Facebook a hostile place for those committed to acts of violence. We understand, however, that simply working to keep violence off Facebook is not an adequate solution to the problem of online extremism and violence, particularly because bad actors can leverage a variety of platforms. We believe our partnerships with other companies, civil society, researchers, and governments are crucial to combatting 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 counterspeech campaigns from students in 75 countries, engaged over 6,500 students, and reached over 200 million people. We're also partnering with Life After Hate, an organization founded by former violent extremists, to connect people who search for terms associated with white supremacy to resources focused on helping people leave behind hate groups.

Our work to combat violence is never complete. Individuals and organizations intent on violent acts come in many ideological stripes—and the most dangerous among them are deeply resilient. We know that bad actors will continue to attempt to skirt our detection with more sophisticated efforts, and we are dedicated to continuing to advance our work and share our progress.

IV. Conclusion

Facebook is committed to helping people build a vibrant community that encourages and fosters free expression. At the same time, we want to do what we can to protect our users from real-world harm and stop terrorists, extremists, hate groups, and any others from using our platform to promote or engage in violence. We recognize that there is always more work to do in combatting the abuse of our site by bad actors, but we are proud of the progress we have made over the last few years. We know that people have questions about what we are doing to continue that progress, and we look forward to working with this Committee, regulators, and others in the tech industry and civil society to continue working on these issues. I appreciate the opportunity to be here today, and I look forward to your questions.

The CHAIRMAN. Thank you very much. Mr. Pickles.

STATEMENT OF NICK PICKLES, DIRECTOR, PUBLIC POLICY STRATEGY, TWITTER, INC.

Mr. PICKLES. Chairman Wicker, Ranking Member Cantwell, members of the Committee, thank you for the opportunity to appear today to discuss these important issues. Twitter has publicly committed to improving the collective health, openness, and civility of public conversation on our platform. Our policies are designed to keep people safe on Twitter and they continuously evolve to reflect the realities of the world we operate in. We are working faster, we are investing to remove content that distracts from healthy conversation before it is reported, including terrorists and violent extremist content.

Tackling terrorism, violent extremism, and preventing violent attacks requires a whole of society response including from social media companies. Let me be clear, Twitter is incentivized to keep terrorists and violent content off our service both from a business standpoint and then the current legal frameworks. Such content does not serve our business interests, it breaks our rules, but is fundamentally contrary to our values. Communities in America and around the world have been impacted by instance of mass violence, terrorism, and violent extremism with tragic frequency in recent years. These events demand a robust public policy response from every quarter.

We acknowledge that technology companies have a role to play. However, it is important to recognize content removal alone cannot solve these issues. I would like to outline four of Twitter's key policies in this area. Firstly, Twitter takes a zero tolerance approach to terrorists content on our service. Individuals may not promote terrorism, engage in terrorist recruitment, or terrorist acts. Since 2015, we have suspended more than 1.5 million accounts for violations of our rules related to terrorism and continue to see more than 90 percent of these accounts suspended through our own proactive measures.

In the majority of cases, we take action at the account creation stage before account has even tweeted, and the remaining 10 percent is identified through a combination of user reports and partnerships. Second, we prohibit the use of Twitter by violent extremist groups. These are defined in our rules as groups that whether by statements on or off the platform use or promote violence against civilians to further their cause whatever their ideology. Since the introduction of this policy in 2017, we have taken action on more than 186 groups globally and suspended more than 2,000 unique accounts. Third, Twitter does not allow hateful conduct on our service.

An individual on Twitter is not permitted to threaten or promote violence or directly attack people based on their protected characteristics. Where any of these rules are broken, we will take action to remove the content and will permanently remove those promoting terrorism or violent extremism from Twitter. Fortunately, our rules prohibit the selling, buying, or facilitating transactions in weapons, including firearms, ammunition, and explosives, and instructions on making weapons. So are explosive devices or 3D printed weapons.

We will take appropriate action on any account found to be engaged in this activity, including a permanent suspension where appropriate. Additionally, we prohibit the promotion of weapons and weapon accessories globally through our paid advertising policies. Collaboration with our industry peers and civil society is critically important to addressing the common threats from terrorism globally.

In June 2017, we launched the Global Internet Forum to Counter Terrorism, GCT, a partnership with Twitter, YouTube, Facebook, and Microsoft. This facilitates, among other things, information-sharing, technical cooperation, research collaboration, including with academic institutions. Twitter and technology companies have a role to play in addressing mass violence, ensuring our platforms cannot be exploited by those promoting violence. This cannot be the only public policy response and removing content alone will not stop those who are determined to cause harm.

Quite often, when we remove content from our platforms, it moves those views, these ideologies into the darker corners of the

Internet where they cannot be challenged and held to account. As our pair companies are improving 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. Addressing mass violence requires a whole of society response. We welcome the opportunity to continue to work with industry peers, Government institutions, legislators, law enforcement, academics, and civil society to find the right solutions. Thank you for your time today.

[The prepared statement of Mr. Pickles follows:]

PREPARED STATMENT OF NICK PICKLES, DIRECTOR, PUBLIC POLICY STRATEGY, TWITTER, INC.

Chairman Wicker, Ranking Member Cantwell, and Members of the Committee:

At Twitter, our mission is to serve the public conversation. Twitter is a place where people from around the world come together in an open and free exchange of ideas. We have made the health of our service the top priority. Conversely, abuse, malicious automation, hateful conduct, violent extremist and terrorist content terrorism, and manipulation will detract from the health of our platform.

Tackling terrorism, violent extremism, and preventing violent attacks require a whole of society response, including from social media. It has long been a priority of Twitter to remove this content from the service. Let me be clear: Twitter has no incentive to keep terrorist and violent extremist content available on our platform. Such content does not serve our business interests, breaks our rules, and is fundamentally contrary to our values.

Communities in America and around the world have been impacted by incidents of mass violence, terrorism, and violent extremism with tragic frequency in recent years. These events demand a robust public policy response from every quarter. We acknowledge that the technology companies play a critical role, however, it is important to recognize content removal online cannot alone solve these issues.

We welcome the opportunity to continue to work with you on the Committee, our industry peers, government, academics, and civil society to find the right solutions. Partnership is essential.

My statement today will provide information and deeper context on: (I) Twitter's work to protect the health of the public conversation, including combating terrorism, violent extremist groups, and hateful conduct; (II) our policies relating to weapons and weapon accessories; and (III) our partnerships and societal engagement.

I. TWITTER'S POLICIES ON TERRORIST CONTENT, VIOLENT EXTREMIST GROUPS, AND HATEFUL CONDUCT

All individuals accessing or using Twitter's services must adhere to the policies set forth in the Twitter Rules. Accounts under investigation or that have been detected as sharing content in violation with the Twitter Rules may be required to remove content, or in serious cases, will see their account permanently suspended. Our policies and enforcement options evolve continuously to address emerging behaviors online.

A. Policy on Terrorism

Individuals on Twitter are prohibited from making specific threats of violence or wish for the serious physical harm, death, or disease of an individual or group of people. This includes, but is not limited to, threatening or promoting terrorism.

We suspended more than 1.5 million accounts for violations related to the promotion of terrorism between August 1, 2015, and December 31, 2018. In 2018, a total of 371,669 accounts were suspended for violations related to promotion of terrorism. More than 90 percent of these accounts are suspended through our proactive measures.

We have a zero-tolerance policy and take swift action on ban evaders and other 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.

Government and law enforcement reports constituted less than 0.1 percent of all suspensions in the last reporting period. Continuing the trend we have seen for some time, the number of reports we received from governments of terrorist content from the second half of last year decreased by 77 percent compared to the previous reporting period covering January through June 2018.

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."

B. Policy on Violent Extremist Groups

In December 2017, we broadened our rules to encompass accounts affiliated with violent extremist groups. Our prohibition on the use of Twitter's services by violent extremist groups—*i.e.,* identified groups subscribing to the use of violence as a means to advance their cause—applies irrespective of the cause of the group.

Our policy states:

Violent extremist groups are those that meet all of the below criteria:

- *identify through their stated purpose, publications, or actions as an extremist group;*
- *have engaged in, or currently engage in, violence and/or the promotion of violence as a means to further their cause; and*
- *target civilians in their acts and/or promotion of violence.*

An individual on Twitter may not affiliate with such an organization—whether by their own statements or activity both on and off the service—and we will permanently suspend those who do so.

We know that the challenges we face are not static, nor are bad actors homogenous from one country to the next in how they behave. Our approach combines flexibility with a clear, consistent policy philosophy, enabling us to move quickly while establishing clear norms of unacceptable behavior.

Since the introduction of our policy on violent extremist groups, we have taken action on 186 groups under this policy and permanently suspended 2,217 unique accounts. Ninety-three of these groups advocate violence against civilians alongside some form of extremist white supremacist ideology.

C. Policy on Hateful Conduct

People on Twitter are not permitted to promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. We also do not allow accounts whose primary purpose is inciting harm toward others on the basis of these categories.

We do not allow individuals to use hateful images or symbols in their profile image or profile header. Individuals on the platform are not allowed to use the username, display name, or profile bio to engage in abusive behavior, such as targeted harassment or expressing hate toward a person, group, or protected category.

Under this policy, we take action against behavior that targets individuals or an entire protected category with hateful conduct. Targeting can happen in a number of ways, for example, mentions, including a photo of an individual, or referring to someone by their full name.

When determining the penalty for violating this policy, we consider a number of factors including, but not limited to the severity of the violation and an individual's previous record of rule violations. For example, we may ask someone to remove the violating content and serve a period of time in read-only mode before they can Tweet again. Subsequent violations will lead to longer read-only periods and may eventually result in permanent account suspension. If an account is engaging primarily in abusive behavior, or is deemed to have shared a violent threat, we will permanently suspend the account upon initial review.

D. Investing in Tech: Behavior vs. Content

Twitter's philosophy is to take a behavior-led approach, utilizing a combination of machine learning and human review to prioritize reports and improve the health of the public conversation. That is to say, we increasingly look at how accounts behave before we look at the content they are posting. This is how we seek to scale our efforts globally and leverage technology even where the language used is highly context specific. Twitter employs extensive content detection technology to identify potentially abusive content on the service, along with allowing users to report content to us either as an individual or a bystander.

For abuse, this strategy has allowed us to take three times the amount of enforcement of action on abuse within 24 hours than this time last year. We now proactively surface over 50 percent of abusive content we remove using our technology compared to 20 percent a year ago. This reduces the burden on individuals to report content to us. Since we started using machine learning three years ago to reduce the visibility on abusive content:

- 80 percent of all replies that are removed were already less visible;
- Abuse reports have been reduced by 7.6 percent;
- The most visible replies receive 45 percent less abuse reports;
- 100,000 accounts were suspended for creating new accounts after a suspension during January through March 2019 — a 45 percent increase from the same time last year;
- 60 percent faster response to appeals requests with our new in-app appeal process;
- 3 times more abusive accounts suspended within 24 hours after a report compared to the same time last year; and
- 2.5 times more private information removed with a new, easier reporting process.

II. TWITTER POLICIES REGARDING WEAPONS AND WEAPON ACCESSORIES

Although Twitter's service does not have an e-commerce function, our Rules prohibit the selling, buying, or facilitating transactions in weapons, including firearms, ammunition, and explosives, and instructions on making weapons, such as bombs or 3D printed weapons. We will take appropriate action on any account found to be engaged in this activity, including permanent suspension of accounts where appropriate.

As stated publicly in our advertising policies, Twitter does not allow the use of our promoted products for the purpose of promoting weapons and weapon accessories globally. We explicitly ban advertising of guns, including airsoft guns, air guns, blow guns, paintball guns, antique guns, replica guns, and imitation guns. Twitter also prohibits the use of our promoted products for gun parts and accessories, including gun mounts, grips, magazines, and ammunition. We also do not allow the advertising of the rental of guns (other than from shooting ranges), stun guns, taser guns, mace, pepper spray or other similar self defense weapons. Additionally, we do not permit the advertising of a variety of weapons including swords, machetes, and other edged/bladed weapons; explosives, bombs and bomb making supplies and/or equipment; fireworks, flamethrowers and other pyrotechnic devices; and knives, including butterfly knives, fighting knives, switchblades, disguised knives, and throwing stars.

We do allow advertising related to the discussion of public policy issues pertaining to firearms. Twitter requires extensive information disclosures of any account involved in political issue advertising and provides specific information to the public via our Ads Transparency Center. Such advertisements are distinctly labeled as political issue promoted tweets. Organizations on both sides of the debate have utilized Twitter's promoted products and continue to do so, within the boundaries of our advertising policies.

III. PARTNERSHIPS AND SOCIETAL ENGAGEMENT

We work closely with the Federal Bureau of Investigation, along with law enforcement and numerous public safety authorities around the world. As our partnerships deepen, we are able to better respond to the changing threats we all face, sharing valuable information and promptly responding to valid legal requests for information.

A. Cooperation with Law Enforcement

We have well-established relationships with law enforcement agencies, and we look forward to continued cooperation with them on these issues, as often they have access to information critical to our joint efforts to stop bad faith actors. The threat we face requires extensive partnership and collaboration with our government partners and industry peers. We each possess information the other does not have, and our combined information is more powerful in combating these threats together. We have continuous coverage to address reports from law enforcement around the world and have a portal to swiftly handle law enforcement requests rendered by appropriate legal process.

Twitter informs individuals using the platform that we may preserve, use, or disclose an individual's personal data if we believe that it is reasonably necessary to comply with a law, regulation, legal process, or governmental request; to protect the safety of any person; to protect the safety or integrity of our platform, including to help prevent spam, abuse, or malicious actors on our services, or to explain why we have removed content or accounts from our services; to address fraud, security, or technical issues; or to protect our rights or property or the rights or property of those who use our services.

Twitter retains different types of information for different time periods, and in accordance with our Terms of Service and Privacy Policy. Given Twitter's real-time nature, some information (*e.g.,* Internet Protocol logs) may only be stored for a very brief period of time.

Some information we store is automatically collected, while other information is provided at the user's discretion. Though we do store this information, we cannot guarantee its accuracy. For example, the user may have created a fake or anonymous profile. Twitter doesn't require real name use, e-mail verification, or identity authentication.

Once an account has been deactivated, there is a very brief period in which we may be able to access account information, including Tweets. Content removed by account holders (*e.g.,* Tweets) is generally not available.

Twitter accepts requests from law enforcement to preserve records, which constitute potentially relevant evidence in legal proceedings. We will preserve, but not disclose, a temporary snapshot of the relevant account records for 90 days pending service of valid legal process.

Twitter may honor requests for extensions of preservation requests, but encourage law enforcement agencies to seek records through the appropriate channels in a timely manner, as we cannot always guarantee that requested information will be available.

Our biannual Twitter Transparency Report highlights trends in enforcement of our Rules, legal requests, intellectual property-related requests, and e-mail privacy best practices. The report also provides insight into whether or not we take action on these requests. The Transparency Report includes information requests from governments worldwide and non-government legal requests we have received for account information. In 2018, we received 4,323 requests from United States authorities, relating to 13,086 accounts,

B. Industry Collaboration

Collaboration with our industry peers and civil society is also critically important to addressing common threats from terrorism globally. In June 2017, we launched the Global Internet Forum to Counter Terrorism (the "GIFCT"), a partnership among Twitter, YouTube, Facebook, and Microsoft.

The GIFCT facilitates, among other things, information sharing; technical cooperation; and research collaboration, including with academic institutions. In September 2017, the members of the GIFCT announced a significant financial commitment to support research on terrorist abuse of the Internet and how governments, tech companies, and civil society can respond effectively. Our goal is to establish a network of experts that can develop platform-agnostic research questions and analysis that consider a range of geopolitical contexts.

Technological collaboration is a key part of GIFCT's work. In the first two years of GIFCT, two projects have provided technical resources to support the work of members and smaller companies to remove terrorist content.

First, the shared industry database of "hashes"—unique digital "fingerprints"—for violent terrorist propaganda now spans more than 100,000 hashes. The database allows a company that discovers terrorist content on one of its sites to create a digital fingerprint and share it with the other companies in the forum, who can then use those hashes to identify such content on their services or platforms, review against their respective policies and individual rules, and remove matching content as appropriate or block extremist content before it is posted.

Second, a year ago, Twitter began working with a small group of companies to test a new collaborative system. Because Twitter does not allow files other than photos or short videos to be uploaded, one of the behaviors we saw from those seeking to promote terrorism was to post links to other services where people could access files, longer videos, PDFs, and other materials. Our pilot system allows us to alert other companies when we removed an account or Tweet that linked to material that promoted terrorism hosted on their service. This information sharing ensures the hosting companies can monitor and track similar behavior, taking enforcement action pursuant with their individual policies. This is not a high-tech approach, but it is simple and effective, recognizing the resource constraints of smaller companies.

Based on positive feedback, the partnership has now expanded to 12 companies and we have shared more than 14,000 unique URLs with these services. Every time a piece of content is removed at source, it means any link to that source—wherever it is posted—will no longer be operational.

We are eager to partner with additional companies to expand this project, and we look forward to building on our existing partnerships in the future.

Finally, GIFCT has established a real-time crisis response process that allows us to respond to a violent act quickly to ensure that we share valuable information to limit the spread of terrorist and violent extremist content.

C. The Christchurch Call to Action

In the months since a terrorist attack in Christchurch, New Zealand, New Zealand Prime Minister Jacinda Ardern has led the international policy debate, and that work has culminated in the Christchurch Call. Twitter's Chief Executive Officer Jack Dorsey attended the launch of the Christchurch Call in Paris, meeting with the Prime Minister to express our support and partnership with the New Zealand Government.

Because terrorism cannot be solved by the tech industry alone, the Christchurch Call is a landmark moment and an opportunity to convene governments, industry, and civil society to unite behind our mutual commitment to a safe, secure open, global Internet. It is also a moment to recognize that however or wherever evil manifests itself, it affects us all.

In fulfilling our commitments in the Call, we will take a wide range of actions. We continue to invest in technology to prioritize signals, including user reports, to ensure we can respond as quickly as possible to a potential incident, building on the work we have done to harness proprietary technology to detect and disrupt bad actors proactively.

As part of our commitment to educate users about our rules and to further prohibit the promotion of terrorism or violent extremist groups, we have updated our rules and associated materials to be clearer on where these policies apply. This is accompanied by further data being provided in our transparency report, allowing public consideration of the actions we are taking under our rules, as well as how much content is detected by our proactive efforts.

Twitter will take concrete steps to reduce the risk of livestreaming being abused by terrorists, while recognizing that during a crisis these tools are also used by news organizations, citizens and governments. We are investing in technology and tools to ensure we can act even faster to remove video content and stop it spreading.

Finally we are committed to continuing our partnership with industry peers, expanding on our URL sharing efforts along with wider mentoring efforts, strengthening our new crisis protocol arrangements, and supporting the expansion of GIFCT membership.

D. Partnerships with Civil Society

In tandem with removing content, our wider efforts on countering violent extremism going back to 2015 have focused on bolstering the voices of non-governmental organizations and credible outside groups. These organizations and groups can use our uniquely open service to spread positive and affirmative campaigns that seek to offer an alternative to narratives of hate. Ideologies can only be successfully countered by those who have the credibility to take on the core messages being propagated, and if these core messages go unchallenged the removal of content will always be an incomplete response. These groups do critical work and policy makers should continue to find ways to broaden support for these efforts.

We have partnered with organizations delivering counter and alternative narrative initiatives across the globe and we encourage the Committee to consider the role of government in supporting the work of credible messengers in this space at home and abroad. Twitter has also delivered capacity building workshops to a range of organizations who seek to provide positive, alternative messages and work with communities and individuals at risk.

E. A Whole of Society Response

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, there is no common catalyst to action 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.

We are committed to playing our part. We will continue to seek to proactively remove terrorist and violent extremist content, work with industry peers to respond quickly in a crisis and to support smaller companies in tackling these challenges.

While we strictly enforce our policies, removing all discussion of particular viewpoints, no matter how uncomfortable society may find them, does not eliminate the ideology underpinning them. There is a risk such an approach 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 often not at the forefront of public discussions. We are committed to learning and improving, but every part of the online ecosystem has a part to play.

Furthermore, not every issue will be one where the underlying factors can be addressed by public policy interventions led by technology companies.

* * *

We stand ready to assist the Committee in its important work regarding the issue of the tools that Internet companies can employ to stop the spread of mass violence on our services.

The CHAIRMAN. Thank you very much. Mr. Slater.

STATEMENT OF DEREK SLATER, GLOBAL DIRECTOR, INFORMATION POLICY, GOOGLE LLC

Mr. SLATER. Chairman Wicker, Ranking Member Cantwell, distinguished members of the Committee, thank you for the opportunity to appear before you today. My name is Derek Slater. I am the Global Director of Information Policy at Google. In that capacity, I lead a team that advises the company on public policy frameworks for dealing with online content, including hate speech, extremism, and terrorism. Before I begin, I would like to take a moment on behalf of everyone at Google to express our horror in learning of the tragic attacks in Texas, Ohio and elsewhere, and share our sincere condolences to the affected families, friends, and communities.

All Google services were not involved in these recent incidents. We have engaged with the White House, Congress, and governments around the globe on steps we are taking to ensure that our platforms are not used to support hate speech or incite violence. In my testimony today, I will focus on three key areas where we are making progress to help protect people. First, how we work with governments and law enforcement, second, how our efforts to prohibit the promotion of products that causes damage, harm, or injury, and third, the enforcement of our policies around terrorism and hate speech.

First, Google engages in ongoing dialogue with law enforcement agencies to understand the threat landscape and respond to threats that affect the safety of our users and the broader public. For example, when we have a good faith belief that there is a threat to life or serious bodily harm made on our platform in the United States, the Google cybercrime investigation group will report it to the Northern California Regional Intelligence Center. In turn, that Intelligence Center quickly gets the report into the hands of officers to respond.

The cybercrime investigation group is on call 24/7 to make these reports. We are also deeply committed to working with Government, the tech industry, and experts from civil society and academia. Since 2017, we have done this in particular through the Global Internet Forum to Counter Terrorism of which YouTube is a founded company, and Google was its first chair. Recently, GIFCT introduced joint content incident protocols for responding to emerging or active events. The GIFCT also released its first-ever Transparency Report and a new counter speech campaign toolkit.

Second, we take the threat posed by gun violence in the United States very seriously and our advertising policies have long prohibited the promotion of weapons, ammunition, and similar products that cause damage, harm, or injury. Similarly, we also prohibit the promotion of instructions for making guns, explosives, or other harmful products, and we employ a number of proactive and reactive measures to ensure that our policies are appropriately enforced. We know that we must be vigilant on these issues and are constantly improving our enforcement procedures, including implementing enhancements to our automated systems and updating our incident management and manual review procedures.

Third, on YouTube, we have rigorous policies and programs that defend against the use of our platform to spread hate or incite violence. Over the past two years, we have invested heavily in machines and people to quickly identify and remove content that violates our policies. This includes machine learning technology to effectively enforce our policies at scale, hiring over 10,000 people across Google tasked with detecting or viewing and removing content.

An intel desk of experts that proactively looks for new trends and improves escalation pathway for expert NGOs and governments to notify us about content in bulk through our trusted flagger program, and finally going beyond removals by actively creating programs to promote beneficial counter speech, such as the creators for change program and alphabets jigsaw groups use for a redirect method. This broad, cross sectional work has led to tangible results. Over 87 percent of the 9 million videos we removed in the second quarter of 2019 were first flagged by our automated systems.

More than 80 percent of those auto flagged videos were removed before they received a single view, and overall, videos that violate our policies generate a fraction of a percent of the views on YouTube. Our efforts do not end there. As we are constantly evolving to new challenges and looking for ways to improve our policies. For example, YouTube recently further updated its hate speech policy. The updated policy specifically prohibits videos alleging that a group is superior in order to justify discrimination, segregation, or exclusion based on qualities like age, gender, race, caste, religion, sexual orientation, or veteran status.

It can take months for us to ramp up enforcement of our new policies. We have already seen five times spike and removals and channel terminations on hate speech. In conclusion, we take the safety of our users very seriously and value our close and collaborative relationships with law enforcement and government agencies.

We understand these are difficult issues of great interest to Congress and want to be responsible actors who are part of the solution. As these issues evolve, Google will continue to invest in the people and technology to meet the challenge. We look forward to continued collaboration with the Committee as it examines these issues. Thank you for your time, and I look forward to taking your questions.

[The prepared statement of Mr. Slater follows:]

Prepared Statement of Derek Slater, Director, Information Policy, Google LLC

Chairman Wicker, Ranking Member Cantwell, and distinguished members of the Committee: Thank you for the opportunity to appear before you today. I appreciate Congress' work in looking closely at how to prevent tragic episodes of mass violence.

My name is Derek Slater, and I am the Global Director of Information Policy at Google. In that capacity I lead a team that advises the company on public policy frameworks for dealing with online content—including hate speech, extremism, and terrorism. Prior to my role at Google, I worked on Internet policy at the Electronic Frontier Foundation and at the Berkman Center for Internet and Society.

Before I begin, I would like to take a moment on behalf of everyone at Google to express our horror in learning of the tragic attacks in Texas and Ohio and to share our sincere condolences to the affected families, friends, and communities. While Google services were not involved in these recent incidents, we have engaged with the White House, Congress, and governments around the globe on steps we are taking to ensure that our platforms are not used to support hate speech or incite violence.

We believe the free flow of information and ideas has important social, cultural and economic benefits, though society has always recognized that free speech must be subject to reasonable limits. This is true both online and off, and it is why, in addition to respecting the law, we have additional policies, procedures, and community guidelines that govern what activity is permissible on our platforms.

In my testimony today, I will focus on three key areas where we are making progress to help protect people: (i) how we work with governments and law enforcement; (ii) our efforts to prohibit the promotion of products that cause damage, harm, or injury; and (iii) the enforcement of our policies around terrorism and hate speech.

Working with Government and Law Enforcement

Google appreciates that law enforcement agencies face significant challenges in protecting the public against crime and terrorism. Google engages in ongoing dialogue with law enforcement agencies to understand the threat landscape and respond to threats that affect the safety of our users and the broader public. When we become aware of statements on our platform that constitute a threat to life or that reflect that someone's life may be in danger, we report this activity to law enforcement agencies.

For example, when we have a good faith belief that there is a threat to life or serious bodily harm made on our platform in the United States, the Google CyberCrime Investigation Group (CCIG) will report it to the Northern California Regional Intelligence Center (NCRIC). In turn, NCRIC quickly gets the report into the hands of officers to respond. CCIG is on call 24/7 to make these reports.

Under U.S. law, the Stored Communications Act allows Google and other service providers to voluntarily disclose user data to governmental entities in emergency circumstances where the provider has a good faith belief that disclosing the information will prevent loss of life or serious physical injury to a person. Our team is staffed on a 24/7/365 basis to respond to these emergency disclosure requests (EDRs). We have seen significant growth in the volume of EDRs that we receive from U.S. governmental entities, as illustrated in our *transparency report covering government requests for user data.* In fact, the number of EDRs submitted from agencies in the U.S. almost doubled from 2017 to 2018. We have grown our teams to accommodate this growing volume and to ensure we can quickly respond to emergency situations that implicate public safety.

We are also deeply committed to working with government, the tech industry, and experts from civil society and academia to protect our services from being exploited by bad actors. The recent tragic events in Christchurch presented unique challenges, and we had to take unprecedented steps to address the sheer volume of new videos related to the events. In the months since, Google and YouTube signed the Christchurch Call to Action, a series of commitments to quickly and responsibly address terrorist content online. This is an extension of our ongoing commitment to working with our colleagues in the industry to address the challenges of terrorism online. Since 2017, we've done this through the Global Internet Forum to Counter Terrorism (GIFCT), of which Google is a founding company and was its first chair. Recently, GIFCT introduced joint content incident protocols for responding to emerging or active events. The GIFCT also released its first-ever Transparency Report and a new counterspeech campaign toolkit that will help activists and civil society organizations challenge the voices of extremism online.

Prohibiting the Promotion of Products That May Cause Damage, Harm, or Injury

We take the threat posed by gun violence in the United States very seriously and our advertising policies have long prohibited the promotion of weapons, ammunition, explosive materials, fireworks, and similar products that cause damage, harm, or injury. Similarly, we also prohibit the promotion of instructions for making guns, explosives, or other harmful products.

On platforms like Google Ads and Google Shopping Ads, we employ a number of proactive and reactive measures to ensure that our policies are appropriately enforced. For example, we run automated and manual checks to detect content that violates our policies. If an advertiser or merchant violates our policies, we will take appropriate action up to and including suspension of their account. Users can also provide direct feedback on ads that potentially violate Google policies via an external form using the 'Report a violation' link or via the feedback link on Google.com and other Google properties to report any products that may violate our policies. This feedback is reviewed by our teams and appropriate action is taken.

We know that we must be vigilant on these issues and are constantly improving our enforcement procedures, including implementing enhancements to our automated systems and updating our incident management and manual review procedures.

Policies and Enforcement on YouTube for Terrorism and Hate Speech

We have robust policies and programs to defend our platforms to spread hate or incite violence. This includes prohibitions on: terrorist recruitment, violent extremism, incitement to violence, glorification of violence, and instructional videos related to acts of violence. We apply these policies to violent extremism of all kinds, whether inciting violence on the basis of race or religion or as part of an organized terrorist group.

In order to improve the effectiveness of our policy enforcement, we have invested heavily in both technology and people to quickly identify and remove content that violates our policies against incitement to violence and hate speech:

(1) YouTube's enforcement system starts from the point at which a user uploads a video. If our technology detects that the video is similar to videos that we know already violate our policies, it is sent for humans to review. If they determine that it violates our policies, they remove it and the system makes a "digital fingerprint" or hash of the video so it can't be uploaded again.

(2) Machine learning technology also helps us more effectively identify this content and enforce our policies at scale. However, because hate and violent extremism content is constantly evolving and can sometimes be context-dependent, we also rely on experts to help us identify policy-violating videos. Some of these experts sit at our intel desk, which proactively looks for new trends in content that might violate our policies. We also developed an improved escalation pathway for expert NGOs and governments to notify us of bad content in bulk through our Trusted Flagger program. We reserve the final decision on whether to remove videos they flag, but we benefit immensely from their expertise.

(3) This broad cross-sectional work has led to tangible results. Over 87 percent of the 9 million videos we removed in the second quarter of 2019 were first flagged by our automated systems. More than 80 percent of those auto-flagged videos were removed before they received a single view. And overall, videos that violate our policies generate a fraction of a percent of the views on YouTube.

Our efforts do not end there, as we are constantly evolving to new challenges and looking for ways to improve our policies. For example, YouTube recently updated its Hate Speech policy to specifically prohibit videos alleging that a group is superior in order to justify discrimination, segregation or exclusion based on qualities like age, gender, race, caste, religion, sexual orientation or veteran status. This would include, for example, videos that promote or glorify Nazi ideology, because it is inherently discriminatory. YouTube also updated its policies to prohibit content denying that well-documented violent events, like the Holocaust or the shooting at Sandy Hook Elementary, took place.

The updated Hate Speech policy was launched in early June, and as our teams review and remove more content in line with the new policy, our machine learning algorithms will improve in tandem to help us identify and remove such content. Though it can take months for us to ramp up enforcement of a new policy, the profound impact of our Hate Speech policy update is already evident in the data released in this quarter's *Community Guidelines Enforcement Report:* the number of individual video removals for hate speech saw a 5x spike to over 100,000, the number of channel terminations for hate speech also saw a 5x spike to 17,000, and the total comment removals nearly doubled in Q2 to over 500 million due in part to a large increase in hate speech removals.

Finally, we go beyond removing policy-violating content by actively creating programs to promote beneficial counterspeech. These programs present narratives and elevate credible voices speaking out against hate, violence, and terrorism. For example, our Creators for Change program supports creators who are tackling tough issues, including extremism and hate by building empathy and acting as positive role models. We launched our most recent Creators for Change global campaign videos in November 2018. As of June 2019 they already had 59 million views; the creators involved have over 60 million subscribers and more than 8.5 billion lifetime views of their channels; and through 'Local Chapters' of Creators for Change, creators tackle challenges specific to different markets.

Alphabet's Jigsaw group, an incubator to tackle some of the toughest global security challenges, has deployed the Redirect Method, which uses targeting tools and curated YouTube playlists to disrupt online radicalization. The method is open to anyone to use, and NGOs have sponsored campaigns against a wide-spectrum of ideologically-motivated terrorists and violent extremists.

Conclusion

We take the safety of our users very seriously and value our close and collaborative relationships with law enforcement and government agencies. We have invested substantial resources to tackle the problem of hate speech. At present, we spend hundreds of millions of dollars annually and have more than 10,000 people working across Google to address content that might violate our policies, which include our policies against promoting violence and terrorism.

We understand these are difficult issues of great interest to Congress and want to be responsible actors who are a part of the solution. As these issues evolve, Google will continue to invest in the people and technology to meet the challenge. We look forward to continued collaboration with the Committee as it examines these issues. Thank you for your time. I look forward to taking your questions.

The CHAIRMAN. Thank you very much. Mr. Selim, your group prefers to be known as ADL these days, is that correct?

Mr. SELIM. Correct. The Anti-Defamation League goes by ADL for short.

Mr. CHAIRMAN. Great. Well, we appreciate you being with us today and we are happy to receive your testimony.

STATEMENT OF GEORGE SELIM, SENIOR VICE PRESIDENT, NATIONAL PROGRAMS, ADL (ANTI-DEFAMATION LEAGUE)

Mr. SELIM. Thank you, Mr. Chairman, Ranking Member Cantwell, thank you for the opportunity to be here with the distinguished members of this Committee this morning. My name is George Selim and I serve as the Senior Vice President for Programs at the ADL or the Anti-Defamation League, and for decades the ADL has fought against bigotry and anti-Semitism by exposing extremist groups and individuals who spread hate to incite violence.

Today, the ADL is the foremost non-governmental authority on domestic terrorism, extremism, hate groups, and hate crimes. I have personally served in several roles and the Government's National Security apparatus at the Department of Justice, the Department of Homeland Security, at the White House on the National Security Council, and now outside Government on the frontlines of combating anti-Semitism and all forms of bigotry at the ADL. In my testimony, I would like to share with you some key data, findings, analysis, and urge this Committee to take action to counter a severe national security threat, the threat of online white supremacist extremism that threatens our communities.

The alleged El Paso shooter posted a manifesto to 8chan prior to the attack. He expressed support for the accused shooter in Christchurch, New Zealand who also posted on 8chan. Before the massacre in Poway, California, the alleged shooter posted a link to his manifesto on 8chan, citing the terrorists in New Zealand and in the Pittsburgh Tree of Life attack, three killing sprees, three white supremacist manifestos—one targeted Muslims, another targeted Jews, and a third targeted Latino and other immigrant communities. One thing these three killers had in common was 8chan, an online platform that has become the go-to for many bigots and extremists.

Unfettered access to online platforms, both fringe and mainstream, has significantly driven the scale, speed, and effectiveness of these forms of extremist attacks. Our ADL research shows that domestic extremist violence is trending up, and that anti-Semitic hate is trending up. The FBI and DOJ data shows similar trends. The online environment today amplifies hateful voices worldwide and facilitates the coordination, recruitment, and propaganda that fuels the extremism that terrorizes our communities, all of our communities.

Whether through Government, the private sector, or civil society, immediate action is paramount to prevent the next tragedy that could take innocent lives. ADL has worked with the platforms represented on this table to try to address that hate and its rampant nature online. We have been part of the conversations to improve the terms of service, content moderation programs, and better support for those individuals experiencing hate and harassment on those platforms.

We appreciate this work greatly but much more needs to be done. ADL has called on these companies at this hearing as well as many others to be far more transparent about the prevalence and nature of hate on their platforms. We need meaningful transparency to give actionable information to policymakers and stakeholders, but the growth of hate and extremist violence will not be solved by addressing these issues online alone. We urge this Committee to take immediate action.

First, our Nation's leaders must clearly and forcefully call out bigotry in all its forms at every opportunity. Our Nation's law enforcement leadership must make enforcing hate crimes laws a top priority. Our communities need this Congress's immediate action on a range of ways, notably to codify Federal offices to address domestic terrorism and extremism and create transparent and comprehensive reporting such as that required in the Domestic Terrorism Prevention Act and similar measures in the Domestic Terrorism Data Act. Our Federal legal system currently lacks the means to prosecute a white supremacist terrorist as a terrorist. Congress should explore whether it is possible to craft a rights protecting domestic terrorism statute.

Any statute that Congress should consider would need to include specific, careful Congressional and civil liberties oversight to ensure the spirit of such protections is faithfully executed. In addition, the State Department should examine whether certain foreign, white supremacist groups meet the criteria for designation an FTO, foreign terrorist organizations. For technology and social media companies, we look forward to companies expanding their terms of service and exploring accountability and governance challenges, aspiring to greater transparency in how you address these issues and partnering with civil society groups to help in all of these efforts.

ADL stands ready both with both the Government and the private sector to better address all forms and threats online. This is an all-hands-on-deck moment to protect all of our communities. I look forward to your questions. Mr. Chairman, Ranking Member, and other distinguished members of this Committee. Thank you.

[The prepared statement of Mr. Selim follows:]

PREPARED STATEMENT OF GEORGE SELIM, SENIOR VICE PRESIDENT, NATIONAL PROGRAMS, ADL (ANTI-DEFAMATION LEAGUE)

Introduction

Since 1913, the mission of ADL (Anti-Defamation League) has been to "stop the defamation of the Jewish people and to secure justice and fair treatment to all." For decades, ADL has fought against bigotry and anti-Semitism by exposing extremist groups and individuals who spread hate and incite violence. Today, ADL is the foremost non-governmental authority on domestic terrorism, extremism, hate groups, and hate crimes. ADL plays a leading role in exposing extremist movements and activities, while helping communities and government agencies alike in combating them. ADL's team of experts—analysts, investigators, researchers, and linguists—use cutting-edge technologies and investigative techniques to track and disrupt extremists and extremist movements worldwide. ADL provides law enforcement officials and the public with extensive resources, including analytic reports on extremist trends and databases of Hate Symbols and Terror Symbols that can help alert online platforms of problematic content.

White Supremacy and Mass Shootings [1]

When white supremacist Robert Bowers entered the Tree of Life Synagogue in Pittsburgh in October 2018 to launch a killing spree against Jews attending services, taking 11 lives and wounding seven more, his senseless and hate-fueled violence directly impacted not just the victims' families, friends and neighbors, but all residents of Pittsburgh—and communities nationwide and around the world. The deadliest attack against American Jews, unfortunately, was only one of many in the past year tied to a white supremacist ideology that has found fertile ground online with consequences affecting not only Americans but people around the world. Extremist-related killings are comparatively few when compared to the total number of homicides in the U.S. each year. Nevertheless, such killings, especially when they are committed as hate crimes or terrorist attacks, can send shock waves through entire communities—and beyond. A list of selected white supremacist shooting sprees is included at the end of this document.

Recent analysis by ADL's Center on Extremism shows that domestic extremists took the lives of at least 50 people in 2018, a sharp increase from the 37 people killed by extremists in 2017. In fact, 2018 is the fourth-deadliest year since 1970, behind only 1995 (which saw 184 deaths, most attributed to the Oklahoma City bombing), 2016 (72 deaths) and 2015 (70 deaths).

2018's high death toll is due in large part to the number of shooting sprees by extremists. In 2017, only one extremist-related shooting spree occurred; in 2018, there were five shooting sprees collectively responsible for 38 deaths and 33 wounded. There were fewer lethal incidents in 2018 than in 2017 (17 compared to 21), but the events were significantly deadlier—and the 2018 shooting sprees were responsible for most of the deaths.

These attacks are in large part intensified by the use of guns. In both high-and low-casualty attacks, domestic extremists used guns in 42 of the 50 murders they committed in 2018, far outpacing edged weapons or physical assaults. Over the past

[1] Datasets for this section are available on ADL's HEAT Map: ADL, *ADL H.E.A.T Map,* updated June 19, 2019, *https://www.adl.org/education-and-resources/resource-knowledge-base/adl-heat-map.*

ten years, firearms were used in 73 percent of domestic extremist-related killings in the United States. Guns are the weapon of choice among America's extremist murderers, regardless of their ideology.

White supremacists were responsible for the great majority of extremist-related killings in 2018, which is the case almost every year. Right-wing extremists were responsible for 49 (or 98 percent) of the 50 domestic extremist-related killings in 2018, with white supremacists alone accounting for 39 (or 78 percent) of those murders.

Hate Crimes in America

While most anti-Semitic incidents are not directly perpetrated by extremists or white supremacists, there are important connections between the trends. We found in our annual *Audit of Anti-Semitic Incidents* that in 2018, 249 acts of anti-Semitism (13 percent of the total incidents) were attributable to known extremist groups or individuals inspired by extremist ideology, making it the highest level of anti-Semitic incidents with known connections to extremists or extremist groups since 2004.[2] Of those, 139 incidents were part of fliering campaigns by white supremacist groups. Another 80 were anti-Semitic robocalls allegedly perpetrated by anti-Semitic podcaster Scott Rhodes in support of the candidacy of Patrick Little, an unabashed white supremacist who ran an unsuccessful campaign for U.S. Senate in California.

The Audit also noted spikes at several points during the year. The final three months of the year were unusually active, with 255 incidents in October, 300 in November and 194 in December. The high number in October included 45 propaganda distributions by white supremacists. The incidents in November and December immediately followed the Pittsburgh massacre, which likely drew more attention to anti-Semitic activities. Incidents first spiked in May, when 209 anti-Semitic acts were reported, including 80 anti-Semitic robocalls sent by white supremacists, which targeted Jewish individuals and institutions with harassing messages.

Hate crimes are only an element of the anti-Semitic incidents that we track. The most recent data about hate crimes made available by the FBI is for 2017.[3] The FBI has been tracking and documenting hate crimes reported from federal, state, and local law enforcement officials since 1991 under the Hate Crimes Statistics Act of 1990 (HCSA). Though clearly incomplete, the Bureau's annual HCSA reports provide the best single national snapshot of bias-motivated criminal activity in the United States. The Act has also proven to be a powerful mechanism to confront violent bigotry, increasing public awareness of the problem and sparking improvements in the local response of the criminal justice system to hate violence—since in order to effectively report hate crimes, police officials must be trained to identify and respond to them.

The FBI documented 7,175 hate crimes reported by 16,149 law enforcement agencies across the country—the highest level of participation since the enactment of the HCSA, and a 6 percent increase over 2016 participation of 15,254. Of the 7,175 total incidents:

- Religion-based crimes increased 23 percent, from 1,273 in 2016 to 1,564 in 2017—the second highest number of religion-based crimes ever [only 2001, after 9/11, recorded more—1,828].
- Crimes directed against Jews increased 37 percent—from 684 in 2016 to 938 in 2017. Crimes against Jews and Jewish institutions were slightly more than 13 percent of all reported hate crimes—and 60 percent of the total number of reported religion-based crimes. Every year since 1991, crimes against Jews and Jewish institutions have been between 50 and 80 percent of all religion-based hate crimes.
- Race-based crimes were the most numerous (as they have been every year since 1991), totaling 4,131 crimes, almost 58 percent of the total. Crimes against African-Americans, as always, were the plurality of these crimes—2,013, about 28 percent of all reported hate crimes.
- Reported crimes against Muslims decreased 11 percent, from 307 in 2016 to 273 in 2017. However, the 273 anti-Muslim hate crimes recorded was the highest reported number of crimes against Muslims ever—behind 2016's 307 and 481 in 2001, after the 9/11 terrorist attacks.
- Crimes directed against LGBTQ people increased from 1,076 in 2016 to 1,130 in 2017. Crimes directed against individuals on the basis of their gender iden-

[2] ADL, *2018 Audit of Anti-Semitic Incidents, https://www.adl.org/audit2018,* April 2019.

[3] FBI, *2017 Hate Crime Statistics, 2017 https://ucr.fbi.gov/hate-crime/2017,* November 2018.

tity decreased slightly, from 124 in 2016 to 119 in 2017, slightly less than two percent of all hate crimes.

Importantly, only 2,040 of the 16,149 reporting agencies—less than 13 percent—reported one or more hate crimes to the FBI. That means that about 87 percent of all participating police agencies affirmatively reported zero (0) hate crimes to the FBI (including at least 92 cities over 100,000). And more than 1,000 law enforcement agencies did not report any data to the FBI (including 9 cities over 100,000).

Moreover, we need to remember that these are only reported crimes. Many communities and individuals do not feel comfortable going to law enforcement for a variety of reasons, so there is likely an undercount of hate crimes resulting from unwillingness to report.

The Role of Online Platforms in White Supremacist Violence

The real-world violence of extremists does not emerge from a vacuum. In many cases the hatred that motivates extremist violence, and especially these documented white supremacist murders, is nurtured in online forums such as Gab, 4chan, 8chan, and other platforms.[4]

Extremist groups are empowered by access to the online world; the Internet amplifies the hateful voices of the few to reach millions around the world. The online environment also offers community: while most extremists are unaffiliated with organized groups, online forums allow isolated extremists to become more active and involved in virtual campaigns of ideological recruitment and radicalization. As Internet proficiency and the use of social media are nearly universal, the efforts of terrorist and extremist movements to exploit these technologies and platforms to increase the accessibility of materials that justify and instigate violence are increasing exponentially. Both terrorist and extremist movements, here at home and abroad, use online and mobile platforms to spread their messages and to actively recruit adherents who live in the communities they target.

Individuals can easily find sanction, support, and reinforcement online for their extreme beliefs or actions, and in some cases neatly packaged alongside bomb-making instructions. This enables adherents like violent white supremacist mass shooters such as Bowers to self-radicalize without face-to-face contact with an established terrorist group or cell.

Perhaps the most important contributor to the subculture of white supremacists are the so-called "imageboards," a type of online discussion forum originally created to share images. One of the most prominent is 4chan, a 15-year-old imageboard whose influence extends far beyond the alt right, as a key source of Internet memes. Its/pol subforum is a disturbing site, an anarchic collection of posts that range from relatively innocuous to highly offensive, with most users posting content anonymously.

Due in part to its extremely lax content moderation policies, 4chan has become home to many racists and openly and vocal white supremacists. Some of its imitators, such as 8chan, lean even more towards racism and white supremacy. Parts of Reddit, a popular website that contains a massive collection of subject-oriented discussion threads, also share the "chan" subculture.

ADL has assessed that individuals do not primarily utilize 8chan for sharing hateful images and messages, but they also use it to turn real-world killings into entertainment, canonizing the perpetrators of previous massacres and keeping track of their respective body counts, like scores in a video game.

The current ADL assessment is that at its core, 8chan is a haven for both violent daydreamers and real-life murderers to virtually meet, network and recruit more followers. This intersection poses considerable risk both online and in the physical world.

Patrick Crusius, the alleged El Paso shooter charged with killing 22 people and injuring many more, is believed to have posted a four-page manifesto to 8chan prior to the attack. His justification for the deadly spree was that he was defending his country from "cultural and ethnic replacement brought on by an invasion."[5]

One of the most telling elements of Crusius's post is that in it, he also expressed support for Australian, white supremacist, mass-murderer Brenton Tarrant, the ac-

[4] Anti-Defamation League, " Hatechan: The Hate and Violence-Filled Legacy of 8chan," *ADL Blog,* August 7, 2019, *https://www.adl.org/blog/hatechan-the-hate-and-violence-filled-legacy-of-8chan;* ADL, *Gab and 8chan: Home to Terrorist Plots Hiding in Plain Sight, https://www.adl.org/resources/reports/gab-and-8chan-home-to-terrorist-plots-hiding-in-plain-sight.*

[5] Anti-Defamation League, "Mass Shooting in El Paso: What We Know," *ADL Blog,* August 4, 2019, *https://www.adl.org/blog/mass-shooting-in-el-paso-what-we-know.*

cused shooter in the March 2019 mosque attacks in Christchurch, New Zealand that left 51 people dead.[6]

Like the El Paso shooter, we assess that Tarrant likely turned to 8chan to post what he referred to as a "explanation" for his deadly rampage, providing links to his own manifesto, which he called "The Great Replacement." In it, he fixated on the white supremacist theory that white European society will be overrun by migration from Muslim and African nations.[7]

In his manifesto, Tarrant addressed the 8chan community directly—as if they were co-conspirators—explicitly directing them to "do your part."

Just one month later, someone did. Before his massacre at the Chabad Congregation in Poway, California, the shooter posted a link to his own manifesto on 8chan, offering the same kind of white supremacist tropes and cited the Christchurch and Pittsburgh shooters for inspiring his own deadly attacks.

Three white supremacist manifestos, three killing sprees. One targeted Muslims, another Jews, the third Latinx and immigrants. What these three men had in common was 8chan, the platform for their final messages.

While the most extreme forms of online content normally thrive on platforms like 8chan, Gab, and 4chan, larger social media platforms like Facebook, Twitter, and YouTube must also remain vigilant. Extremists leverage larger mainstream platforms to ensure that the hateful philosophies and messages that begin to germinate on message boards like Gab and 8chan find a new and much larger audience. Twitter's 300 million users and Facebook's 2.4 billion dwarf the hundreds of thousands on 8chan and Gab. Extremists make use of mainstream platforms in specific and strategic ways to exponentially increase their audience while avoiding content moderation activity that Facebook and Twitter use to remove hateful content. These include creating private pages and events, sharing links that directly lead users to extreme content on websites like 8chan and using coded language called "dogwhistles" to imply and spread hateful ideology while attempting to circumvent content moderation systems.

Since the white supremacist rally in Charlottesville in 2017 and subsequent attacks and murders by extremists to date, there have been many well-publicized efforts by the technology and social media companies that run mainstream social platforms and services to stem the tide of hate and extremism online. After Charlottesville, tech companies ranging from large social platforms like Facebook to payment processors like Paypal to cybersecurity services like Cloudflare took action to expel white supremacists from their services. Even so, these same companies and others in this market sector have been forced to repeatedly respond to violent white supremacist activity on their platforms in the past 12 months. The Christchurch video was streamed on Facebook live, leading Facebook to change its livestreaming policy.[8] Paypal provided payment services to the fringe platform Gab, where the Pittsburgh shooter was believed to be radicalized, but cut off its services after the massacre.[9] Cloudflare provided cybersecurity services to 8chan, and publicly cut it off after the site was implicated in the shooting in El Paso (among others).[10] Although it appears that these companies and others took significant action to address white supremacy and hate in 2017 and claim to have continued to do so, ADL assesses that the above-mentioned platforms are still being abused, including today, by people espousing this hateful and violent ideology even two years later.

Scoping the Problem

One of the key drivers of these complicated and at times deadly issues is the size and scale of these platforms. For example, on Twitter approximately 6,000 tweets are posted every second and approximately 500 million tweets are posted every day. If the company's policies and systems operated at 99 percent effectiveness in detecting and responding to violent hate and extremist rhetoric, that would still leave five million tweets unaddressed every day. Imagine that each of those tweets, on the low end, reached just 60 people: those tweets would reach the number of people equal roughly to the population of the United States (330 million people) every day.

[6] Ibid.

[7] Anti-Defamation League, "White Supremacist Terrorist Attacks at Mosques in New Zealand," March 15, 2019, *https://www.adl.org/blog/white-supremacist-terrorist-attack-at-mosques-in-new-zealand.*

[8] "Christchurch Attacks: Facebook Curbs Live Feature," *BBC News,* May 15, 2019, *https://www.bbc.com/news/technology-48276802.*

[9] Adam Smith, "GoDaddy and PayPal Ban Gab After Pittsburgh Shooting," *PCMag,* October 28, 2018, *https://www.pcmag.com/news/364650/godaddy-and-paypal-ban-gab-after-pittsburgh-shooting.*

[10] Matthew Prince, "Terminating Service for 8chan," *Cloudflare,* August 5, 2019, *https://blog.cloudflare.com/terminating-service-for-8chan/.*

The policies and systems of these companies are very likely not operating with a high degree of accuracy, leaving possibly millions of users exposed and impacted by hateful and extreme content every day. As an example, YouTube in June 2019 announced a policy change focusing on prohibiting white nationalist and other extremist content from existing on its platform.[11] In August 2019, an ADL investigation found a number of prominent white nationalists and other forms of hateful extremists still active and easily found on the platform, despite the policy change.[12] Similarly, after Facebook very publicly banned Alex Jones from its platforms in May 2019, Jones was quickly able to shift his operations to another account on the platform.[13] These instances raise alarming questions about the degree to which social media platforms, through their own internal policies and systems, are able to meaningfully detect, assess, and act on hateful content at the global scale their platforms operate.

The U.S. Congress and American public admittedly have limited knowledge of just how well platforms are dealing with the problem of white supremacist extremism. To evaluate their efforts, civil society organizations like ADL can conduct limited external research similar to the manner mentioned above, in which we use the platform information that is publicly available to objectively assess the stated actions and policy implications of a given platform. Or we can look to the platforms' own limited efforts at transparency about their policies and practices. The mainstream social media platforms have several potentially relevant metrics related to the issue of extremism, especially white supremacist extremism, that they share in their regular transparency reports. These differ slightly as described by each platform. The metrics are self-reported by the companies, and there is no way to fully understand the classification of content categories outside of the brief descriptions given by the platforms as part of this reporting.

For example, the platforms provide information related to terrorism. Facebook reported 6.4 million pieces of content related to terrorist propaganda removed from January to March 2019. This may seem meaningful, but it is not a particularly insightful datapoint. Typically, the social media platform companies are only looking at international terrorism from designated groups such as Al Qaeda and ISIS and are not including white supremacist violence and related activity as part of this terrorism classification.

White supremacist content could fall under the category of hate speech or violent content on a platform. Twitter reported 250,806 accounts actioned for hateful conduct and 56,577 accounts actioned for violent threats from July to December 2018. Yet a wide variety of other types of content not associated with extremism or white supremacy might also fall in this category, making it difficult to glean meaningful analysis about white supremacist content from these metrics.

Additionally, when Facebook claims in its transparency report that it took action on four million pieces of hate speech from January to March 2019, it is difficult to understand what this means in context as we do not know how that compares to the level of hate speech reported to them, which communities are impacted by those pieces of content, or whether any of that content is connected with extremist activity on other parts of their platform.

In order to truly assess the problem of hate and extremism on social media platforms, technology companies must provide meaningful transparency with metrics that are agreed upon and verified by trusted third parties, like ADL, and that give actionable information to users, civil society groups, governments, and other stakeholders. Meaningful transparency will allow stakeholders to answer questions such as: "How significant is the problem of white supremacy on this platform?" "Is this platform safe for people who belong to my community?" "Have the actions taken by this company to improve the problem of hate and extremism on their platform had the desired impact?" Until tech platforms take the collective actions to come to the table with external parties and meaningfully address these kinds of questions through their transparency efforts, our ability to understand the extent of the problem of hate and extremism online, or how to meaningfully and systematically address it, will be extremely limited.

[11] Casey Newton, "YouTube Just Banned White Supremacist Content, and Thousands of Channels are About to be Removes," *The Verge,* June 5, 2019, *https://www.theverge.com/2019/6/5/18652576/youtube-supremacist-content-ban-borderline-extremist-terms-of-service.*

[12] Anti-Defamation League, "Despite YouTube Policy Update, Anti-Semitic, White Supremacist Channels Remain," *ADL Blog,* August 15, 2019, *https://www.adl.org/blog/despite-youtube-policy-update-anti-semitic-white-supremacist-channels-remain.*

[13] Craig Timberg, "Alez Jones Banned from Facebook? His videos are still there—and so are his followers," *The Washington Post,* November 5, 2018, *https://beta.washingtonpost.com/technology/2018/11/05/alex-jones-banned-facebook-his-videos-are-still-there-so-are-his-followers/.*

What We Know About Online Hate and Harassment

One way in which ADL has tried to address this gap in knowledge is by conducting a national representative survey on the hate and harassment experienced by Americans online. Our survey found that over half of respondents (53 percent) experienced some type of online harassment; 37 percent of American adults reported experiencing severe harassment (including physical threats, sexual harassment, stalking and sustained harassment), up from 18 percent in 2017.[14]

We also found that identity-based harassment was most common against LGBTQ+ individuals, with 63 percent of LGBTQ+ respondents experiencing harassment because of their sexual orientation. Religious-based harassment was very common against Muslims (35 percent) and, to a lesser extent, Jewish (16 percent) respondents.

Harassment was also common among other minority groups, with race-based harassment affecting 30 percent of Hispanics or Latinos, 27 percent of African-Americans, and 20 percent of Asian-Americans. Finally, women also experienced harassment disproportionately, with gender identity-based harassment affecting 24 percent of female-identified respondents, compared to 15 percent of male-identified.[15]

Hate and harassment are also endemic to online games. Fifty-three percent of the total population of the United States and 64 percent of the online population of the United States plays video games. Following our wider online survey, we surveyed Americans who play online games and found that 74 percent of respondents experienced some form of harassment while playing games online. Sixty-five percent of players experienced some form of severe harassment, including physical threats, stalking, and sustained harassment.[16]

We are also seeing an increase in extremist and white supremacist content within online games and gaming forums. Scholars have observed white supremacist recruiters actively prey on disaffected youth within the gaming community, and use these channels to plant seeds of hate by invoking sentiments of “us versus them.” Our survey found that nearly a quarter of players (23 percent) are exposed to discussions about white supremacist ideology and almost one in ten (9 percent) are exposed to discussions about Holocaust denial in online multiplayer games. These are alarming insights into an industry that has managed to avoid the intense media scrutiny that more traditional social media platforms have experienced.[17]

Online hate and harassment, whether carried out by extremists or simply by those who feel freer to harm others by the distance and anonymity of being online have real-life, sometimes devastating consequences. Our online game survey found that 23 percent of harassed players become less social and 15 percent felt isolated as a result of in-game harassment. One in ten players had depressive or suicidal thoughts as a result of harassment in online multiplayer games, and nearly one in ten took steps to reduce the threat to their physical safety (8 percent).[18] Alarmingly, nearly a third of online multiplayer gamers (29 percent) had been doxed—had their personal information shared with the goal of harassment.[19]

Our wider survey found that among those who had been targeted, or feared being targeted, approximately 38 percent stopped, reduced or changed their activities online, such as posting less often, avoiding certain sites, changing privacy setting, deleting apps, or increasing filtering of content or users. Some 15 percent took steps to reduce risk to their physical safety, such as moving locations, changing their commute, taking a self-defense class, avoiding being alone, or avoiding certain locations.[20]

Our survey also found societal consequences among respondents. More than half (59 percent) said that online hate and harassment were making hate crimes more common, and half said that they are increasing the use of derogatory language. More than one-third (39 percent) thought that online hate and harassment are making young Americans lose faith in the country, and 30 percent believed that they are making it harder to stand up to hate. Some felt less comfortable in their more immediate environments: approximately 22 percent of Americans report that online

[14] ADL, *Online Hate and Harassment: The American Experience,* 2019, *https://www.adl.org/onlineharassment.*

[15] Ibid.

[16] ADL, *Free to Play? Hate, Harassment, and Positive Social Experiences in Online Games,* July 2019, *https://www.adl.org/free-to-play,* page 18.

[17] Ibid, page 7.

[18] Ibid, page 27.

[19] Ibid, page 18.

[20] ADL, *Online Hate and Harassment: The American Experience.*

hate and harassment makes them feel less safe in their community while 18 percent feel that it makes family members trust each other less.[21]

Critically, those surveyed wanted to see private technology companies take action to counter or mitigate online hate and harassment. Eighty-four percent said that platforms should do more, including making it easier for users to filter (81 percent) and report (76 percent) hateful and harassing content. In addition, Americans want companies to label comments and posts that appear to come from automated "bots" rather than people. Finally, a large percentage of respondents were in favor of platforms removing problematic users as well as having outside experts independently assess the amount of hate on a platform.[22]

Over 80 percent of those surveyed wanted government to act by strengthening laws and improving training and resources for police on cyberhate. Strong support exists for these changes regardless of whether an individual has previously experienced online hate and harassment and regardless of political belief. Although respondents identifying as liberal reported even greater agreement with the actions, those identifying as conservatives overwhelmingly supported all the actions as well.[23]

Moving Forward: Policy Recommendations to Counter the Threat

1. *Bully Pulpit* The President, cabinet officials, and Members of Congress must call out bigotry *at every opportunity.* The right to free speech is a core value, but the promotion of hate should be vehemently rejected. Simply put, you cannot say it enough: America is no place for hate.
2. *Enforcement of Existing Laws* The Administration must send loud, clear, and consistent messages that violent bigotry is unacceptable and ensure that the FBI and the Justice Department's Civil Rights Division will enforce relevant Federal laws and vigorously investigate and prosecute hate crimes.
3. *Improve Federal Hate Crime Training and Data Collection* The Department of Justice should incentivize and encourage state and local law enforcement agencies to more comprehensively collect and report hate crimes data to the FBI, with special attention devoted to large underreporting law enforcement agencies that either have not participated in the FBI Hate Crime Statistics Act program at all or have affirmatively and not credibly reported zero hate crimes. More comprehensive, complete hate crime reporting can deter hate violence and advance police-community relations. In addition, the administration, DHS and DOJ should take steps to ensure that it is efficient and safe for all victims of hate crimes to contact the police. If marginalized or targeted community members—including immigrants, people with disabilities, LGBTQ community members, Muslims, Arabs, Middle Easterners, South Asians and people with limited language proficiency—cannot report, or do not feel safe reporting hate crimes, law enforcement cannot effectively address these crimes, thereby jeopardizing the safety of all.
4. *Legislation to Address White Supremacy and Domestic Terrorism* Congress must act to counter the threat of domestic terrorism and prevent more attacks. No legislative action is perfect, but inaction should not be an option. Congress should enact the following measures:
   - *Domestic Terrorism Prevention Act (DTPA) (S. 894/HR 1931)* This legislation would enhance the Federal government's efforts to prevent domestic terrorism by authorizing into law the offices addressing domestic terrorism, and would require Federal law enforcement agencies to regularly assess those threats. The bill would also provide training and resources to assist non-federal law enforcement in addressing these threats, requiring DOJ, DHS, and the FBI to provide training and resources to assist state, local, and tribal law enforcement in understanding, detecting, deterring, and investigating acts of domestic terrorism.
   - *Domestic Terrorism Documentation and Analysis of Threats in America (DATA) Act (HR 3106)* Data on extremism and domestic terrorism is being collected by the FBI, but not enough, and the reporting is insufficient and flawed. Data drives policy; we cannot address what we are not measuring. The DATA Act focuses on increasing the coordination, accountability, and transparency of the Federal government in collecting and recording data on domestic terrorism.

[21] Ibid.
[22] Ibid.
[23] Ibid.

- ○ *The Khalid Jabara and Heather Heyer National Opposition to Hate, Assault, and Threats to Equality Act of 2019 (NO HATE Act of 2019 S. 2043/H.R. 3545)* This legislation would authorize incentive grants to spark improved local and state hate crime training, prevention, best practices, and data collection initiatives—including grants for state hate crime reporting hotlines to direct individuals to local law enforcement and support services.
- ○ *Disarm Hate Act (S. 1462/H.R. 2708)* This legislation would close the loophole that currently permits the sale of firearms to individuals who have been convicted of threatening a person based on their race, religion, gender, sexual orientation, or disability. The measure would prohibit individuals convicted of a misdemeanor hate crime from obtaining a firearm.

In addition, more consideration is needed for two additional initiatives that could help address white supremacy and domestic terrorism in the United States.

- Congress should examine whether a rights-protecting domestic terrorism criminal charge is needed—and could be appropriately crafted. Our Federal legal system currently lacks the means to prosecute a white supremacist terrorist as a terrorist. Perpetrators can be prosecuted for weapons charges, acts of violence (including murder), racketeering, hate crimes, or other criminal violations. But we cannot legally prosecute them for what they are: terrorists. Many experts have argued that, without being so empowered, there is a danger that would-be domestic terrorists are more likely to be charged with lesser crimes and subsequently receive lesser sentences. Congress should begin immediate hearings and consultations with legal and policy experts, marginalized communities, and law enforcement professionals on whether it is possible to craft a rights-protecting domestic terrorism statute. Any statute Congress would seriously consider should include specific, careful Congressional and civil liberties oversight to ensure the spirit of such protections are faithfully executed.
- The State Department should examine whether certain white supremacist groups operating abroad meet the specific criteria to be subject to sanctions under its Designated Foreign Terrorist Organization (FTO) authority. The criteria, set out in 8 U.S.C. § 1189(a)[1] are: (1) the organization must be foreign; (2) the organization must engage in terrorist activity or retain the capability and intent to engage in terrorist activity or terrorism; and (3) the terrorist activity or terrorism of the organization must threaten the security of U.S. nationals or the national security of the U.S.
- None of the current 68 organizations on the FTO list is a white supremacist organization.[2] And while the possibility of designating white supremacist organizations under the State Department's FTO authority holds promise, there are some important considerations that must be taken into account.
- First, while several countries have added white supremacist groups to their own designated terrorist lists in recent days—including Canada[3] and England[4]—white supremacist groups do not operate exactly like other FTOs, such as ISIS and al-Qaeda. For example, individual white supremacists that carry out attacks—wherever they are—very rarely receive specific operational instructions from organized white supremacist groups abroad to carry out these attacks.
- These groups generally do not have training camps in Europe or elsewhere where individuals travel to learn tactics and then return home to carry out an attack. Instead, individuals in the United States are typically motivated to act based on their own white supremacist ideology, which primarily stems from domestic sources of inspiration but which can sometimes also stem from inspirational sources abroad—including the violent actions of white supremacists—whether that foreign source is associated with an organization or not. Second, in the United States, unlike in Canada and England, the First Amendment provides unique, broad protection for even the most vile hate speech and propa-

[1] "8 U.S. Code § &#x202F;1189.Designation of foreign terrorist organizations," Cornell Law School Legal Information Institute, accessed September 16, 2019; (*https://www.law.cornell.edu/uscode/text/8/1189)*

[2] State Department, "Foreign Terrorist Organizations," accessed September 16, 2019; (*https://www.state.gov/foreign-terrorist-organizations/)*

[3] Harmeet Kaur, "For the first time, Canada adds white supremacists and neo-Nazi groups to its terror organization list," *CNN,* June 28, 2018, (*https://www.cnn.com/2019/06/27/americas/canada-neo-nazi-terror-organization-list-trnd/index.html*)

[4] Emma Lake, "Terror Crackdown: Which terror groups are banned under UK law and when was National Action added to the list?" *The Sun (UK),* October 26, 2017 (*https://www.thesun.co.uk/news/4569388/banned-terror-groups-uk-national-action)*

ganda. While clearly criminal conduct would not be protected under the First Amendment, a great deal of non-criminal association, speech, and hateful propaganda would be protected speech. The First Amendment's assembly and speech protections would not permit designation of white supremacist organizations operating here, but designating *foreign* white supremacist groups could make knowingly providing material support or resources to them a crime—extending authority for law enforcement officials to investigate whether such a crime is being planned or is occurring.[5]

5. Address Online Hate and Harassment

- *Strengthen laws against perpetrators of online hate* Hate and harassment translate from real-world to online spaces, including in social media and games, but our laws have not kept up. Many forms of severe online misconduct are not consistently covered by cybercrime, harassment, stalking and hate crime law. Congress has an opportunity to lead the fight against cyberhate by increasing protections for targets as well as penalties for perpetrators of online misconduct. Some actions Congress can take include revising Federal law to allow for penalty enhancements based on cyber-related conduct; updating Federal stalking and harassment statutes' intent requirement to account for online behavior; and legislating specifically on cybercrimes such as doxing, swatting, non-consensual pornography, and deepfakes.
- *Urge social media platforms to institute robust governance* Government officials have an important role to play in encouraging social media platforms to institute robust and verifiable industry-wide self-governance. This could take many forms, including Congressional oversight or passing laws that require certain levels of transparency and auditing. The Internet plays a vital role in allowing for innovation and democratizing trends, and that should be preserved. At the same time the ability to use it for hateful and severely harmful conduct needs to be effectively addressed.
- *Improve training of law enforcement* Law enforcement is a key responder to online hate, especially in cases when users feel they are in imminent danger. Increasing resources and training for these departments is critical to ensure they can effectively investigate and prosecute cyber cases and that targets know they will be supported if they contact law enforcement.

6. Platform Responsibility to Address Online Hate and Harassment

- *Terms of Service* Every social media and online game platform must have clear terms of service that address hateful content and harassing behavior, and clearly define consequences for violations. These policies should state that the platform will not tolerate hateful content or behavior based on protected characteristics. They should prohibit abusive tactics such as harassment, doxing and swatting. Platforms should also note what the process of appeal is for users who feel their content was flagged as hateful or abusive in error.
- *Responsibility and Accountability* Social media and online game platforms should assume greater responsibility to enforce their policies and to do so accurately at scale. They should improve the complaint and flagging process so that it provides a more consistent and speedy resolution for targets. They should lessen the burden of the complaint process for users, and instead proactively, swiftly, and continuously addressing hateful content using a mix of artificial intelligence and humans who are fluent in the relevant language and knowledgeable in the social and cultural context of the relevant community.

  Additionally, given the prevalence of online hate and harassment, platforms should offer far more services and tools for individuals facing or fearing online attack. They should provide greater filtering options that allow individuals to decide for themselves how much they want to see likely hateful comments. They should consider the experience of individuals who are being harassed in a coordinated way, and be able to provide aid to these individuals in meaningful ways. They should allow users to speak to a person as part of the complaint process in certain, clearly defined cases. They should provide user-friendly tools to help targets preserve evidence and report problems to law enforcement and companies.

[5] Mary B. McCord, "White Nationalist Killers Are Terrorists. We Should Fight Them Like Terrorists," *Washington Post,* Aug. 8, 2019, (*https://www.washingtonpost.com/outlook/white-nationalist-killers-are-terrorists-we-should-fight-them-like-terrorists/2019/08/08/3f8b761a-b964-11e9-bad6-609f75bfd97f_story.html*)

- *Governance and Transparency* Perhaps most importantly, social media and online game platforms should adopt robust governance. This should include regularly scheduled external, independent audits so that the public knows the extent of hate and harassment on a given platform. Audits should also allow the public to verify that the company followed through on its stated actions and assess the effectiveness of company efforts over time. Companies should provide information from the audit and elsewhere through more robust transparency reports. Finally, companies should create independent groups of experts from relevant stakeholders, including civil society, academia and journalism, to help provide guidance and oversight of platform policies.

Beyond their own community guidelines, transparency efforts and content moderation policies, features available on social media and online game platforms need to be designed with anti-hate principles in mind. Companies need to conduct a thoughtful design process that puts their users first, and incorporates risk and radicalization factors before, and not after, tragedy strikes. Today, the most popular method of developing technology tools is through a Software Prototyping approach: an industry-wide standard that prompts companies to quickly release a product or feature and iterate on it over time. This approach completely devalues the impact of unintended design consequences. For example, the Christchurch shooter used Facebook's livestreaming feature to share his attack with the world. The feature could have been designed to limit or lock audiences for new or first-time streamers or prevent easy recording of the video.

These kinds of attacks, designed to leverage social media to attract maximum attention and encourage the next attack, force us to reassess the threat of hateful echo chambers like 8chan as well as the exploitable features in mainstream platforms like Facebook—and how they help drive extremist violence.

Conclusion

ADL data clearly and decisively illustrates that hate is rising across America. Hate has found fertile ground on online platforms, which disrupt societal norms, lowering the barrier of entry to peddlers of hate by making it anonymous and virtual. The Internet also gives extremists a platform and amplifies their reach, giving them easy access to each other and to those who might be radicalized.

All technology and social media companies have a responsibility to address this hate, through the tools they use, the guidelines they set, the transparency they offer, their engagement with civil society and the way they design their platforms.

But we cannot solve the scourge of hate in America simply by fixing online platforms. First, everyone who has a bully pulpit must speak out against such hate. We must also look at our education systems, at our law enforcement capacity and training and at our laws. And we must hold perpetrators accountable for the harm that they cause online and off.

Addendum: Ideological Extremist shooting sprees, 2009–2019

The following is a sampling of white supremacist shooting sprees which took place between 2009 and 2019 compiled by ADL's Center on Extremism. More information and statistics about extremist violence of all ideological backgrounds in the U.S. is available at *https://www.adl.org/education-and-resources/resource-knowledge-base/adl-heat-map*

El Paso, Texas, August 2019. White supremacist Patrick Crusius was arrested following one of the deadliest white supremacist attacks in modern U.S. history, a shooting spree at an El Paso Wal-Mart targeting people of perceived Mexican origin or ancestry that left 22 dead and 24 injured.

Gilroy, California, July 2019. Santino Legan opened fire at the Gilroy Garlic Festival killing 3 and injuring 15 before being fatally wounded by police. In an Instagram post, which appears to have been made by Legan, he asked why towns were overcrowded and open space paved over to make room for "hoards [sic] of mestizos and Silicon Valley white tweets." Legan also urged people to read the book Might is Right, by Ragnar Redbeard. Might is Right, or The Survival of the Fittest is a book argues in favor of self-interest and the primacy of the individual. It also attacks Christianity and Judaism, as religions that weaken people; non-Anglo-Saxons, as lesser races; women, as greatly inferior beings compared to men; urban-dwellers, as weak creatures; and the American concept of government based on the notion that all people are created equal.

Poway, California, April 2019. White supremacist John T. Earnest allegedly opened fire at a synagogue in Poway, California, killing one person and injuring three before fleeing. He was reportedly emulating white supremacist Brenton Tarrant's killing spree in New Zealand in March 2019. Shortly after Tarrant's spree, Earnest allegedly set fire to a mosque in Escondido, California, leaving behind graffiti that referenced Brenton Tarrant's attack. People inside the mosque were able to put out the fire. Earnest's connection to the Escondido mosque attack was not known before the Poway attack.

Pittsburgh, Pennsylvania, October 2018. White supremacist Robert Bowers murdered 11 people and injured seven more, including four police officers, during services at the Tree of Life Synagogue. Bowers was a virulent anti-Semite who, among other things, blamed Jews for orchestrating the immigration of non-whites into the United States.

Parkland, Florida, February 2018. Nikolas Cruz launched a deadly shooting spree at his former high school, Marjory Stoneman Douglas High School, killing 17 people and wounding 17 more. According to CNN, Cruz, 19, belonged to a racist Instagram group and hated blacks and Jews, even claiming Jews wanted to destroy the world. Cruz also allegedly referred to women who engaged in interracial relationships as "traitors." A South Florida Sun-Sentinel article reported that Cruz had racist and Nazi symbols on his backpack and that he had etched swastikas onto ammunition magazines left behind at the school after the shooting. However, little evidence has so far emerged to suggest that the MSDHS shooting spree itself was conducted as a white supremacist attack.

Reston, Virginia, December 2017. Accused white supremacist teen Nicholas Giampa allegedly shot and killed his girlfriend's parents after they became upset by his rumored neo-Nazi views. Giampa, was, at the very least, influenced by Atomwaffen and praised Mason's book, Siege, a book based on a collection of newsletters written by neo-Nazi James Mason in the 1980s. Giampa retweeted material from the "Siege Culture" website and at least one Atomwaffen photo. He also admired someone named "Ryan Atomwaffen" for his white supremacist book collection.

Aztec, New Mexico, December 2017. White supremacist David Atchison disguised himself as a student in order to conduct a school shooting at a local high school, where he killed two students before killing himself.

Mesa, Arizona, March 2015. White supremacist Ryan Elliott Giroux killed one and injured five others during a shooting spree in Mesa. The shootings began at a hotel where two people were shot, one fatally. Giroux then went to a nearby restaurant where he shot a woman and stole a car. Other shootings occurred as he tried to evade apprehension.

Charleston, South Carolina, June 2015. White supremacist Dylann Storm Roof conducted a deadly shooting spree at the AME Emanuel Church in Charleston, killing nine people. Roof deliberately targeted the church because its parishioners were African-American; he hoped to incite a "race war" that he thought whites would win. Roof had written a racist and anti-Semitic manifesto prior to carrying out the attack. Both Federal and state authorities charged Roof in connection with the massacre; in January 2017, Roof was convicted of the Federal charges against him and sentenced to death.

Lafayette, Louisiana, July 2015. White supremacist John Russell Houser killed himself after conducting a vicious shooting spree at a movie theater in Lafayette, Louisiana, that left two people dead and nine others injured. Houser, obsessed at the perceived moral decay of the United States, may have chosen the movie theater as his target because it was showing the Amy Schumer movie Trainwreck.

Minneapolis, Minnesota, November 2015. Police arrested Allen "Lance" Scarsella in November 2015 after Scarsella and others travelled to a Black Lives Matter protest in north Minneapolis, where Scarsella opened fire on protesters there, shooting five people, though none fatally. During his trial in early 2017, prosecutors showed jurors text messages in which Scarsella had described his intent to kill black people. Scarsella was convicted of 12 counts of first-degree assault and one count of riot.

Austin, Texas, November 2014. Larry Steve McQuilliams of Austin, Texas, a suspected adherent of the racist and anti-Semitic religious sect known as Christian Identity, launched a shooting attack in downtown Austin, Texas, firing over 100 rounds of ammunition at targets including the Austin Police Department, a Federal court house and the Mexican consulate. According to police reports, McQuilliams had improvised explosive devices, a map of 34 other targets, including churches, and a copy of the Christian Identity-related book Vigilantes of Christendom: The Story of the Phineas Priesthood in his rental van. McQuilliams died at the scene after an Austin police officer shot him at long range.

Overland Park, Kansas, April 2014. Long-time Missouri white supremacist Frazier Glenn Miller launched an attack on Jewish institutions in the greater Kansas City area, opening fire at two institutions in a shooting spree that took the lives of three people, including one child, before police were able to take him into custody. Miller told police and the media that he launched the attacks "for the specific purpose of killing Jews." Prosecutors have indicted Miller on capital murder charges.

Oak Creek, Wisconsin, August 2012. Racist skinhead Wade Michael Page opened fire at a Sikh temple in Oak Creek, Wisconsin, killing six people and wounding four others, including a police officer responding to the shootings. Page killed himself at the scene after being shot by police. Page was a member of the Hammerskins, a racist skinhead group. He also played in the white power bands End Apathy and Definite Hate.

Washington, Oregon, and California, September 2011. White supremacists David Pedersen and Holly Grigsby engaged in a multi-state killing spree that resulted in four murders in three states. The couple murdered Pedersen's father and stepmother in Washington, a white man in Oregon as part of a carjacking, and an African-American male in California as part of another carjacking. In court, Pederson said he targeted the Oregon man because he believed he was Jewish and the Californian man because he was black. After their arrest, the couple admitted they had been headed to Sacramento to find a prominent Jewish person to kill.

Washington, D.C., June 2009. White supremacist James von Brunn attacked the United States Holocaust Memorial Museum in Washington, D.C., entering the facility and opening fire on security guards inside, shooting and killing one of them. Two other security guards returned fire, wounding von Brunn and preventing further deaths. Von Brunn was arrested and charged with murder. He died of natural causes while awaiting trial.

Boston, Massachusetts, January 2009. White supremacist Keith Luke embarked upon a spree of murderous violence against ethnic and religious minorities in the Boston area in early 2009. He raped and shot an African immigrant, and shot and killed her sister, who had tried to help her. Shortly thereafter, he shot and killed a homeless African immigrant. Although he planned to go to a synagogue that evening to kill as many Jews as possible, then commit suicide, police intercepted him before he could do so. Luke fired at police during a chase before he crashed his vehicle. Police subsequently arrested him without incident. Luke was convicted of murder in 2013 and killed himself in prison the following year.

The CHAIRMAN. Thank you, Mr. Selim. To Ms. Bickert, Mr. Pickles and Mr. Slater, on your platforms, how do you define violent content? How do you define extreme content Ms. Bickert?

Ms. BICKERT. Thank you, Mr. Chairman. We will remove any content that celebrates a violent act, and this is a serious physical injury or death of another person. We also will remove any organization that has proclaimed a violent mission or is engaged in acts of violence. We also don't allow anybody who has engaged in organized hate to have a presence on the site, and we remove hate speech. And hate speech we define as an attack on a person based on his or her characteristics, like race, religion, sexual orientation, gender. We list them out in our policies.

The CHAIRMAN. Harder to define extreme than violent, is that correct?

Ms. BICKERT. Yes, and we see different people use that word in different ways. Senator, so what we do is any organization that has proclaimed violent mission or engaged in documented acts of violence, we remove them. It doesn't matter what the reason is for the violence, we just do not allow the violence period.

The CHAIRMAN. Mr. Pickles, what is your platform's definition of extreme?

Mr. PICKLES. So, similar to Facebook. Agree that the word extremism itself is very subjective. And in some context can be a positive thing. People who are extremely active on this issue and itself is not a bad thing. And so we have a three stage test that defines violent extremist groups, and that test is that we identify through their stated purpose publications or actions as extremists, then engage in violence, so they actually may currently be involved in violence presently, or they promote violence as a means to further their cause, and they target civilians.

So we have got that three-stage test of both the ideology and the violence, because we believe that that framing allows us to protect speech and to protect debates but also remove violent extremist from our platform. We then have a broader framework that prohibits, for example, threats of violence, call for harm, and wish of harm against people that is much broader. And again not dependent on ideology.

The CHAIRMAN. Mr. Slater, can you add any nuances to——

Mr. SLATER. Thank you, Chairman. Broadly similar in that we ban designated foreign terrorist organizations from using our platform as well as incitement of violence, glorification of violence, encouragement to violence, and of course hate speech. So broadly similar lines.

The CHAIRMAN. Now, Mr. Selim has suggested that your three platforms need to be more transparent. What do you say to that, Mr. Slater?

Mr. SLATER. Thank you Chairman. And I think transparency is the bedrock of the work we do, particularly around online content and to try and help people understand both what the rules are and how we are enforcing them. It is something we need to continue to get better on. Look forward to working with this Committee, and Mr. Selim and others on that. We have in the last year on YouTube provided our YouTube community guidelines enforcement report, where you can go and see how many videos we have removed in a quarter, for what reasons, which were flagged by machines versus users, and we break that down by violent extremism, hate speech, child safety, and other issues. So I think this is a really key issue and we look forward to continuing to improve.

The CHAIRMAN. Mr. Selim before I ask Ms. Bickert and Mr. Pickles to respond, perhaps you could help them understand how you frankly don't believe they are quite transparent enough at this point.

Mr. SELIM. Mr. Chairman, thank you for your question. To be clear, the point I am making on transparency is to make sure that there are more clearly delineated categories between the point that Mr. Slater was making in terms of what the machines or the algorithms use to remove certain types of content or stop it from going up in the first place and what users on any of these platforms go on to say, like we think this is a violation of the terms of service.

There are degrees of inconsistencies across these platforms that are at the table as well as others. And so to get a holistic picture of what a certain issue may be while individuals may flag versus what some algorithms pull down, there are different consistencies in that. And so when we are asking for transparency, we are really looking for a much more balanced approach in that across all the platforms.

The CHAIRMAN. So, Mr. Pickles, is he touching on something that has a point?

Mr. PICKLES. Yes. Absolutely. I think the balance between particularly for companies who are investing in technology understanding what came down because a person saw it and reported it versus did the content come down because technology found it is very important. We have now published a breakdown of six policy areas and the number of user reports we receive. It is about 11 million reports every year, but 40 percent of the content that we remove, we removed because technology found it, not because of user reports.

The CHAIRMAN. 40 percent?

Mr. PICKLES. Yes. So telling that story in a meaningful way is absolutely a challenge and one that we are certainly investing in.

The CHAIRMAN. What is that percentage in Facebook, Ms. Bickert?

Ms. BICKERT. Mr. Chairman, when it comes to violent content and terror content, more than 99 percent of what we remove is flagged by our technical tools, and we have had a productive——

The CHAIRMAN. The artificial intelligence?

Ms. BICKERT. Some of it is artificial intelligence, some of it is image matching. So known videos where we use a software to reduce that to basically a digital fingerprint, and we are able to stop uploads of that video again. And we have worked with the ADL for years on this, and I think transparency is key. I think we would all agree. We, for the past year and a half, have published not only our detailed implementation guidelines for exactly how we define hate speech and violence, but also reports on exactly how much we are removing in each category, and how much of that like, Mr. Pickles said, how much of it is actually flagged by our technical tools before we get user reports.

The CHAIRMAN. Thank you very much. Senator Cantwell.

Senator CANTWELL. Thank you, Mr. Chairman. Mr. Selim, I think you mentioned 8chan, but what do you think we need to do to monitor incitement on 8chan and other dark websites?

Mr. SELIM. So I think you can really approach this issue from two categories. There are a number of increased measures, some of which I noted in my written statement submitted to this Committee, that these companies as well as others can take to create a greater degree of transparency and standards so that we can have a really accurate measure of the types of hatred and bigotry that exists in the online environment writ at large. As a result of that increased or better data, we can make better policies that apply to content moderation, terms of service, et cetera. So I think really having the good data is a framework for better policies and better applications and content moderation programs.

Senator CANTWELL. So you are saying there is more that they can do? Social media companies, there is more that they can do?

Mr. SELIM. Yes, ma'am. There is much more that they can do.

Senator CANTWELL. I look in your statement, you include auditing and you know third-party evaluation for that transparency as well as you know responsibility, but as I mentioned in my opening statement, basically then drive all of this to a dark web that we have less access to. I am going to get to them and ask them a question, but what more do you think we should be doing together to address the hate that is taking place on these darker websites too?

Mr. SELIM. So a number of measures. I mean, the first is having our public policy be very starting from place where we are victim focused. We know that whether it is Pittsburgh, Poway, El Paso, or any of the number cities that other panelists and members of this committee have mentioned in their statements, we need to start to make measures that combat extremism or domestic terrorism be from preventing other such horrific tragedies. And in order to do that we really need to start from a place that prevents and has a better accounting of hate crimes, bias-motivated crimes, hate related incidents, etc.

And when we start from that place, I think we can make better policy and better programs at the Federal Government, and State and local, and also in the private industry levels as well.

Senator CANTWELL. Well, one of the reasons I am definitely going to be, you know, calling on the Department of Justice to ask what more we can do in this coordination is several years ago Interpol, Microsoft, the others worked on trying to address on an international basis child pornography to better skill law enforcement at policing crime scenes online. And I would assume that the representatives today would be supportive, maybe helpful, maybe even financially helpful in trying to address these crimes as they exist today as hate crimes on the dark side of the web. Is that—do I have any responses from our tech companies here?

Ms. BICKERT. Thank you, Senator Cantwell. This is something that across the industry we have been working on for the past few years in a manner very similar to how the industry came together against child exploitation online. We launched the global GIFCT, the Global Internet Forum to Counter Terrorism, which both of my colleagues referred to as a way of getting industry to create sort of a no-go zone for this terrorist and violent content.

As part of that, we trained hundreds of smaller companies on best practices and we make technology available to them. The reality is for the bigger companies, we often are able to build technical tools that will stop videos at the time of upload. It is much harder for smaller companies, which is why we provide technology to them. We now have 14 companies that are involved in a hash sharing consortium so that we can help even these small companies stop terrorist content at the time of upload.

Senator CANTWELL. Well, I appreciate, and I agree with Mr. Selim. There is more that you can do on your own sites. But setting that aside for a minute, what do you think we should do about 8chan and the dark websites? What are what do you all think we should do?

Ms. BICKERT. I can tell you what we do on Facebook, Senator, which is we ban any link that connects to 8chan pol where these manifestos have appeared. So those manifestos with the El Paso shooting, with Poway, were not available through Facebook.

Senator CANTWELL. I am saying what more do you think in Government and law enforcement working together, besides what you do to address this, anybody else? Mr. Pickles?

Mr. SLATER. Well, I think, to follow up on Mr. Selim's point, I think certainly if this criminal activity is happening on these platforms then a law enforcement response is primary. As I say, add the tools we have in our toolbox related to content and if people are promoting violence against individuals, that is criminal offenses, a law enforcement intervention at that point is something I think should be looked at. And I think if we can strengthen this industry, our cooperation with law enforcement, we can make sure that the information sharing is a strong as it needs to be to support those interventions.

Senator CANTWELL. So, you think we need more law enforcement resources addressing this issue?

Mr. SLATER. I think it is a question of both resources and I think again to follow Mr. Selim's point, there was a paper from George Washington University last week looking at the statutory framework around some of these spaces and if there are opportunities to strengthen them? And in many of the areas Mr. Selim mentioned, and again, I think that is a worthwhile public health policy conversation to have.

Senator CANTWELL. I definitely believe you need more law enforcement resources on this issue, and I look at what progress we made with Interpol and the tech industry fighting on other issues. I think this is something, and I hear that from Mr. Selim, more resources. So, thank you all very much. Mr. Chairman.

The CHAIRMAN. Thank you, Senator Cantwell.

Senator Fischer.

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

Senator FISCHER. Thank you, Mr. Chairman. In June, Senator Thune held a subcommittee hearing on persuasive design, and as we discussed, Facebook, Twitter, and YouTube are engineered to track, capture, and keep our attention, whether it is through predictions of the next video to keep us watching or what content to push, to the top of our news feeds. I think we have to realize that when social media platforms fail to block extremist content online, this content doesn't just slip through the cracks, it is amplified, and it is amplified to a wider audience.

And we saw those effects during the Christchurch shooting. The New Zealand's terrorists Facebook live broadcast was up for an hour, that was confirmed by *The Wall Street Journal,* before it was removed, and it gained thousands of views during that timeframe. Ms. Bickert, how do you concentrate on the increased risk from how your algorithms boost content while gaps still exists in getting dangerous content off the platform? You touched on that a little bit in your response to Senator Wicker, but how are you targeting solutions to address that specific tension that we see?

Ms. BICKERT. Senator, thank you for the question. It is a real area of focus, and there are three things that we are doing. Probably the most significant is technological improvements, which I will come back to in a second. Second is making sure that we are staffed to very quickly review reports that come in. So the Christchurch video, once that was reported to us by law enforcement, we were able to remove it within minutes. That response time is critical to stopping the virality you mentioned.

And finally, partnerships. We have hundreds of safety and civil society organizations that we partner with. So if they are seeing something, they can flag it for us through a special channel. Now, going back to the technology briefly, with the horrific Christchurch video, one of the challenges for us was that our artificial intelligence tools did not spot violence in the video. What we are doing going forward is working with law enforcement agencies, including in the U.S. and the UK, to try to gather videos that could be helpful training data for our technical tools, and that is just one of the many efforts.

We have to try to improve these machine learning technologies so that we can stop the next viral video at the time of upload or the time of creation.

Senator FISCHER. When you talk about working with law enforcement, you said law enforcement contacted you, is that reciprocal? Do you see something show up and then you in turn try to get it to law enforcement as soon as possible so that individuals can be identified? What is the working relationship there?

Ms. BICKERT. Absolutely. Senator. We have a team that is our law enforcement outreach team. Anytime that we identify a credible threat of imminent harm, we will reach out proactively to law enforcement agencies. And we do that regularly. Also when there is some sort of mass violence incident, we reach out to them, even if we have no indication that our service is involved at all, we want to make sure the lines of communication are open. They know how to submit emergency process to us. We respond around the clock in a very timely fashion because we know that every minute is critical in this type of situation. I am a former prosecutor myself and so these things are very personal to me.

Senator FISCHER. I know that the platforms that are represented here today, you have increased your efforts to take down this harmful content, but as we know there are still shortfalls that exist in order to get that response made in not just a timely manner but one that is really going to truly have an effect. Mr. Slater, when it comes to liability, do media platforms—you guys need more skin in the game so that you can ensure better accountability and be able to incentivize some kind of timely solution?

Mr. PICKLES. Thank you, Senator, for the question. I think if you look at the practices that we are all investing in, certainly looking from our perspective, and the way we are getting better over time. The current legal framework strikes a reasonable balance.

In particular, it both provides protection from liability that would go too far that would be overbroad but also acts as a sword not just a shield, empowering us and giving us the legal certainty that we need to invest in these technologies, the people to monitor or detect, review, and remove this sort of violative content. That way the legal framework continues to work well.

Senator FISCHER. Mr. Selim, can you comment on this as well? Do you think there is enough legal motivation for social media platforms to prioritize some kind of solutions out there? I mean, that is what this hearing is about to find the solutions so that we can curb that online hate that I think continues to grow.

Mr. SELIM. When thinking through the issues of content moderation, the authorities that exist within the current legal frameworks that reside within the companies represented at this table is sufficient for them to take actions on issues of content moderation, transparency reporting, etc. So there certainly is a degree of legal authorities that affords these companies as well as others the opportunity to take any number of measures.

Senator FISCHER. Ms. Bickert, in your testimony you say that Facebook live will ban a user for 30 days for first-time violation of its platform policies. Is that enough? Can users be banned permanently? Would that be something to look at?

Ms. BICKERT. Senator, thank you for the question. One serious violation will lead to a temporary removal of the ability to use live. However, if we see repeated serious violations, we simply take that person's accounts away, and that is something that we do across the board not just with hate and inciting content, but other content as well.

Senator FISCHER. Thank you.

The CHAIRMAN. Thank you so much, Senator Fisher.

Senator Blumenthal.

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

Senator BLUMENTHAL. Thank you, Mr. Chairman. Thank you all for being here today and thank you for outlining the increased attention and intensity of effort that you are providing to this very profoundly significant area. I welcome that you are doing more and trying to do it better, but I would suggest that even more needs to be done and it needs to be better, and you have the resources and technological capabilities to do more and better.

And just to take the question that Senator Fischer asked of you, Mr. Selim, about incentives. Your answer was that they have authority to provide them with opportunities. The question is, really don't they need more incentives to do more and do it better, to prevent this kind of mass violence that may be spurred by hate speech appearing on the site or in fact may actually be a signal of violence to come?

And I just want to highlight that 80 percent of all perpetrators of mass violence provide clear signals and signs that they are about to kill people. That is the reason that Senator Graham and I have a bipartisan measure to provide incentives to more states to adopt extreme risk protection order laws that will, in fact, give law enforcement the information they need to take guns away from people who are dangerous to themselves or others.

And that information is so critically important to prevent mass violence, but also suicides, domestic violence, and the keys and information and signals often appear on the internet. In fact just this past December in Monroe, Washington a clearly troubled young man made a series of anti-Semitic rants and violent posts online. He bragged about planning to "shoot up an expletive school" in a video while armed with an AR–15 style weapon, and on Facebook posted that he was "shooting for 30 Jews."

Fortunately, the ADL saw that post, it went to the FBI, and the ADL's vigilance prevented another Parkland or Tree of Life attack. Fred Gutenberg of Coral Springs, Florida met with me yesterday, told me about a similar incident involving a young man in Coral Springs who said he was about to shoot up the high school there, and law enforcement was able to foresaw it using an extreme risk protection order statute.

So my question is to Facebook, Twitter, and Google, what more can you do to make sure that these kinds of signs and signals involving references to guns, it may not be hate speech, but it is references to possible violence with guns or use of guns, to make that available to law enforcement? Ms. Bickert, and Mr. Pickles, and Mr. Slater.

Ms. BICKERT. Thank you, Senator Blumenthal. One of the biggest things we can do is engage with law enforcement to find out what is working in our relationship and what isn't, and that is the dialogue that over the past years has led to us establishing a portal through which they can electronically submit request for content with legal process and we can respond very quickly——

Senator BLUMENTHAL. But what are you doing proactively? And I apologize for interrupting, but my time is limited. Proactively, what are you doing with the technology you have to identify the signs and signals that somebody is about to use a gun in a dangerous way? That someone is dangerous to himself or others and is about to use a gun?

Ms. BICKERT. Senator, we are now using technology to try to identify any of those early signs, including gun violence, but also suicide or self-injury.

Senator BLUMENTHAL. Do you report it to law enforcement?

Ms. BICKERT. We do. In 2018, we referred a number of many cases of suicide or self-injury, but we detected them using artificial intelligence to law enforcement so that they were able to then intervene, and in many cases, save lives.

Mr. PICKLES. We have is a very similar approach where we have a credible threat that something, someone is at risk to others or themselves. We work with the FBI to ensure they have the information they need.

Senator BLUMENTHAL. Mr. Slater?

Mr. SLATER. Thank you, Senator. Similarly, when we have a good faith belief of a credible threat, then we will proactively refer to the Northern California Regional Intelligence Center who will then fan that out to the right authorities.

Senator BLUMENTHAL. Because my time has expired, I am going to ask each of you if you would please give me more details in writing as a follow up for how you—what identification signs you use, what kind of technology, and how you think it can be improved assuming that the Congress approves, as I hope it will, the emergency risk protection order statute to provide incentives to more than just the 18 states that have them now, but others to do the same. Thank you.

The CHAIRMAN. Thank you so much, Senator Blumenthal.

Senator Thune.

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

Senator THUNE. Thank you, Mr. Chairman. And thank all of you for being here today. Your participation in this hearing is appreciated as this Committee continues its oversight of the difficult tasks each of your companies face preserving an openness on your platforms while seeking to responsibly manage and thwart the actions of those who use your services to spread extremist and violent content. Last Congress, we held a hearing looking at terrorist recruitment propaganda online.

We discussed the cross sharing of information between Facebook, Microsoft, Twitter, and YouTube which allowed each of those companies to identify potential extremism faster and more efficiently. So I would just direct this question and ask that how effective is that shared database of hashes been?

Ms. BICKERT. Senator, thank you for the question. Through the shared data base, we now have more than 200,000 distinct hashes of terror propaganda, and that has allowed—I can speak for Facebook only, but that has allowed us at Facebook to remove a lot more than we otherwise would have been able to do.

Mr. PICKLES. I would just add, since that hearing actually, I think the reassuring thing is that we don't just share hashes now. We have grown that partnership, so we share URLs. So if we see a link to a piece of content like a manifesto, we are able to share that across industry. And furthermore, I think an area that after Christchurch we recognize we need to improve, we now have real-time communications in a crisis.

So industry can talk to each other in real time, operationally to say, even you know, not content related but situational awareness, that partnership between industry now also involves law enforcement. That wasn't there when I think we had that hearing last, and so I think it not just about the hash program but broadening our new programs that are developing that work further.

Mr. SLATER. Yes, I think broadly, I would say look at how we have been improving over time. Surely systems are not perfect. We are always going to have to evolve to deal with bad actors, but I think on the whole, we are doing a better job in part because of this technology sharing, this information sharing, in removing the sort of content before it has wide exposure of any sort or is viewed widely.

Mr. SELIM. Senator, I would only add that the threat environment that we are in today as a country has changed and evolved in the past 24 to 36 months. And likewise, the tactics and techniques that these platforms as well as others use to evolve, the evolving nature of the terrorist landscape online, whether it be foreign or domestic, needs to keep pace with the threat environment that we are in today.

Senator THUNE. And so just as a follow-up, are there similar partnerships among your companies as well as the smaller platforms to specifically identify mass violence?

Ms. BICKERT. Senator, one of the things that we have done over time is expand the mandate of the Global Internet Forum to Counter Terrorism. So we relatively recently expanded to include mass violent incidents, and we are now sharing both through our crisis incident protocol and our hash sharing, we are sharing a broader variety of violent incidents.

Senator THUNE. Mr. Slater, YouTube's, I should say, automated recommendation systems comes under criticism for potentially steering users toward increasingly violent content, and earlier this year, I led a subcommittee hearing examining the use of persuasive technologies on Internet platforms, algorithm transparency, and algorithmic content selection. I asked the witness that Google provided at that time for that hearing several specific questions for the record about YouTube that were not thoroughly answered, and I would just say that providing complete answers to questions members submit for the record is essential as we look to work together as partners to combat many of the issues discussed here today.

So I would like your commitment to provide thorough responses to any questions you might get for the record. Do I have that?

Mr. SLATER. Surely, Senator, to the best of our ability.

Senator THUNE. OK. In addition, I would like to just explore the nexus between persuasive technologies and today's topic, specifically what percentage of YouTube video views are the result of YouTube automatically suggesting or playing another video after the user finishes watching the video?

Mr. SLATER. So I do not have a specific statistic there, but I can say the purpose of our watch next, our recommendation system, is to show people videos they may like that are similar to what they have watched before. At the same time we do recognize this concern about recommendations for borderline content that is content that maybe is not removed but brushes right up against those lines. And we have introduced changes this year to reduce recommendations for those sort of borderline videos.

Senator THUNE. Could you get the number? And I assume you have that somewhere. That has got to be available and furnished for the record, but so the question again is to ask you specifically what is YouTube doing to address the risk that some of these features which as, you note, are pointing a user in the direction of increasingly violent content?

Mr. SLATER. Yes, and that change we made in January to reduce recommendations has been key. And it is still in its early days, but it is working. We have reduced the views from those recommendations for that borderline content by 50 percent just since January. As those systems get better, we hope that that will improve and happy to discuss it further.

Senator THUNE. Thank you. Thank you, Mr. Chairman.

The CHAIRMAN. Thank you, Senator Thune. Now based on presence at the gavel, we next have Senator Blackburn followed by Senator Scott.

Senator Blackburn.

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 being here this morning and for talking with us. This Committee has looked at this issue on the algorithms and their utilization for some time and we are going to continue to do this. Looking at content and the extremists content that is online is certainly important. We know there are a host of solutions that are out there, and we need to come to an agreement and an understanding of how you are going to use these technologies to really protect our citizens.

And social media companies are in a sense open public forums, and they should be where people can interact with one another. And part of your responsibility in this vein is to have an objective cop on the beat and be able to see what is happening because you are looking at it in real time. But what has unfortunately happened many times is you don't get an objective view, you don't get a consistent view, you get a subjective view. And this is problematic, and it leads to confusion by the public that is using the virtual space for entertainment, for their transactional life, for obtaining their news.

So indeed as we look at this issue, we are looking for you to approach it in a consistent and objective manner. And we welcome the opportunity to visit with you today. Ms. Bickert, I have got a couple of things that I wanted to talk with you about. We have all heard about these third-party facilities where contractors are working long hours and they are looking at grotesque and violent images, and they are doing this day in and day out. So talk a little bit about how you transition from that to using modern technologies.

What Facebook is going to do in order to capture this, to extract it and to minimize harm. You have talked about you have got 30,000 employees that are working on safety and security, and then there are third-party entities that are working on this. So let's talk about that impact on the individuals and then talk about the use of technologies to speed up this process and to make it more consistent and accurate.

Ms. BICKERT. Thank you for the question, Senator. Making sure that we are enforcing our policies is a priority for us, making sure that our content reviewers are healthy and safe in their jobs is paramount. And so one of the things that we do is we make sure that we are using technology to make their jobs easier and to limit the amount of content, types of content that they have to see. I will give you a couple examples with child exploitation videos, with graphic violence, with terror propaganda. We are now able to use technology to review a lot of that content so that people don't have to. And in situations where——

Senator BLACKBURN. Let me ask you this, I am sorry to interrupt, but we need to move forward, your 30,000 reviewers, are they all located in Palo Alto or are they scattered around the country, or around the globe?

Ms. BICKERT. No Senator, the more than 50—we have 30,000 people working in safety and security. Some of them are engineers or lawyers. The content reviewers, we have more than 15,000. They are based around the world.

Senator BLACKBURN. OK. Yes, great.

Ms. BICKERT. And for any of them, not only are we using technology, and there are ways that we are using even where we cannot make a decision on the content using technology alone, there are things we can do like removing the volume or separating a video into still frames, that can make the experience better for the reviewer.

Senator BLACKBURN. OK. Now, let me ask you about this. Mark Zuckerberg in a *Washington Post* op-ed had called for us to regulate, to define "lawful but awful" speech. So tell me how you think you could define, or we could define lawful but awful speech but not overreach or infringe on somebody's First Amendment, free speech rights?

Ms. BICKERT. Senator, one of the things that we are looking to with our dialogue with Government is clarity on the actions that Government wants us to take. So we have our set of policies that lays out very clearly how we define things, but we don't do that in a vacuum. We do that with a lot of input from civil society organizations and academics around the world but we also like to hear the views from governments so we can make sure we are mindful of all of the different safety——

Senator BLACKBURN. No, ours are constitutionally based. I am out of time. Mr. Pickles, I am going to submit a question to you for the record. Mr. Selim, I have got one that I am going to send to you. Mr. Slater, I always have questions for Google, so you can depend on me to get one to you and we do hope that you all are addressing your prioritization issues also. With that, Mr. Chairman, I yield back.

The CHAIRMAN. Thank you very much.

Senator Scott.

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

Senator SCOTT. Thank you for being here today. I am glad we are having a meaningful conversation about what is happening in our Nation. It is time we face the fact that our culture has produced an underclass of predominantly white young men who place no value on human life. These individuals live purposeless lives of anonymity and digital dependency, and increasingly act on their most evil desires, sometimes with racial hatred. As you all know, while I was Governor, we had the horrible shooting at the school in Parkland.

Within three weeks we passed historic legislation, including the risk protection orders that Senator Blumenthal was talking about. We did it by sitting down with law enforcement, mental health counselors, and educators to come up with the right solution. Now with regard to the shooting at Parkland, the killer, Nicholas Cruz, had a long, long history of violent behavior. In September 2017, the FBI learned that someone with the username Nicholas Cruz had posted a comment on a YouTube video that said, "I am going to be a professional school shooter."

And Nicholas Cruz made other threatening comments on various platforms. The individual whose video Nicholas Cruz posted this comment on reported it to the FBI. Unfortunately, the FBI closed the investigation after 16 days without ever contacting Nicholas Cruz. The FBI claimed they were unable to identify the person who made the comment. Unfortunately, we now have 17 innocent lives that were lost because of Nicholas Cruz.

My question is to Mr. Slater: How was it a platform like YouTube which is owned by Google not able to track down the IP address and identity of the person who made that comment? When did YouTube remove the comment? Did YouTube report this comment to law enforcement? If so, who and when? If you did report this comment to law enforcement, did you follow-up? What was the process, and was there any follow up to see if there was any corrective action?

Mr. SLATER. Senator, thank you for the question. First, it was a horrendous event. And you know, we strive to be vigilant, to invest heavily, to proactively report where we see an imminent threat. I don't have the details on the specific facts you are describing. I will be happy to get back to you, but let me say this going forward, looking ahead, Parkland was a moment that did spur us to proactively reach out to law enforcement to start talking about, how can we do this better?

And that is part of how we then reached out and started working more closely with the Northern California Regional Intelligence Center to make sure that when we did have these good faith beliefs, we could go to a one-stop shop who could get it to the right law enforcement, locally rather than us trying to call the right people. And this is something we are just this month in fact, or in the last month, there was an incident where PBS was streaming the NewsHour on YouTube, somebody put a threat in the live chat.

We refer that to the Regional Intelligence Center, and they refer it to the Orlando Police who then took the person into custody appropriately. And this was reported in the news. So that is not to say things are perfect. We always have to strive to get better and I look forward to working with you and law enforcement on that. But I do think that we continue to improve over time.

Senator SCOTT. So with regard to Nicholas Cruz, you will give me the information of, you know, who did you contact, when did you contact, when was it taken down? So to this day I cannot get an answer on what anybody did with regard to this shooter. What YouTube did, what the FBI did, nobody wants to talk about it, which is fascinating to me. So if you give me that information.

And then second, are you comfortable that if another Nicholas Cruz put something up, you have the process now that you will contact somebody and there will be a follow-up process?

Mr. SLATER. Senator, I think our processes are getting better all the time. They are robust. I think this is an area where it is an evolving challenge, both because technology evolves, because people's tactics evolve. They might use code words, and so on, but I would be happy to follow up with the team and get more information on how those practices operate and how we continue to work together.

Senator SCOTT. Thank you. Mr. Pickles, how can Nicolas Maduro, who is committing genocide against his citizens, who is withholding clean water, food, and medicine still have a Twitter account with 3.7 million followers?

Mr. PICKLES. Well, you rightly highlight that the behavior that is being taken there is abhorrent and the question for us, as a public company that provides a public space for dialogue is, is someone breaking our rules on our service? We recognize that there are situations where there are geopolitical circumstances where there are world leaders who have Twitter accounts in countries where Twitter has blocked, where there is no free speech, and so we do take a view that and we hope that the dialogue that that person being on the platform starts, helps contribute to solving the challenges that you have outlined.

Senator SCOTT. But he has been doing it for a long time and it is not getting better in Venezuela, it is getting worse.

Mr. PICKLES. And I think this is a good illustration of how the role technology companies along with other parts of public policy responses. And If we remove that person's account, it would not change the facts on the ground. And so we need to bear in mind how did the other levers come into play.

Senator SCOTT. I completely disagree. Maduro sits there and talks about things and continues to act like he is a world leader, and he is a pariah. And it sure seems to me that what you are doing is allowing him to continue to do that.

Mr. PICKLES. Well, as I said, his current account has not broken all the rules. Were he to break all rules, he would be treated the same as every other user, and we would take action when necessary.

Senator CANTWELL. Mr. Chairman, I know that we have votes already starting and you are trying to get other people. I would be happy to work with the Senator from Florida on this issue. I do think that we are not doing enough, and I think this specific case I mentioned in my opening statement about the Rohingya and what happened on Facebook is another example, so happy to work with you on this issue.

The CHAIRMAN. Well, yes, and thank you, Senator Cantwell, and thank you Senator Scott for raising this. I am told there is a vote on, and I am shocked to hear that they are going to leave it open till 11:30 a.m., which is generally what happens.

Senator Duckworth.

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

Senator DUCKWORTH. Thank you, Mr. Chairman. While I do appreciate this Committee's consideration of issues at the intersection of extremism and social media, many I think would agree that today's hearing is another data point on a long history of congressional hand-wringing on gun violence.

According to the gun violence archive, since 2019 began, 260 days ago, we have witnessed 318 mass shootings in the U.S., more than one per day. Mass shootings are those in which at least four people are shot, excluding the shooter. After 20 children, 6 adults, and a shooter lost their lives at Sandy Hook Elementary School in 2012, many elected officials including myself declared an end to Congressional inaction. No more we said, but since that day, our Nation has endured 2,226 mass shootings. Think about that number for a minute. But here we are not focused on ways to stop gun violence, but rather the scourge of social media.

I am not going to say that there is no connection but every other country on the planet has social media, video games, online harassment, hate groups, crime, and mental health issues, but they do not have mass shootings like we do. Nothing highlights the absurdity of Congress's inability to solve the gun violence crisis than seeing 318 mass shootings in 260 days, and then holding our hearings on extremism and social media. Ms. Bickert and Mr. Pickles, this is a chart from the Digital Marketing Institute that according to their website highlights the average number of hours that social media users spend on platforms like Facebook and Twitter.

As you will see, the United States and our users are relatively middle of the pack when it comes to time spent online. My question to you both is this, do you agree that Americans use of social media is not especially unique on a per capita basis? In other words, are you aware of specific trends on your platforms to explain the amount of gun violence in the United States?

The CHAIRMAN. Senator Duckworth, and this will not come out of your time, do sort of explain to us, because some of us cannot see the detail.

Senator DUCKWORTH. Sure, this is how much time average number of hours that social media users spend using social media each day via any device.

The CHAIRMAN. And the arrow points to the United States?

Senator DUCKWORTH. To the United States. The highest is the Philippines. The lowest is Japan. The U.S. is right in the middle. So American users and I have got a four and a half year old and I have an 18 month old and when I get home says iPhone, iPhone and she is on it. She knows how to select YouTube kids on my phone, and she knows how to go right to what she wants to watch. OK, so I am just as concerned that the United States in terms of social media usage, which you both agree, is somewhere in the middle of the pack compared to the rest of the world.

Ms. BICKERT. Yes, Senator, according to the study which I am not more familiar with, yes.

Senator DUCKWORTH. In other words, are you aware, are either of you aware of specific trends on your platforms to explain the amount of gun violence in the United States?

Mr. PICKLES. No, I think your study reflects our view, about 80 percent of our users are outside the United States. And so I think you are right. The image speaks for itself.

Senator DUCKWORTH. Thank you. Mr. Selim, you brought up the role that video games can play on online hate and harassment. I agree with you that any dissemination of hate must be addressed regardless of the platform used. But if a meaningful connection between video games and gun violence exists, you think that the widespread use of video games in Japan and South Korea would reflect that connection, correct? If you look at this chart, I think there is something to be said for the availability of guns in the U.S.

If you look into the amount of time that the folks in Japan and South Korea spend on video games is far greater than anywhere else. We are third, and yet if you look at the number of incidents of gun violence and gun deaths per every 100,000 people in 2017, here is the U.S., but we are not the biggest users of video games. Would this be accurate?

Mr. SELIM. Senator, thank you for your question. I have not read this specific study, but I do have one data point, if I may share with you for just a moment, according to an ADL report looking at extremists related murders and homicides over the past decade, our research shows that 73 percent of extremist related murders and homicides were in fact committed with firearms. So to the extent that you are making the point that extremists with weapons results in violence and homicide, we have the data that backs that point up.

Senator DUCKWORTH. Thank you. As we are reminded daily, the world is full of individuals who use social media platforms to disparage others, cast false equivalencies, and question facts. Some will use the unanimity of online platforms to spread hate but our use of social media, video games, and other variables does little to explain the 2,226 mass shooting since Sandy Hook. The Internet has emboldened and empowered hate by allowing individuals to develop online communities and share their warped ideas, but it is our weak gun laws here in the U.S. that allows that hate to become lethal. There is a clear and undeniable connection between the number of guns in the United States and the number of gun deaths in our community.

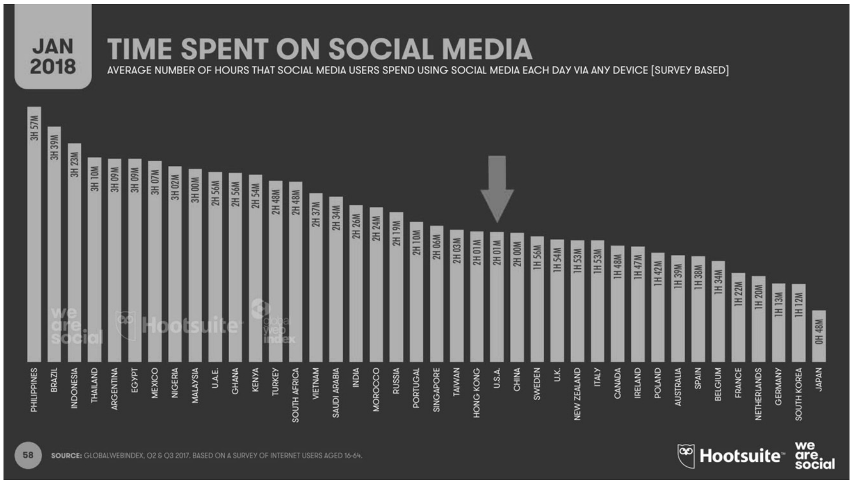
Look at this platform. This is the number of guns per 100 people and this is the number of gun related deaths per 100,000 people. We are up here. Here is the rest of the world. Some of whom use more social media than we do. Some of whom actually engage in more video games than we do. We are saturated in weaponry that was designed for war but is made available to nearly anyone who attends a local gun show.

A Dayton shooter has hundred round drum. I didn't have a hundred round drum when I served in Iraq. We did not send Marines into Fallujah with hundred round drums, but yet you can buy them at gun shows. Look, 90 percent of Americans that agree that Congress should expand background checks and red flag laws. 60 percent of Americans agree that banning high-capacity ammunition clips is what we need to do.

This is not controversial. It is well past time that Leader McConnell brings HRA to the House and passed bipartisan background checks back to the Senate floor for a vote. I hope Leader McConnell will also allow votes on to keep American Safe Act, the Extreme Risk Protection Act, the Disarm Hate Act, and the Domestic Terrorism Prevention Act. Each of these bills will keep our children and our neighbors safer. I hope my Republican colleagues will join in these bipartisan efforts. Thank you and I yield back.

The CHAIRMAN. Senator Duckworth, let's do this so we can have a complete record. If you would reduce those three posters to a size that we can copy, and they will be admitted in the record at this point in the hearing without objection.

[The information referred to follows:]

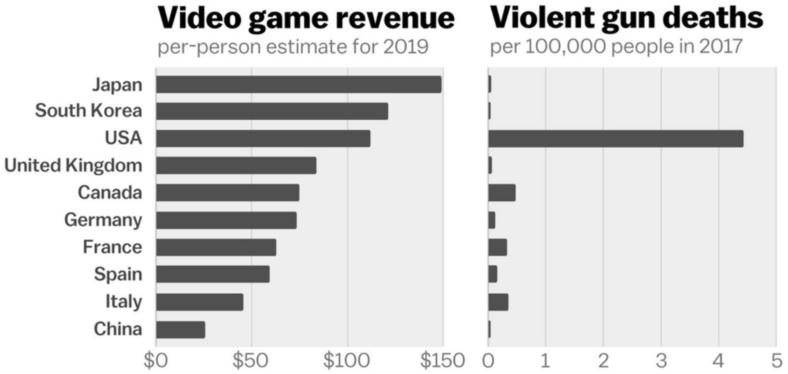


JAN 2018

TIME SPENT ON SOCIAL MEDIA

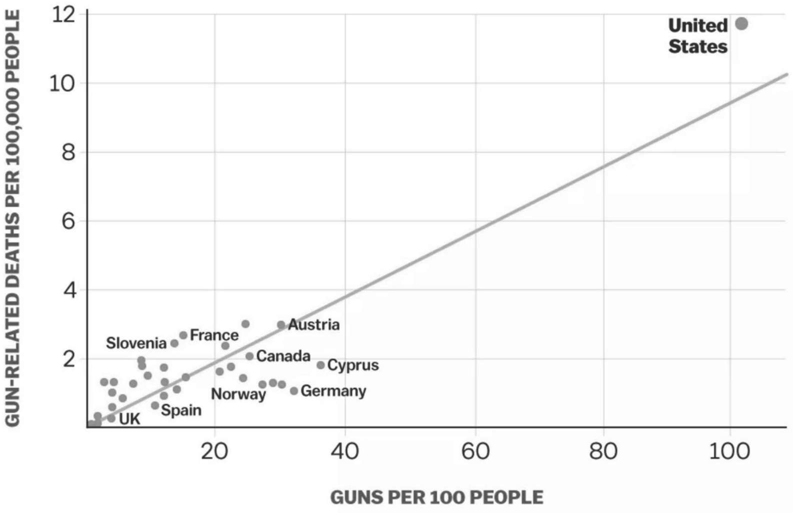
AVERAGE NUMBER OF HOURS THAT SOCIAL MEDIA USERS SPEND USING SOCIAL MEDIA EACH DAY VIA ANY DEVICE [SURVEY BASED]

SOURCE: GLOBALWEBINDEX, Q2 & Q3 2017. BASED ON A SURVEY OF INTERNET USERS AGED 16-64.



Violent gun death data from the Institute for Health Metrics and Evaluation; video game revenue data (which does not include hardware sales) from Newzoo, a gaming analytics company

Vox



Source: Gunpolicy.org, United Nations Development Programme

Vox

Senator DUCKWORTH. Thank you very much, Mr. Chairman, that is generous of you.

The CHAIRMAN. Senator Young.

STATEMENT OF HON. TODD YOUNG, U.S. SENATOR FROM INDIANA

Senator YOUNG. Thank you, Mr. Chairman. I want to thank all of our panelists for being here today. I really do appreciate your testimony and your answering our questions. Look, we all need to collaborate in curbing online extremism, which I understand to be one of multiple causes that we could cite as we all think about the issue of mass casualty events and extremist events, or generally. The Nation is wrestling with mass violence extremism and issues of responsibility, digital responsibility, for some of these events.

In fact, in my home state of Indiana, Hoosiers and Crown Point, Indiana recently experienced firsthand how a person can become radicalized over the internet, something I know that many of your companies have studied and are working on. In 2016, a Crown Point man was arrested and convicted for planning a terrorist attack after becoming radicalized by ISIS over the internet. Thankfully the FBI in the Indianapolis Joint Terrorism Task Force intervened before any violent attack occurred. However, that is not always the case as we know, and we have seen this across the country.

And that is why it is critically important that we have this hearing, that we continue to work together collaboratively, knowing that your products and platforms provide incredible value to consumers and they obviously were not intended for this purpose. So it is our responsibility in Congress, it is definitely your responsibility as business people, to make sure that we monitor how the great value that you provide can be used in an illicit, improper, dangerous, and nefarious manner.

In one minute or less because I have three minutes less left, I would request that the representatives from Google and Facebook and Twitter tell us why Americans should be confident that each of your companies are taking this issue seriously, and why Americans should be optimistic about your efforts going forward?

The CHAIRMAN. One minute each?

Senator YOUNG. Yes, indeed.

Mr. SLATER. Thank you, Senator. I would start by pointing to YouTube community guidelines enforcement report, which details every quarter videos we have removed, the reasons why, and indeed how much is being flagged first by machines in dealing with this issue, removing violative content, as a combination of technology and people. Technology can get better and better at identifying patterns.

People can help deal with the right nuances and we have seen over time that the technology is getting better and better at taking down the content faster and before people have viewed it. As I've said at the outset, of the 9 million videos that we removed in the second quarter of this year, 87 percent of those were first flagged by our machines, and 80 percent of those were removed before a single view. When we talk about violent extremism which it is generally better in terms of removable before wide viewing.

So, you know, we are already seeing advancements in machine learning not just in this area but across the industry broadly, and the thing about machine learning is as it is fed more data, as it learns from mistakes, as we say, you got to learn here. Those systems will get better, and so why one should be optimistic if those systems ideally will continue to get better. Will they be perfect? No, bad actors will continue to evolve, but I do think there is room for optimism, and I think there is reason for optimism based on the collaboration between all of us today.

Senator YOUNG. Thank you. Facebook.

Ms. BICKERT. Thank you, Senator. The first thing I will say is Facebook will not work as a service if it is not a safe place, and this is something that we are keenly aware of every day. If we want people to come together to build this community, they have to know they are safe. And so the incentives are there for us to make sure we are doing our part.

One of the things that we have on our team of more than 350 people who are primarily dedicated in their jobs to countering terrorism and hate is expertise. So I lead this team, my background is with more than a decade as a Federal criminal prosecutor and safety and security are personal to me. But the people that I have hired onto this team have backgrounds in law enforcement, in academia, studying terrorism and radicalization. This is something that people come to work on at Facebook because this is what they care about. They are not assigned to work on it while they are at Facebook. This is bringing in expertise, and I want to make that very clear.

And then finally, similar to my colleagues here, we have taken steps to make what we are doing very transparent. The reports we published in the past year and a half show a steady increase in our ability to detect terror, violence, and hate much earlier when it is uploaded to the site and before anybody reports it to us. Now more than 99 percent of the violent videos and the terrorist propaganda that we removed from the site we are finding ourselves before anybody reports it to us.

Senator YOUNG. Thank you. Twitter.

Mr. PICKLES. Thank you, Senator. I think people can be optimistic. A few years ago, at the peak of Islamic caliphate so-called, people challenged our industry to do more be better. I now look at a time where 90 percent of the terrorist content Twitter removes is detected through technology. I look at independent academics like Professor Morecambe who talked about the IS community being decimated on Twitter. I look at the collaboration that we have between our companies, which didn't exist when I first joined Twitter five and a half years ago. All of those areas have been driven by better technology, faster response, and a much more aggressive posture toward bad actors.

Twitter is now showing benefit in other areas, but I think we can also take confidence that no one is going to tell this committee our work is done. And every one of us will leave here today knowing we have more to do and we can never sleep. These actors are adversarial, and we have to keep it active.

Senator YOUNG. Thank you so much. I could spend five days, five weeks, maybe five months, or five years in this. I only had five minutes. I am already one minute over, Mr. Chairman.

The CHAIRMAN. Thank you. Senator Rosen, you are next. I am going to go vote and I can assure you I will not let them close that vote until you have asked your questions and get over there.

Senator Rosen.

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

Senator ROSEN. I appreciate that, Senator. Thank you for holding this important hearing. I want to thank all the witnesses for being here to talk about this very real and difficult issue. The rise of extreme on extremism online is a serious threat and the Internet is unfortunately proven of valuable tool to extremists who are connecting with one another through various forms to spread hate and dangerous ideologies. While we are here to focus today on the proliferation of extremism online, which of course is incredibly important, we must not lose sight of the fact that violent individuals who find communities online to fuel their hatred have also acted in the name of hate.

We cannot ignore the fact that the absence of sensible commonsense gun safety measures like background checks are allowing individuals to access dangerous weapons far too easily. And so we know the majority of Americans want us to support that, but I represent the great State of Nevada, and as we approach unfortunately the 2-year anniversary of the one October shooting in Las Vegas, the deadliest mass shooting in modern American history, we know that coordination with and between law enforcement is more important than ever. The Southern Nevada Counterterrorism Center also known as our Fusion Center is an example of a dynamic partnership between 27 different law enforcement agencies to rapidly and accurately respond to terrorists and other threats.

With Las Vegas hosting nearly 50 million tourists and visitors each year, the Fusion Center is responsible for preventing countless crimes and even acts of terrorism. So to all of you, can you please discuss with us your coordination efforts with law enforcement when violent or threatening content is identified on your platforms, and what do you need from us as a legislative body to promote and enable, facilitate, whatever word you want to use, to facilitate this partnership to keep our communities safe from another shooting like the one in October? Please.

Ms. BICKERT. Thank you, Senator. The attack was incredibly tragic, and our hearts are with those who have suffered and did suffer in that attack. Our relationship with law enforcement first is an ongoing effort. We have a team that does trainings to make sure that law enforcement understand how they can best work with us. And that is something that we do proactively, we reach out and offer those.

Anytime there is a mass violence incident, we reach out to law enforcement immediately even if we are not aware of any connection between our service and the incident. We want to make sure that they know where we are and how to reach us. We also have an online portal through which they can submit legal process, including emergency requests, and we have a team that office is staffed 24 hours a day so that we can respond quickly.

And finally, we proactively refer imminent threat of serious physical harm to law enforcement whenever we find them.

Senator ROSEN. Thank you.

Mr. PICKLES. Thank you, Senator, and I just wanted to echo firstly Monika's sympathies for your constituents who were victims of that horrible tragedy. The lessons I think we have learned since that attack have continued to inform our thinking, and for example, not waiting for the ideological intent of the shooter to be known before acting. I think one of the challenges we have is in the traditional terrorist space, we might look for an organization affiliation before we would say, this is a terrorist attack.

We don't wait for that anymore. We act first to stop people using our services. As Monika said, we do cooperate with law enforcement and provide credible threats. I think one of the questions and I along with colleagues from other companies actually met with a number of agencies yesterday to discuss how we can further deepen our collaboration, and one of the questions we had there is a huge amount of information within the law enforcement community, within the DHS umbrella, that is classified that might help us understand the threats, the trends, the situational awareness.

So understanding how more information can be shared with our industry to inform us about the threats——

Senator ROSEN. Can you provide us in writing some of the tools that you think you might need to help you better cooperate to protect our communities?

Mr. PICKLES. Absolutely, and that was the subject of the meeting yesterday and we had a very productive conversation.

Senator ROSEN. Thank you.

Mr. SLATER. Senator, broadly similar here, both in horror and sympathy. Tragedies like that one and in the ways that we proactively cooperate with law enforcement refer credible threats as well as receive valid request emergency disclosure request and respond to them expeditiously.

Senator ROSEN. Thank you. I see my time is up. I am going to submit a question for the record about combating violent anti-Semitism online. I know other people are waiting. We have votes. I appreciate your time and your commitment to solving, working on this issue.

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

Senator LEE [presiding]. Thank you, Senator Rosen. Your questions will be submitted for the record. I want to start with a simple yes or no question. I don't mean this to be a trick yes or no question answer. It is either yes or no, or yes or no with a brief one sentence caveat if you need to. I would like to hear from each of the three of you, from Ms. Bickert, and then Mr. Pickles, and then Mr. Slater: Do you provide a platform that you regard and present to the public as neutral in the political sense?

Ms. BICKERT. Yes, Senator, our rules are politically neutral, and we apply them neutrally.

Senator LEE. So you aspire to political neutrality as to left versus right?

Ms. BICKERT. We want to be a service for political ideas across the spectrum.

Senator LEE. Mr. Pickles?

Mr. PICKLES. We enforce our rules impartially and our rules are crafted without ideology included.

Senator LEE. Mr. Slater?

Mr. SLATER. Similarly, we craft our services without regard to political ideology though as we have discussed today, we are not neutral against terrorism or violent extremism.

Senator LEE. Yes, and I appreciate you pointing that out that is of course not what I am talking about. And that leads into the next question I wanted to raise with each of you. I think it is important the work each of you are doing in this area is important. It is important for anyone occupying this space to be conscious of those things. You do a service to those who access your services by removing things like pornography, terrorism advocacy, and things like that. There is a lot of debate that surrounds this issue and surrounds some of the legal framework surrounding it.

As you know Section 230 of the Communications Decency Act has received a lot of criticism. It protects a website from being held liable as a publisher of information by another information content provider. And significantly, Section 230 is a good Samaritan provision. It gives you the promise that you won't be held liable for taking down this type of objectionable content that we are talking about, whether it is something that is constitutionally protected or not. And so for each of the same witnesses, again, I would ask you, each of you represents a private company and each of you are accountable to your consumers within your company. This means that in some sense, that you have incentives to provide a safe and enjoyable experience on your respective platform. So I have got a question about Section 230.

Does Section 230, particularly the Good Samaritan provisions, help you in your efforts to swiftly take down things like pornography and terrorist content off your platforms? And would it be more difficult without the legal certainty that Section 230 provides?

Ms. BICKERT. Absolutely, Senator. Section 230 is critical to our efforts in safety and security.

Senator LEE. Mr. Pickles?

Mr. PICKLES. Absolutely. I would go further and say that Section 230 has been critical to the leadership of American industry in the information technology sector.

Senator LEE. Mr. Slater?

Mr. SLATER. Absolutely. Yes.

Senator LEE. On a related point, imagine a world where this is suddenly taken away, where those provisions no longer exist. Large companies like yours might be able to—I strongly suspect still would be able to and still probably would filter out this content between the artificial intelligence capabilities at your disposal and the human resources that you have.

I suspect you could and probably would still do your best to perform the same function. What about a startup, what about a company trying to enter into the space that each of your companies entered into when they were created not very many years ago? What would happen to them? Ms. Bickert?

Ms. BICKERT. Senator, thank you for that question. This reminds me of industry conversations involving smaller companies back before we formed the Global Internet Forum to Counter Terrorism in June 2017. We were having closed door sessions with companies, large and small, to talk about the best ways to combat the threat of terrorism online, and the smaller companies were very concerned about liability. Section 230 is very important for them to be able to begin to proactively act and assess content.

Mr. PICKLES. I would say it is a fundamental part of maintaining a competitive online ecosystem. And without it, the ecosystem is less competitive.

Senator LEE. Mr. Slater?

Mr. SLATER. Yes, and I just add, the U.S. has Section 230 and that is part of the reason why we have been a leader in economic growth and innovation and technological development. Other countries that don't have something like it suffer, and study after study has shown that. And we will be happy to discuss that more.

Senator LEE. If it were to be taken away—so all three of your companies, in particular Mr. Slater, not exactly known for being a small business or a business with a modest economic impact, but you can identify, I assume, with this concern I am expressing if we were to take that away Google might be able to keep up with what it needs to do, but wouldn't it be harder for someone to start say a new search engine company, a new tech platform of one sort or another, as somebody starting out in the same position where your company was a couple of decades ago. Wouldn't that be exponentially more difficult?

Mr. SLATER. I think it would create problems for innovators of all stripes, but certainly small, medium sized businesses would have a lot of trouble potentially getting their arms around that sort of significant change to the fundamental legal framework of the internet.

Senator LEE. Thank you. My time has expired.

Senator Baldwin.

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

Senator BALDWIN. Thank you. I wanted to begin by thanking our full committee Chairman Wicker for holding this hearing. I think it is a vital conversation for us to be having. We need to be taking a hard look at how we address the rising tide of online extremism and its real world consequences in our country. I do have some questions for you on this important topic, but first I wanted to echo some of what my colleagues have already said, which is there is much more that the Senate must do to address gun violence, whether or not it is connected to hatred espoused on the internet.

So more than 200 days ago, the House of Representatives passed a bipartisan universal background check bill and this commonsense gun safety measure has an extraordinary level of public support. It deserves a vote on the Senate floor, and I feel like we can't simply have hearings, but we have to act to reduce gun violence. Mr. Selim, ADL Center on Extremism has closely studied hate crimes and extremist violence in this country. Is it fair to say that there has been an alarming increase in bias-motivated crimes including extremist killings in the last several years?

Mr. SELIM. Yes, Senator, that is accurate.

Senator BALDWIN. In the case of extremist killings, what role do you feel that access to firearms has played in that increase?

Mr. SELIM. Senator, thank you for that question. As I briefly alluded to earlier just to expand on what I was mentioning, according to our recent ADL report, extremists of all ideological spectrums that committed murders or homicides in the United States, 73 percent of those acts were committed with firearms.

Senator BALDWIN. Thank you. What impact do you believe this increase in hate crimes, including extremist killings, have on the minority communities whose members have been the targets of these attacks, and let me just add to that question. One of the unique aspects of a hate crime is that it now not only victimizes the targeted victim, but it strikes fear among those who share the same characteristics with the victim or victims.

Mr. SELIM. Senator, thank you for making this point. In the past 24 months, we saw a calendar year 2017 with a 57 percent increase of anti-Semitic incidents across the country. The FBI and DOJ's own hate crime data showed a 17 percent increase in hate crimes and bias-motivated crimes in calendar 2017. We continue to see these troubling statistics year after year and so it is imperative, and part of my testimony today, both the submitted written and my oral testimony, speaks to the need for greater enhancement and enforcement of hate crime laws and protections for victims.

Senator BALDWIN. I am an original co-sponsor of Senator Bob Casey's legislation to disarm hate crime, hate act, which would bar those convicted of misdemeanor hate crimes from obtaining firearms. Do you agree that this measure could help keep guns out of the hands of individuals who might engage in extremist violence?

Mr. SELIM. Yes, Senator. Thank you for your leadership and all members who have supported this legislation. ADL supports this legislation.

Senator BALDWIN. Thank you. I appreciate the efforts that our witnesses from the social media companies have described regarding their company's efforts to combat online extremism, including to provide some transparency to their users and the general public. It is of course critically important to understand how you are addressing problems within your existing services and platforms. I would actually like to learn more from you about how you are thinking about this issue as you develop and introduce new products.

In other words, I think a lot of us feel that the approach of rapidly introducing a new product and then assessing the consequences later is a problem. So I would like to ask you how do you plan to build combating extremism into the next generation of ways in which individuals engage online, and why don't we start with you Ms. Bickert?

Ms. BICKERT. Thank you for the question, Senator. Safety by design is an important part to building new products at our company. One of the things we have built in the past maybe 5 years is a new products policy team that is under me. Their responsibility is to make sure they are aware of new products and features that are being built and explaining to these engineers who are thinking of all the wonderful ways that the service could be used, all of the abuse scenarios that we could also envision and making sure that we have reporting mechanisms or other safety features in place.

Mr. PICKLES. I think as I said earlier, we are in a very adversarial space. We know that bad actors will change the behavior. And so every time we have a feature, a policy decision, one of the key processes in that part of the discussion is how can this be used against us? How can this be gamed? How will people change their behavior to try and circumvent the policy? And you are absolutely right. We need to take that learning and share it with smaller companies. Certain the work that FCT has done, working with more than I think 200 small companies around the world to share that knowledge with them, to help them understand the challenges, is also invaluable.

Mr. SLATER. Similarly, our trust and safety teams are at the table with product managers and engineers from the conception of an idea all the way through the development and possible release. So from ground up, it is safety by design.

Senator BALDWIN. Thank you.

STATEMENT OF HON. DAN SULLIVAN, U.S. SENATOR FROM ALASKA

Senator SULLIVAN [presiding]. So I want to thank the witnesses, and I am going to be taking over as the Chair, and I will call on myself as the next witness. I want to actually ask all of you, you know, your companies, your technology, you are famous for its algorithms, which seemed to have the ability to pinpoint on what people want. You know, you can put an e-mail out or even some people think, talk about say your interest in yellow sweaters, and next thing you know, you have ads popping up on your Facebook or other accounts that talk about yellow sweaters. Who knows how that happens but to a lot of us it has. It is pretty impressive.

But here is my question. If your algorithm technology is so good at, kind of, pinpointing things like that, what people are interested in, particularly as it relates to ads, what are the challenges with regard to directing that kind of technology to help us and help you find what is being talked about here on both sides of the aisle which is the people who are committing this kind of violence are typically disaffected young males, and aren't there signs, aren't there things that you can do with the technology that you do so well in other spaces to at least provide more warning signs of this kind of violence from these kind of individuals who in some ways already have a profile online? Throw that out to any of you. And are you working on that?

Ms. BICKERT. Thank you for the question, Senator. Technology plays a huge role in what we are doing to enforce our safety policies at Facebook. In the area of terrorism to extremism, and violence, it is not just the matching software that we have to stop things like organized terror propaganda videos.

We are now using artificial intelligence machine learning to get better at identifying new content that we have not seen before that might be promoting violence or trying to incite violence or engage in other harmful behavior. Anytime that we find a credible threat of imminent physical harm, we proactively send that out to law enforcement. And these systems are getting better every day.

Senator SULLIVAN. And are you using algorithms and the advanced technologies that you use in other spaces to help identify those threats?

Ms. BICKERT. There are certainly cross learnings across the company. There are different products that work in different ways, but——

Senator SULLIVAN. But is it a priority of yours, the way it would be for selling yellow sweaters?

Ms. BICKERT. Oh, absolutely. And this is something that we do——

Senator SULLIVAN. Can I ask that of all the companies here?

Mr. PICKLES. Absolutely, investing in technology to find content that is terrorist content, violent extremist content, is absolutely a priority.

Mr. SLATER. It is a top priority. Yes.

Mr. SELIM. Senator, I would only add to this part of the conversation as someone who studied the research in the data around these issues for nearly two decades, the threat environment that we are in today has changed significantly. White supremacist terrorists in the United States do not have training camps in the same way that foreign terrorist groups do like Al-Qaeda or Isis. Their training camp where they connect, learn, and coordinate with one another is in the online space.

So it is imperative that the question you are asking about the machine learning, the technology, the artificial intelligence continue to advance to disrupt that environment and make it an inhospitable place for individuals that want to promote violent content of any ideological spectrum to be disrupted.

Senator SULLIVAN. Let me ask another question. This is kind of a bigger kind of policy question, but you all of your companies kind of have this tension between you want eyeballs, on right, you want more clicks, you want more time on, and yet—with Facebook or Google or Twitter, and yet there I think there is increasing studies that are showing for example the amount of young men and women, young girls, who feel kind of a sense of loneliness from their time online.

You know, there is indications that among teenagers, the suicide rates are increasing particularly for young girls. One of the things that I worry about, you know, we are all dealing with this opioid epidemic right now and we are looking back going, my God, how did we how did we do that? How did we get to this position in the 90s and the policies, and other things that you know, 72,000 Americans died of overdoses last year.

And so we are, kind of, looking backward saying, how did this happen? Do you, in your kind of c-suites of policymaking, do you ever wonder why we can be looking back in 20 years going, how in the hell did we addict a bunch of young Americans to look at their damn iPhones 8 hours a day and 20 years from now we are going to be seeing the social and physical and psychological ramifications where we all might be kicking ourselves in the head saying, why did we allow that to happen?

Do you guys ever think about that? Because I think about that and it worries me, but you have tension because you want—don't you want more Facetime, don't you want young teenagers spending 7 hours a day staring at their iPhones because that helps your revenues? Do you worry that 15, 20 years from now, we are going to be in the same spot that we are with opioids and saying, what did we do to our kids? What did we do to our citizens? Do any of you guys worry about that? Your power, your negative implications of what is happening in society right now.

Ms. BICKERT. Senator, thank you for the question. As a mother, I take these questions about wellness very seriously and our company does as well. And this is something that we look at and we talk to youth wellness groups to make sure that we are crafting products and policies that are in the best long-term interests of the people who want to come and connect through Facebook.

I also want to say that we have seen social media be a tremendous place for support for those who are thinking of harming themselves or struggling with eating disorders or opioid addiction or getting exposed to hateful content. And so we are also exploring and developing ways of linking people up with helpful resources. We already do that now for opioid addiction, for thoughts of self-harm, for people who are asking or searching for hateful content. We now provide them with help resources. We do think that this can be a really positive thing for overall wellness.

Mr. PICKLES. I just thought we have similar programs in place for both opioids searches and also for people who are using terms referencing self-harm or suicide where we will provide, intervene and provide them with a source of support. And that is something we have rolled out around the world. I think the other thing is we certainly recognize that things like digital literacy are issues that we as industry and certainly we as Twitter need to invest in to make sure that as people using our services, they also have the skills and the awareness to use them discerningly.

And then finally, our CEO is committed to the company, to looking at the health of the conversation, and not just using the kind of metrics that you have referenced but looking at much more broader metrics that measure the health of the conversation rather than just revenue.

Senator SULLIVAN. Thank you, Mr. Chairman.

The CHAIRMAN [presiding]. Thank you, Senator Sullivan.

Senator Cruz.

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

Senator CRUZ. Thank you, Mr. Chairman, and I will say thank you to my friend from Alaska for sharing apparently this deep void and longing in your heart. And I just want to reassure you for Christmas, you will be getting that yellow sweater.

[Laughter.]

Senator CRUZ. Mr. Slater, I want to start with you. I want to talk a little bit about Project Dragonfly. In August 2018 it was reported that Google was developing a censored search engine under the alias of Project Dragonfly. In response to those concerns, Alphabet shareholders requested that the company publish a human rights impact assessment by October 30 of this year examining the actual and potential impacts of censored Google search in China.

However, during Alphabet shareholder meeting on June 19, the proposal for the assessment was rejected. In fact Alphabet's Board of Directors explicitly encouraged shareholders to vote against the proposal and Alphabet commented that "Google has been open about its desire to increase its ability to serve users in China and other countries.

We have considered a variety of options for how to offer services in China in a way that is consistent with our mission and have gradually expanded our offerings to consumers in China." So I want to start with just some clarity. Mr. Slater, has Google ceased any and all development and work on Project Dragonfly?

Mr. SLATER. Senator, to my knowledge, yes.

Senator CRUZ. And has Google committed to foregoing future projects that may be named differently, but would be focused on developing a censored search engine in China?

Mr. SLATER. Senator, we have nothing to announce at this time. And I think whatever we would do, we would look very carefully at things like human rights. In fact, we work with the Global Network Initiative on an ongoing basis to evaluate how our principles, our practices, our products comport with human rights in the law.

Senator CRUZ. So, roughly contemporaneously, Google decided that it didn't want to work with the U.S. Department of Defense. How does Google justify having been willing to work with the Chinese government on complex projects including artificial intelligence under Project Maven and at the same time not being willing to help the Department of Defense develop ways to minimize civilian casualties through better AI? How do you how do you reconcile those two approaches?

Mr. SLATER. Senator, as we have talked about today, we do partner with law enforcement and we do partner with the military in certain ways offering some of our services. Also as a business, we draw responsible lines about where we want to be in business, including limitations on and getting in the field of building weapons and so on, and you know, we will continue to evaluate that over time.

Senator CRUZ. Let me shift to a different topic which is this panelists talked about combating extremism and the efforts of social media to do that. Many Americans, including myself, have a longstanding concern that when big tech says it is combating extremism that that is often a shield for advancing political censorship. Mr. Pickles, I want to talk about recently Twitter extended its pattern of censorship to the level that it took down the Twitter account of the Senate Majority Leader Mitch McConnell.

That I found a pretty remarkable thing for Twitter to do, and it did so because that account, as I understand it, had sent out a video of angry protesters outside of Senator McConnell's house, including an organizer of Black Lives Matter in Louisville, who is heard in the video saying that the Senate Majority Leader "should have broken his little raggedy wrinkled ass neck," and someone else who had a voodoo doll of the Majority Leader and another angry protester said, "just stab the mf's heart," although that person did not abbreviate mf. Senate Majority Leader sent out those threats of violence and found rather remarkably his own Twitter account taken down. How does Twitter explain that?

Mr. PICKLES. Well, thank you Senator for the opportunity to discuss this. Something we have been asked around the world is the climate in many political jurisdictions of safety of people who hold public office. And so when we saw a video posted by numerous users that clearly identified someone's home and clearly contained as you so referenced some quite severe threats out of an abundance of caution, we did remove that video. We didn't remove the accounts. We moved that single tweet that contained the video from everybody who had posted it because the essence of a video with someone's personal home where the Senate Majority Leader may have been residing at the time with several violent references, we felt was something out of an abundance of caution we should remove. We then discussed this further with the Leader's office. We understood their intent was to call attention to those very threats of violence.

And so we did permit the video to be put on Twitter with a warning message saying this is sensitive media, but it is that balance that we are striking between—I have been in many different situations where I have been asked the exact opposite which is similar content should be removed because it contains a clear violent threat, and that balance is something that we strive to get right every day. But our first thought in that instance was the safety of Leader McConnell and his family.

Senator CRUZ. You would agree there is a difference between someone posting video where they are threatening someone else and the target of that threat posting the video. Do you agree that those are qualitatively different?

Mr. PICKLES. I think that is holy fair, but I think in the situation where you have the person's home visible in the video, there is still a risk there and we are motivated by preventing that offline harm that could have occurred because the home was visible. It was a hardcore and we appreciate the Leader's discussion and discussing with his campaign team and his Senate office, and we appreciate their insight. But this was something that our motivation was to prevent harm, not the kind of potentially ideological issues you may allude to.

Senator CRUZ. Thank you.

The CHAIRMAN. But Mr. Pickles, have you rethought your policy since the instance that Senator Cruz asked about? And I would call your attention to Ms. Bickert's testimony, written testimonial, on page 2 which says, and I quote, "we do not allow propaganda or symbols that represent any of these organizations or individuals to be shared on a platform unless they are being used to condemn or inform." Is that language instructive to your platform and don't you think that clearly it was readily evident from the beginning that Senator McConnell and his campaign had posted that video to condemn and inform?

Mr. PICKLES. I think this is an absolutely relevant issue. We as a company have taken a more aggressive posture after the Christchurch attack. We did see people posting both excerpts of the manifesto and content of the video to condemn it, and we decided even in those circumstances we would remove it. And for other attacks more recently in the United States where images have been posted to other manifestos with large chunks of the manifestos even where they are condemning it, we have taken the decision to remove that material.

So this is something that is constantly under tension and I think the case you illustrate highlights for us, the complexity in getting this right. But again if we are going to err on the side of caution, fewer violent threats and fewer people's homes being visible on our platform is notably a good thing. We have to work harder at taking into account the kind of context you outline, but this is something where this is the first time—I have been with the company five and a half years, I have never been asked why didn't we leave something up that contains a violent threat, and so I think that in itself is illustrative of the complexity of the situation.

The CHAIRMAN. Well in terms of the context in this instance it was the owner of the home who chose to inform the world about what was being said against him, and it was the individual himself who posted this. And it seems to be a clear cut case in that instance that differentiates it from the condemnation of the larger incident of the Christchurch violence. I would just suggest that it shouldn't have taken very long for Twitter to understand that. Senator Sullivan, you are recognized.

Senator SULLIVAN. Thank you, Mr. Chairman. I just have a couple of follow-up questions. Mr. Slater, one of Senator Cruz's questions. You know, I think it is—whether a company wants to work with the Pentagon I think is something that leadership of the company, individual companies have to make that decision. I think that is certainly something that is fine.

I think what troubles a number of us is that where there is a declaration that you are not willing to work with the Department of Defense on certain issues and yet there is a willingness to work with one of our country's potential adversaries, particularly on sensitive technological issues that are important to the competition between the two nations.

Do you understand why that has caused bipartisan concern here? And how should we address it? Should Congress take action on those kinds of situations? Not saying everybody has to work for the Pentagon, that is your decision. But if you don't want to work to help with our Nation's defense, but you are working with the country that poses a very significant threat long-term to the United States, do you understand why that causes concern here?

Mr. SLATER. Senator, I do appreciate the concern. We are proudly an American company. We are a business that wants to draw a responsible lines and we look forward to continue to engage with you, the Committee, and others to make sure we are doing that.

Senator SULLIVAN. Do you think if there are instances of that, a clear-cut example of, hey, we are not going to do anything on the Nation's defense with the U.S. Department of Defense, but we are going to work with the Chinese, something very clear and obvious. Do you think there is something that we should do to prevent that or penalize that? We the Congress?

Mr. SLATER. I think it is an important question. I think as a business we try and strike responsible and consistent lines, but the details would certainly have to matter.

Senator SULLIVAN. OK. Mr. Pickles let me ask just a one final question. It is really a follow up to Senator Scott's earlier question. You said that the Twitter account of Maduro in Venezuela has not "broken any of the rules." What are those rules? And at what point would you look to have somebody who is certainly not treating his citizens well? And Senator Scott has been a leader on this issue, but, you know, what are those rules and at what point would you look at what they are doing in their own citizen as a way to maybe not provide them the platform that you have?

Mr. PICKLES. Thank you. Well firstly the rules apply to any user on Twitter at the same. I can make a make a full copy available and it will be, for example, whether it is encouragement of violence. If the Twitter account was used in some of the ways that we have seen around the world to encourage violence against minorities, to organize violence, we would take action on those accounts breaking those rules.

Senator SULLIVAN. Would Twitter allow Putin to have an account or Xi Jinping to have an account?

Mr. PICKLES. If they were acting within our rules, the one thing I would note is, and this is slightly different but important, some worldly, some governments have sought to manipulate our platform to spread propaganda information through breaking our rules. One of those governments is Venezuela, and we have made a public declaration of every account that we removed from Twitter for engaging in information operations covertly that we believe is responsible for that government.

We made that whole archive available to the public and to researchers. We have taken this same step with information operations that have been directed we believe from countries including China, Iran, and Russia because we believe that it is not just those single Twitter accounts, that some governments do also seek to manipulate our platform. And where they do so, we will take action to remove that manipulation and make it public so people can learn——

Senator SULLIVAN. So if the government takes violence against its own citizens, is that breaking the Twitter rules?

Mr. PICKLES. What I think of that act is activities happening offline, and the key question for us is, what is happening on Twitter?

Senator SULLIVAN. Thank you. Thank you, Mr. Chairman.

The CHAIRMAN. Thank you, Senator Sullivan. And thank you to our witnesses. 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 complete written answers to the Committee as soon as possible but no later than Wednesday, October 2, 2019 by close of business.

I thank each and every one of you for appearing today. This hearing is now adjourned.

[Whereupon, at 12:08 p.m., the hearing was adjourned.]

APPENDIX

16 September 2019

To: Chairman Roger Wicker,
Ranking Member Maria Cantwell,
U.S. Senate Committee on Financial Services,
Washington, DC.

Fr: Gretchen Peters and Professor Amr al-Azm
The Alliance to Counter Crime Online

Re: Concerns Facebook Platforms Facilitate Terror, Spread Crime

Dear Chairman Wicker and Ranking Member Cantwell,

As the Senate Committee on Commerce, Science and Transportation prepares to question Facebook's Head of Global Policy Management, we want to express our grave concern that Facebook's platforms are infested by criminal syndicates and terror groups. Facebook has been grossly negligent both in monitoring and removing this toxic content. Moreover, *one of our members has performed research* indicating the firm has knowingly deceived lawmakers, investors and the public about the extent to which the firm is able to remove extremist content.

We want the committee to understand that the world's largest social media company does more than just connect people. The public should not trust Facebook's claim that they have been successful in removing 99 percent of ISIS content because it is only a talking point that they have never been forced to prove. Our research indicates Facebook and its family of platforms are also used by terrorist groups as a megaphone for propaganda, for recruiting new members, and even to fundraise. Just this week, ACCO is preparing to release a report that documents extensive fund-raising activities by designated terror groups such as Lebanese Hezbollah.

ACCO members also include a group of brave Syrian archeologists investigating the *illicit antiquities trade on Facebook.* They have recorded closed groups where almost 2 million regular users log on to trade tens of thousands of artifacts trafficked from conflict regions including Syria, Iraq and Yemen—a *war crime.* Many of the sellers openly declare they are donating proceeds of these sales to ISIS.

The Facebook family of apps are ground zero for organized crime syndicates to connect with buyers, market their illegal goods, and move money, using the same ease of connectivity enjoyed by ordinary users. Instead of acknowledging his technology is being used for illegal purposes and fixing the problem, Facebook CEO Mark Zuckerberg clings to immunities provided by *Section 230 of the Communications Decency Act of 1996,* which courts have interpreted to mean that tech firms shouldn't be held liable for content posted by third-parties.

There is a huge problem with this approach. The algorithms Facebook has touted to connect the world have connected criminals and terrorists faster than Facebook's own *beleaguered moderators* can delete them. The impact of this illegal activity is affecting our communities, our cultures, and our environment, and it's happening in the same digital spaces where our children play, our families connect, and our companies advertise.

In light of all this, Zuckerberg's *announcement* that he plans to alter Facebook to focus on groups—and also launch a cryptocurrency—are downright alarming. Groups are already the epicenter for illicit activity on Facebook. Do we want Facebook to become an even safer place for terrorists and criminals?

There's no reason to believe Facebook's proposed changes will make user data any more secure. After all, Facebook hasn't *changed* its *fundamental business model.* But the changes will make it harder for authorities and civil society groups to track and counter illegal activity on the platform.

The firm's continued negligence in the moderation of criminal and terror content makes clear that the time for self-regulation has passed.

The challenge is that Federal laws take time, something that human trafficking victims, drug addicts and endangered species don't have. But there are other ways

U.S. regulators can address crime on social media. Facebook's IPO may hold the key to effective regulation.

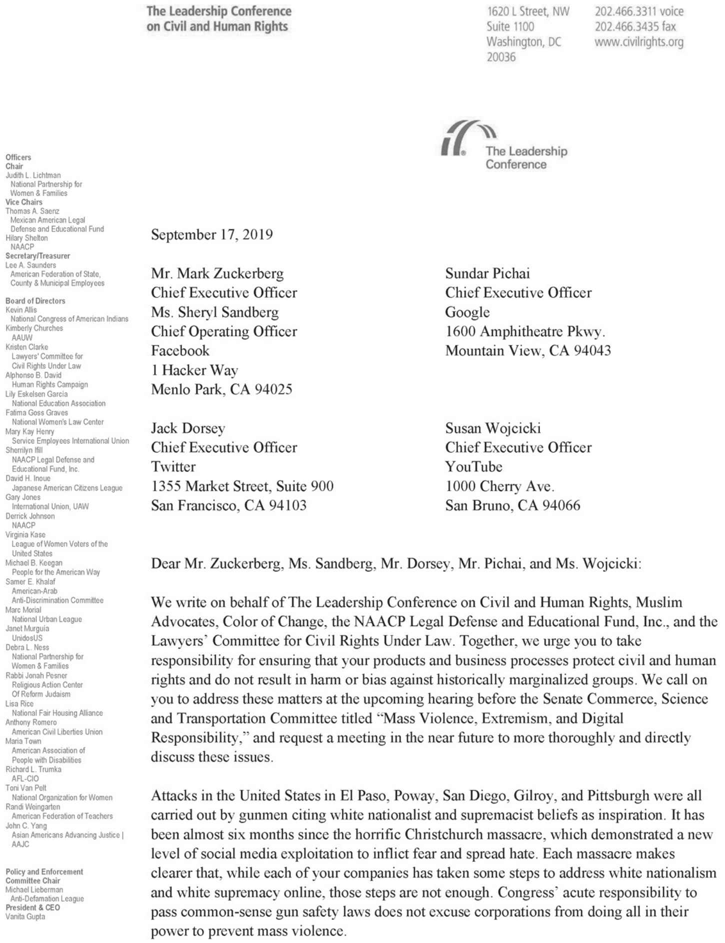
When Facebook went public in 2012, the firm *voluntarily* entered into a strict regulatory regime that negates CDA 230 immunities in the context of Facebook's obligations under securities law. The firm's lack of internal controls and effective compliance programs implicate potentially serious securities law violations. Your committee can influence immediate action by asking the Securities and Exchange Commission (SEC) to utilize its existing regulatory power.

As a result of Facebook's failure to establish appropriate internal controls, criminal activity has accelerated on its platform and continues to grow. Now is not the time to let Facebook launch a cryptocurrency. *It's time* to *make social media a safer space for all.*

Respectfully,

Gretchen Peters, *Executive Director*
Alliance to Counter Crime Online

Dr. Amr AI-Azm, *Co-founder* of ACCO
Director of ATHAR Project



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

The Leadership Conference

September 17, 2019

Mr. Mark Zuckerberg
Chief Executive Officer
Ms. Sheryl Sandberg
Chief Operating Officer
Facebook
1 Hacker Way
Menlo Park, CA 94025

Sundar Pichai
Chief Executive Officer
Google
1600 Amphitheatre Pkwy.
Mountain View, CA 94043

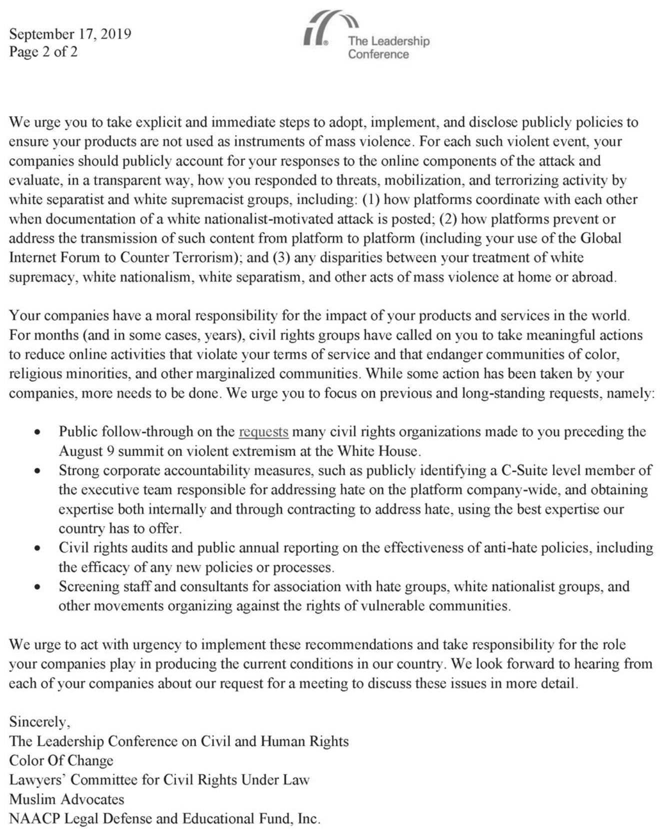
Jack Dorsey
Chief Executive Officer
Twitter
1355 Market Street, Suite 900
San Francisco, CA 94103

Susan Wojcicki
Chief Executive Officer
YouTube
1000 Cherry Ave.
San Bruno, CA 94066

Dear Mr. Zuckerberg, Ms. Sandberg, Mr. Dorsey, Mr. Pichai, and Ms. Wojcicki:

We write on behalf of The Leadership Conference on Civil and Human Rights, Muslim Advocates, Color of Change, the NAACP Legal Defense and Educational Fund, Inc., and the Lawyers' Committee for Civil Rights Under Law. Together, we urge you to take responsibility for ensuring that your products and business processes protect civil and human rights and do not result in harm or bias against historically marginalized groups. We call on you to address these matters at the upcoming hearing before the Senate Commerce, Science and Transportation Committee titled "Mass Violence, Extremism, and Digital Responsibility," and request a meeting in the near future to more thoroughly and directly discuss these issues.

Attacks in the United States in El Paso, Poway, San Diego, Gilroy, and Pittsburgh were all carried out by gunmen citing white nationalist and supremacist beliefs as inspiration. It has been almost six months since the horrific Christchurch massacre, which demonstrated a new level of social media exploitation to inflict fear and spread hate. Each massacre makes clearer that, while each of your companies has taken some steps to address white nationalism and white supremacy online, those steps are not enough. Congress' acute responsibility to pass common-sense gun safety laws does not excuse corporations from doing all in their power to prevent mass violence.



We urge you to take explicit and immediate steps to adopt, implement, and disclose publicly policies to ensure your products are not used as instruments of mass violence. For each such violent event, your companies should publicly account for your responses to the online components of the attack and evaluate, in a transparent way, how you responded to threats, mobilization, and terrorizing activity by white separatist and white supremacist groups, including: (1) how platforms coordinate with each other when documentation of a white nationalist-motivated attack is posted; (2) how platforms prevent or address the transmission of such content from platform to platform (including your use of the Global Internet Forum to Counter Terrorism); and (3) any disparities between your treatment of white supremacy, white nationalism, white separatism, and other acts of mass violence at home or abroad.

Your companies have a moral responsibility for the impact of your products and services in the world. For months (and in some cases, years), civil rights groups have called on you to take meaningful actions to reduce online activities that violate your terms of service and that endanger communities of color, religious minorities, and other marginalized communities. While some action has been taken by your companies, more needs to be done. We urge you to focus on previous and long-standing requests, namely:

- Public follow-through on the requests many civil rights organizations made to you preceding the August 9 summit on violent extremism at the White House.
- Strong corporate accountability measures, such as publicly identifying a C-Suite level member of the executive team responsible for addressing hate on the platform company-wide, and obtaining expertise both internally and through contracting to address hate, using the best expertise our country has to offer.
- Civil rights audits and public annual reporting on the effectiveness of anti-hate policies, including the efficacy of any new policies or processes.
- Screening staff and consultants for association with hate groups, white nationalist groups, and other movements organizing against the rights of vulnerable communities.

We urge to act with urgency to implement these recommendations and take responsibility for the role your companies play in producing the current conditions in our country. We look forward to hearing from each of your companies about our request for a meeting to discuss these issues in more detail.

Sincerely,
The Leadership Conference on Civil and Human Rights
Color Of Change
Lawyers' Committee for Civil Rights Under Law
Muslim Advocates
NAACP Legal Defense and Educational Fund, Inc.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. SHELLEY MOORE CAPITO TO MONIKA BICKERT

Question 1. You mentioned in your testimony the importance of counter speech to prevent people from becoming radicalized. Radicalization comes in many forms and threatens the core values our Nation was created upon. Are there different strategies and best practices in combating domestic vs foreign extremism?

Answer. Terrorists, terrorist content, and hate speech in all forms—including white supremacy and domestic terrorist 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. And we recently updated our definition in consultation with experts in counterterrorism, international humanitarian law, freedom of speech, human rights, and law enforcement. The updated definition still focuses on the behavior, not ideology, of groups. But while our previous definition focused on acts of violence intended to achieve a political or ideological aim, our new definition more clearly encompasses attempts at violence, particularly when directed toward civilians.

In addition to combating foreign terrorism, we are committed to identifying and rooting out domestic hate organizations. We define hate organizations as "any association of three or more people that is organized under a name, sign, or symbol and that has an ideology, statements, or physical actions that attack individuals based on characteristics, including race, religious affiliation, nationality, ethnicity, gender, sex, sexual orientation, and serious disease or disability." In evaluating groups and individuals for designation as hateful, we have an extensive process that takes into account a number of different signals, and we regularly engage with academics and organizations to refine this process.

While we work 24/7 to identify, review, and remove terrorist and violent extremist content, our efforts do not stop there. We have also started connecting people who search for terms associated with white supremacy and hate-based organizations to resources focused on helping people leave behind hate groups. For example, people searching for these terms in the U.S. will be directed to Life After Hate (*https://www.lifeafterhate.org/*), an organization founded by former violent extremists that provides crisis intervention, education, support groups, and outreach. We have also recently expanded this initiative to Australia and Indonesia, where we work with organizations with local expertise on how best to counter hate in their communities. For more information, see *https://newsroom.fb.com/news/2019/03/standing-against-hate.*

People use our platform to speak out against hatred and extremism. They counter hateful content by responding to it directly, raising awareness on important issues, and supporting positive and moderate voices. We believe these efforts to resist and stand up to racism, violence, extremism, and hate are essential. That is why we work closely with local communities, experts 17 in civil society and academia, and policymakers to support counterspeech initiatives across the globe. More information can be found at *https://counterspeech.fb.com/en.*

Question 2. This Committee has held a number of hearings on the rise and importance of artificial intelligence (AI) in today's digital economy. AI has been invaluable in collecting and sorting massive amounts of data. In the case of today's hearing, AI has become critical in order to identify radicalization and terrorist threats. Each company has identified key tools each company uses in identifying bad actors on your platforms, but machine learning being one of the most critical. What factors are given priority when determining radicalized or terrorist content?

a. You also mention the importance of human expertise in determining more nuanced cases. When does human expertise step in after AI has identified or flags content?

b. After content has been flagged for law enforcement involvement, what is the process that takes place afterward? Does that content get sent to the FBI and then disseminated to state law enforcement?

Answer. We use a sophisticated machine learning tool to assess Facebook posts that may signal support for terrorist organizations. The tool produces a score indicating how likely it is that the post violates our policies. In some cases, we will automatically remove posts when the tool indicates with very high confidence that the post contains support for terrorism. But in most cases, we still rely on specialized reviewers to evaluate posts, and we use these scores to prioritize which posts our reviewers assess first.

We are careful not to reveal too much about our automated enforcement techniques, including the specific factors our machine learning prioritizes when evaluating content, because of adversarial shifts by terrorists. But we are seeing real gains as a result of this work: we've removed more than 26 million pieces of content related to global terrorist groups like ISIS and al-Qaeda in the last two years, 99 percent of which we proactively identified and removed before anyone reported it to us.

We reach out to law enforcement whenever we see a credible threat of imminent harm. We contact federal, state, or local law enforcement depending on the specific circumstances of a threat.

We have a long history of working successfully with the Department of Justice, the FBI, state and local law enforcement, and other government agencies to address a wide variety of threats to our platform, including terrorist threats. We have been able to provide support to authorities around the world that are responding to the threat of terrorism, including in cases where law enforcement has been able to disrupt attacks and prevent harm. We have strict processes in place to handle government requests we receive, 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.

Response to Written Questions Submitted by Hon. Amy Klobuchar to Monika Bickert

Online transparency and accountability is a top priority for me. In April, I sent a letter urging the Department of Homeland Security and Federal Bureau of Investigation to create a joint task force to combat election interference and the spread of misinformation, and this week I introduced legislation to create a Center at the Office of the Director of National Intelligence to coordinate the existing efforts of agencies and departments in combating foreign influence campaigns.

Question 1. Can you speak to the importance of joint efforts by agencies, the intelligence community, tech companies, and elections officials in combating the spread of misinformation?

Answer. We work closely with law enforcement, regulators, election officials, other technology companies, researchers, academics, and civil society groups to strengthen our platform against election interference and the spread of misinformation. This coordination is incredibly important—we can't do this alone, and we have worked to strengthen our relationships with government, outside experts, and other technology companies in order to share information and bolster our security efforts.

Our partnerships, as well as our own investigations, help us find and remove bad actors from Facebook. For example, ahead of the U.S. midterm elections on October 26, 2018, we took down 82 Pages, Groups, and accounts linked to Iran. And in the 48 hours ahead of the elections, we also got a tip from the FBI which allowed us to move quickly to take down a coordinated effort by foreign entities on Facebook and Instagram. Based on this tip, we quickly identified a set of accounts that appeared to be engaged in coordinated inauthentic behavior, which is banned on Facebook because we want people to be able to trust the connections they make on our services. So we immediately blocked these accounts and publicly announced what we found and the action we were taking. We also shared that information with the government and other companies to help them with their own investigations.

We're continuing to work closely with the FBI, the Department of Homeland Security (DHS), and other companies on ways to protect elections from interference on our platform. In September, security teams from Facebook and a number of technology companies met at Facebook with representatives from the FBI, the Office of the Director of National Intelligence (DNI), and DHS to further strengthen strategic collaboration regarding the security of the 2020 U.S. state, federal, and presidential elections.

We're constantly following up on thousands of leads of potential bad activity globally, including information shared with us by law enforcement, industry partners, and civil society groups, and insights from past takedowns. Over the past two years, we've seen that threats are rarely confined to a single platform or tech company. That's why we're working closely with our fellow tech companies to deal with the threats we have all seen during and beyond elections. A number of takedowns we have conducted and announced were in close collaboration with other tech platforms, security companies, and law enforcement. We also partner with the Atlantic Council's Digital Forensic Research Lab, Graphika, and other researchers and experts who provide additional analysis of the coordinated inauthentic behavior we identify, remove, and publicly share, including their behavior off-platform, across different Internet services.

Question 2. How could these efforts be improved?

Answer. Our partnerships have been immensely helpful, but it can be challenging to coordinate the operations and timing of these investigations. Timing is a key to our success, and the more entities involved, the harder it inevitably is to get everyone synced seamlessly. That's why it is so important to have open lines of communication with all of these partners so we can ensure we are all aligned, and that we take action pursuant to a timeline that best disrupts the adversary.

The current state of the law also does not make it easy to share information with other entities, which can hamper our partnerships in these areas. Clear authorities or liability protections that allow for sharing between companies and organizations would be helpful to reduce this friction.

Security is never finished, and it will take our continuous efforts to stay one step ahead of bad actors seeking to disrupt our elections. The better we can be at working together, the better we will do by our community.

In April, reports highlighted that the records of more than 540 million Facebook users were publicly exposed on Amazon's cloud service. One provision in the privacy legislation that I lead with Senator Kennedy requires that U.S. consumers are notified of breaches within 72 hours.

Question 3. What are your views on a Federal requirement to ensure that consumers are informed in a timely manner when their personal information has been compromised?

Facebook is committed to continuing to comply with breach notification laws. In some cases, we have gone beyond our legal obligation to notify consumers about instances where their personal information has been compromised, even when the law did not require us to do so.

At present, there are data breach laws in 50 states, with differing notification thresholds and time frames. We support a Federal data breach notification law that would create a consistent nationwide standard for breach notifications. To avoid notification fatigue and ensure consumer attention, legislation should establish clear rules that require notification of the breaches most likely to harm people. Organizations should notify people when there has been a breach affecting their personal information that could cause them a risk of significant harm (for example, identity theft, fraud, real-time location-tracking, or economic loss). Breaches of information that is encrypted, anonymized, or otherwise de-identified do not pose a risk of significant harm and would not require notification, unless the encryption key is also breached or information that would allow the breached information to be re-identified is also breached. Legislation should set forth factors that inform whether a data breach presents a risk of significant harm, such that notification to consumers should be required.

Response to Written Questions Submitted by Hon. Richard Blumenthal to Monika Bickert

NO HATE Act and Reporting

I have introduced legislation, the Jabara-Heyer NO HATE Act, which would help states implement and train officers in the National Incident-Based Reporting Systems. The NO HATE Act would also provide grants to states to better address hate crimes by training law enforcement, establish specialized units, create community relations programs, and run hate crime hotlines.

Question 1. Do you support the Jabara-Heyer NO HATE Act?

Answer. Hate has no place on Facebook, and we have strong relationships with law enforcement, academics, and experts to help us fight hate speech and hate-related violence on our platform. For example, we're partnering with Life After Hate, an organization founded by former violent extremists, to connect people who search for terms associated with white supremacy to resources focused on helping people leave behind hate groups. We also provide training to governments on how best to flag violating content, and we have portals for law enforcement to legally request data in ongoing and crisis scenarios.

We support providing resources to programs that deal with these issues, and we would be happy to discuss the specifics of the proposal with your office.

I understand that it has taken some time for Google and Facebook to establish reliable and timely channels to report threats made on your platform to the proper authorities. Mr. Slater testified that Google now has a strong relationship with the Northern California Regional Intelligence Center, who has been effective at quickly getting reports of threats into the right hands.

Question 2. Would you support adding measures to the Jabara-Heyer NO HATE Act to expand the NCRIC model of integrated threat reporting nationwide?

Answer. As discussed above, we would be happy to discuss the specifics of the proposal with your office.

When it comes to working with law enforcement, we reach out whenever we see a credible threat of imminent harm. We contact federal, state, or local law enforcement depending on the specific circumstances of a threat. We have a long history of working successfully with the Department of Justice, the FBI, state and local law enforcement, and other government agencies to address a wide variety of threats on our platform, including terrorist threats. We have been able to provide support to authorities around the world that are responding to the threat of terrorism, including in cases where law enforcement has been able to disrupt attacks and prevent harm. We also have law enforcement response teams available around the clock to respond to emergency requests.

Question 3. What steps would improve communications channels with law enforcement to make sure the right information gets into the right hands quickly?

Answer. Please see the above responses. As discussed, we work closely with law enforcement. Indeed, we have been able to provide support to authorities around the world that are responding to the threat of terrorism, including in cases where law enforcement has been able to disrupt attacks and prevent harm.

Amplification of 8Chan and Other Hate Sites

We have seen over this year that fringe sites are a breeding ground for racist and violent hate communities. However, extremists then use mainstream platforms to recruit and amplify their hate and ideologies to a larger audience. In particular, the site 8chan has had a repeated role in multiple mass shootings this year. The perpetrators of Christchurch mosque shootings, Poway synagogue shooting, and El Paso massacre each posted manifestos to 8chan before their attacks. It is also sites such as 8chan that facilitate campaigns of harassment and terrorism that target the victims of mass shootings, such as the Sandy Hook families. 8chan is currently offline after webhosting providers finally cut their ties after the El Paso shootings. However, 8chan's owner has said that he plans to revive the site as soon as this week.

Question 1. Has your company taken any steps to limit the spread of 8chan content, including the communities that hosted the manifestos of shooters, on your platforms?

Answer. For years, we've worked to block URLs when we identify that the content at the URL violates our policies when it is shared on Facebook. For example, we blocked links from 8chan and 4chan when the content shared violated our policies. And earlier this year, we started blocking any link that connects to 8chan's/pol/ board, where the Christchurch, El Paso, and Poway attacks were advertised and where a large amount of other hateful content has appeared.

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 our platforms. Since then, the consortium has grown and collaborates closely on critical initiatives focused on tech innovation, knowledge-sharing, and research. Most recently, we reached our 2019 goal of collectively contributing more than 200,000 hashes, or unique digital fingerprints, of known terrorist content into our shared database, enabling each of us to quickly identify and take action on potential terrorist content on our respective platforms.

Question 2. Please describe the specific steps you to restrict the amplification of 8chan and other violent sites on your platforms, including what sites you have taken action to restrict.

Answer. Please see the response to the previous question.

Testing of Consumer Platforms

Question 1. Please describe the process you use to test and evaluate new consumer facing products, including algorithms designed to promote forms of engagement. What methods are employed to assess the impact of these products on individuals and groups, both for an immediate and medium term response?

Answer. While the specific processes that we use to build and evaluate improvements to our services vary by product, we generally work to ensure that the features we build meet the needs and preferences of our community. Methods we use to hear from our community might include inviting people to sit down for one-on-one interviews, join focus groups, try new products and features, or keep diaries about their experiences with apps over time. We also invite large groups of people to take surveys, often via the Facebook app itself—indeed, tens of thousands of people opt into taking surveys every week. And with more than 2 billion people using Facebook every month, we have to carefully consider ways to ensure we are hearing from representative swaths of the community, all over the world.

Before new products and services are launched, company executives perform internal reviews, and we often release features slowly so that we can understand how people are using new features before they are available to everyone on Facebook.

Changes to our products and services that involve people's personal information are also reviewed through a cross-functional evaluation process overseen by the Chief Privacy Officer for Product, which involves our Chief Privacy Officer for Policy, legal compliance experts, and participants from other departments across the company. This process is a collaborative approach to privacy that seeks to promote strong privacy protections and sound decision-making at every stage of the product development process. Moreover, the new FTC Consent Order, which has not yet been finalized, will impose new, rigorous process and documentation requirements in this area. Our privacy program is responsible for reviewing product launches, major changes, and privacy-related bug fixes to products and features to ensure that privacy policies and procedures are consistently applied and that key privacy decisions are implemented. This approach has several key benefits:

First, it is designed to consider privacy early in the product development process. This allows us to consider the benefits that a feature is intended to have for people who use our services, how data will be used to deliver those benefits, and how we can build features from the ground up that include privacy protections to enable those benefits while protecting people's information and putting them in control.

Second, taking a cross-disciplinary approach to privacy encourages us to think about data protection as more than just a compliance exercise. Instead, we evaluate how to design privacy into the features that we build. We consider this from the perspective of things like designing interfaces that make data use intuitive, taking a consistent approach to privacy across our services, and building protections in how our software is engineered. Accordingly, while we scale our privacy review process depending on the complexity of a particular data use, reviews typically involve experts who evaluate proposed data practices from the perspective of multiple disciplines.

Facebook also undergoes ongoing privacy assessments to test the effectiveness of its privacy controls, which are conducted by an independent third-party professional pursuant to the procedures and standards generally accepted in the profession. Facebook's privacy program and related controls are informed by GAPP principles, which are considered industry-leading principles for protecting the privacy and security of personal information. We monitor the privacy program and update the controls as necessary to reflect evolving risks. And, under the new FTC Consent Order, we will continue to undergo these independent reviews on a biennial basis.

Question 2. Do you ever identify unintended consequences of such proposed products and then revise them or decide not to launch?

Answer. As described in the previous response, we test and evaluate new products for impacts of various kinds. If we determine that a negative impact outweighs the product's potential benefit, we do not launch the product.

Question 3. What testing and measurement methodologies are routinely used and how are the product evaluation teams selected? Please submit any criteria you have developed for new or revised data driven products or applications, including their intended impact, demographic reach, and revenue potential.

Answer. Please see the response to your Question 1.

Response to Written Questions Submitted by Hon. Tom Udall to Monika Bickert

Question 1. We cannot talk about mass violence without talking about the social and political climate that is dividing America. Most recently, content that demonizes and spreads hate against immigrant communities is proliferating across social media. This content is too often indistinguishable from social media posts from some elected representatives.

Facebook announced recently that it would exempt politicians from certain rules that prohibit hate speech, incite violence, or post fake news. How did Facebook come to the decision that any content posted by a political figure should be considered newsworthy, even if it clearly espouses hate, incites violence, or is designed to spread misinformation? Why does your commitment to protect users from harmful content end when the poster is a political figure?

Answer. Our recent work builds on an existing policy. Since 2016, we have assessed newsworthiness on a case-by-case basis, balancing the public interest value and the risk associated with the speech.

We've applied this policy to a range of organic content not limited in scope to politicians. We have, for example, allowed as newsworthy images that depict war or famine, or that attempt to raise awareness of issues like indigenous rights. The newsworthiness analysis does not apply to ads.

Even in the case of politicians' speech, no content is automatically deemed newsworthy. Newsworthiness requires a balancing test to make a determination of the public interest value versus the potential for harm. We take a number of factors into consideration, including country-specific context such as whether there is an election underway or the country is at war, as well as the speaker and subject matter of the speech, such as whether it relates to governance or politics.

In evaluating the risk of harm, we will consider the severity of the harm. Content that has the potential to incite violence poses a safety risk that we will take into account. And there are some types of violations—for example, the posting of terrorist propaganda or voter suppression—where the risk of harm will always override any public interest value.

When it comes to fact checking, we rely on third-party fact-checkers to help reduce the spread of false news and other types of viral misinformation, like memes or manipulated photos and videos. We don't believe, however, that it is 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. This 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 politicians are not subject to Facebook's third-party fact-checking program. We have had this policy on the books for over a year now, posted publicly on our site under our eligibility guidelines. This means that we will not send organic content or ads from politicians to our third-party fact-checking partners for review. However, when a politician shares previously debunked content, we will demote that content, display related information from fact-checkers, and reject its inclusion in advertisements.

Question 2. Knowing that the problem of extremism and mass violence extends beyond the screen, I would like you to describe your partnerships with communities and organizations around the country to fight against extremism and hate. What are you doing to promote their voices on your platforms? Moreover, what makes them effective?

Answer. We are proud of the work we have done to make Facebook a hostile place for those committed to acts of violence. We understand, however, that simply working to keep violence off Facebook is not an adequate solution to the problem of online extremism and violence, particularly because bad actors can leverage a variety of platforms and operate offline as well. 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 counterspeech campaigns from students in 75 countries, engaged over 6,500 students, and reached over 200 million people. We're also partnering with Life After Hate, an organization founded by former violent extremists, to connect people who search for terms associated with white supremacy to resources focused on helping people leave behind hate groups.

And we are continuing our work with the Global Internet Forum to Counter Terrorism (GIFCT), an endeavor that focuses on fighting terrorism and extremism through knowledge sharing, support for counterterrorism work, and technical cooperation. In September, GIFCT released a digital Campaign Toolkit produced by the Institute for Strategic Dialogue that instructs NGOs running online counterspeech programs in best practices for utilizing a range of digital platforms. Just as bad actors utilize a range of platforms to get their message out, so must counterspeech practitioners. For more information, please see *https://www.campaigntoolkit.org*.

GIFCT recently announced that it will become an independent organization led by an Executive Director and supported by dedicated technology, counterterrorism, and operations teams. Evolving and institutionalizing GIFCT's structure from a consortium of member companies will build on our early achievements and deepen industry collaboration with experts, partners, and government stakeholders—all in an effort to thwart increasingly sophisticated efforts by terrorists and violent extremists to abuse digital platforms.

Question 3. We are entering another election year and we know that foreign actors have amplified divisive rhetoric on social media and, in some cases, orchestrated actual protests. What specific actions are you taking to prepare for 2020 to prevent Russia and other foreign actors from trying to inflame racial and political tensions through social media?

We have a responsibility to stop abuse and election interference on our platform. That's why we've made significant investments since 2016 to better identify new threats, close vulnerabilities, and reduce the spread of viral misinformation and fake accounts.

Combating Inauthentic Behavior

Over the last three years, we've worked to identify new and emerging threats and remove coordinated inauthentic behavior across our apps. In the past year alone, we've taken down over 50 networks worldwide, many ahead of major democratic elections. As part of our effort to counter foreign influence campaigns, most recently we removed three networks of accounts, Pages, and Groups on Facebook and Instagram for engaging in foreign interference. These manipulation campaigns originated in Russia and targeted a number of countries in Africa. We have identified these manipulation campaigns as part of our internal investigations into suspected Russia-linked inauthentic behavior in the region.

We took down these networks based on their behavior, not the content they posted. In each case, the people behind this activity coordinated with one another and used fake accounts to misrepresent themselves, and that was the basis for our action. We have shared our findings with law enforcement and industry partners. More details can be found at *https://newsroom.fb.com/news/2019/10/removing-more-coordinated-inauthentic-behavior-from-Russia.* As we've improved our ability to disrupt these operations, we've also built a deeper understanding of different threats and how best to counter them. We investigate and enforce against any type of inauthentic behavior.

Protecting the Accounts of Candidates, Elected Officials, and Their Teams

We also recently launched Facebook Protect to further secure the accounts of elected officials, candidates, their staff, and others who may be particularly vulnerable to targeting by hackers and foreign adversaries. As we've seen in past elections, they can be targets of malicious activity. However, because campaigns are generally run for a short period of time, we do not always know who these campaign-affiliated people are, making it harder to help protect them.

Page admins can enroll their organization's Facebook and Instagram accounts in Facebook Protect and invite members of their organization to participate in the program as well. Participants will be required to turn on two-factor authentication, and their accounts will be monitored for hacking, such as login attempts from unusual locations or unverified devices. And, if we discover an attack against one account, we can review and protect other accounts affiliated with that same organization that are enrolled in our program. You can find more information about Facebook Protect at *https://www.facebook.com/gpa/facebook-protect.*

Making Pages More Transparent

We want to make sure people are using Facebook authentically and that they understand who is speaking to them. Over the past year, we've taken steps to ensure Pages are authentic and more transparent by showing people the Page's primary country location, whether the Page has merged with other Pages, and information about the organization that owns the Page. This gives people more context on the Page and makes it easier to understand who is behind it.

Labeling State-Controlled Media

We want to help people better understand the sources of news content they see on Facebook so they can make informed decisions about what they are reading. We will soon begin labeling media outlets that are wholly or partially under the editorial control of their government as state-controlled media. This label will be on both their Page and in our Ad Library. We will hold these Pages to a higher standard of transparency because they combine the opinion-making influence of a media organization with the strategic backing of a state.

Making it Easier to Understand Political Ads

Throughout this year, we've been expanding our work around the world to increase authenticity and transparency around political advertising because we know how important it is that people understand who is publishing the ads that they see. We have now launched our publicly searchable Ad Library in over 190 countries and territories. We allow advertisers to be authorized to purchase political ads and we give people more information about ads that concern social issues, elections, or politics. We require the use of these transparency tools in over 50 jurisdictions, and we make them available for voluntary use in over 140 others, to provide the option of greater transparency and accountability.

We have added a variety of features to our ads transparency tools to help journalists, lawmakers, researchers, and others learn more about the ads they see, including information about how much candidates have spent on ads. And soon we will also begin testing a new database with researchers that will enable them to quickly download the entire Ad Library, pull daily snapshots, and track day-to-day changes.

More Resources for Rapid Response for Elections

We have set up regional operations centers focused on election integrity in California, Dublin, and Singapore. These hubs allow our global teams to better work across regions in the run-up to elections and further strengthen our coordination and response time between staff in Menlo Park and in-country. These teams add a layer of defense against fake news, hate speech, and voter suppression and work cross-functionally with our threat intelligence, data science, engineering, research, community operations, legal, and other teams.

Preventing the Spread of Viral Misinformation

On Facebook and Instagram, we work to keep confirmed misinformation from spreading. For example, we reduce its distribution so fewer people see it—on Instagram, we remove it from Explore and hashtags, and on Facebook, we reduce its distribution in News Feed. On Instagram, we also make content from accounts that repeatedly post misinformation harder to find, for example by filtering content from that account from Explore and hashtag pages. And on Facebook, if Pages, domains, or Groups repeatedly share misinformation, we'll continue to reduce their overall distribution, and we'll place restrictions on the Page's ability to advertise and monetize.

Over the coming weeks, content across Facebook and Instagram that has been rated false or partly false by a third-party fact-checker will start to be more prominently labeled so that people can better decide for themselves what to read, trust, and share. Labels will be shown on top of false and partly false photos and videos, including on top of Stories content on Instagram, and will link out to the assessment from the fact-checker.

Much like we do on Facebook when people try to share known misinformation, we are also introducing a new pop-up that will appear when people attempt to share posts on Instagram that include content that has been debunked by third-party fact-checkers.

In addition to clearer labels, we are 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.

Fighting Voter Suppression and Intimidation

Attempts to interfere with or suppress voting undermine our core values as a company, and we work proactively to remove this type of harmful content. Ahead of the 2018 midterm elections, we extended our voter suppression and intimidation policies to prohibit:

- Misrepresentation of the dates, locations, times, and methods for voting or voter registration (*e.g.,* "Vote by text!");
- 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 (*e.g.,* "If you voted in the primary, your vote in the general election won't count."); and
- Threats of violence relating to voting, voter registration, or the outcome of an election.

We remove this type of content regardless of who it's coming from. Ahead of the midterm elections, our Elections Operations Center removed more than 45,000 pieces of content that violated these policies—more than 90 percent of which our systems detected before anyone reported the content to us.

In advance of the U.S. 2020 elections, we're implementing additional policies and expanding our technical capabilities on Facebook and Instagram to protect the integrity of the election. Following up on a commitment we made in the civil rights audit report released in June, we have now implemented our policy banning paid advertising that suggests voting is useless or meaningless or advises people not to vote. In addition, our systems are now more effective at proactively detecting and removing this harmful content. We use machine learning to help us quickly identify potentially incorrect voting information and remove it.

We are also continuing to expand and develop our partnerships to provide expertise on trends in voter suppression and intimidation, as well as early detection of violating content. This includes working directly with secretaries of state and election directors to address localized voter suppression that may only be occurring in a single state or district. This work will be supported by our Elections Operations Center during both the primary and general elections.

Helping People Better Understand What They See Online

Part of our work to stop the spread of misinformation is helping people spot it for themselves. That's why we partner with organizations and experts in media literacy. We recently announced an initial investment of $2 million to support projects that empower people to determine what to read and share—both on Facebook and elsewhere.

These projects range from training programs to help ensure the largest Instagram accounts have the resources they need to reduce the spread of misinformation, to expanding a pilot program that brings together senior citizens and high school students to learn about online safety and media literacy, to public events in local venues like bookstores, community centers, and libraries in cities across the country. We're also supporting a series of training events focused on critical thinking among first-time voters.

In addition, we're including a new series of media literacy lessons in our Digital Literacy Library. These lessons are drawn from the Youth and Media team at the Berkman Klein Center for Internet & Society at Harvard University, which has made them available for free worldwide under a Creative Commons license. The lessons, created for middle and high school educators, are designed to be interactive and cover topics ranging from assessing the quality of the information online to more technical skills like reverse image search.

Question 4. Regarding the shared industry database of hashes linked to content that promotes terrorism; I would like to understand the thresholds for including certain content in the database. Who makes the decision to include content in that database and how is that decision made? What percent of that database concerns white nationalist or other domestic extremist content?

Answer. Facebook's internal terrorism definition applies to a wide range of terrorist actors, regardless of ideology or designation by governments or intergovernmental entities. We have designated more than 200 white supremacist organizations under our broader Dangerous Organizations policy. For the purposes of the hash-sharing database, GIFCT uses the UN's Consolidated Sanctions List to identify groups for which we will share hashes of any terrorist-related content or propaganda found. Following the Christchurch attack, we have also developed a Content Incident Protocol (CIP), which enables companies to share hashes related to propaganda produced by attackers during a terrorist attack. The CIP was deployed for the first time after the October 9 attack in Halle, Germany.

Companies also agreed upon a basic taxonomy to describe the type of content ingested into the hash-sharing database. The taxonomy includes the following labels that are applied to the content when a company adds hashes to the shared database:

- *Imminent Credible Threat:* A public posting of a specific, imminent, credible threat of violence toward non-combatants and/or civilian infrastructure.
- *Graphic Violence Against Defenseless People:* The murder, execution, rape, torture, or infliction of serious bodily harm on defenseless people (prisoner exploitation, obvious non-combatants being targeted).
- *Glorification of Terrorist Acts:* Content that glorifies, praises, condones, or celebrates attacks after the fact.
- *Recruitment and Instruction:* Materials that seek to recruit followers, give guidance, or instruct them operationally.
- *New Zealand Perpetrator Content:* The GIFCT set a new precedent in the wake of the New Zealand terrorist attack. Due to the virality and cross-platform spread of the attacker's manifesto and attack video, and because New Zealand authorities deemed all manifesto and attack video content illegal, the GIFCT created a crisis bank to mitigate the spread of this content.

GIFCT categorizes the content ingested based on these categories. As of July 2019, the breakdown of the content in the database is as follows:

- Imminent Credible Threat: 0.4 percent
- Graphic Violence Against Defenseless People: 4.8 percent
- Glorification of Terrorist Acts: 85.5 percent
- Radicalization, Recruitment, Instruction: 9.1 percent

New Zealand Perpetrator Content: 0.6 percent

Response to Written Questions Submitted by Hon. Jacky Rosen to Monika Bickert

Question 1. The challenge for social media platforms prohibiting certain types of behavior on their sites is creating clear and concise rules for users to comply. Offensive conduct isn't a static issue, and as technology has evolved, so have our definitions of what constitutes abusive behavior such as cyberbullying and misinformation campaigns.

- Can you explain to us how your companies come up with rules regarding hateful speech and how those rules have evolved? What are your guidelines for determining when charged rhetoric crosses the line into becoming hate speech? For example, how do you determine if rhetoric is anti-Semitic?

Answer. We do not allow hate speech on Facebook because it creates an environment of intimidation and exclusion and in some cases may promote real-world violence.

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, caste, sex, gender, gender identity, and serious disease or disability. We also provide some protections for immigration status. We define "attack" as violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation. We separate attacks into three tiers of severity, as described in our published Community Standards. For more information, please see *https://www.facebook.com/communitystandards/hate_speech.*

- How closely do you work with outside groups, researchers, and users to come up with definitions of what constitutes hate and abusive speech and policies to deal with ambiguous cases? For instance, have you worked with the Anti-Defamation League or other groups combating hate when determining guidelines?

Answer. Facebook has partnerships with a broad range of U.S. and international NGOs, academics, and experts who study organized hate groups. These academics and experts share information with Facebook on how organizations are adapting to social media and give feedback on how Facebook might better tackle these problems.

Question 2. With almost three and a half billion social media users worldwide—and one million users joining every day—social media platforms have turned to a mix of machine learning and human moderators to detect and take down hate speech, terrorist propaganda, cyber-bullying, and disinformation. Machine learning can be a useful tool in identifying objectionable content quickly, preventing it from spreading. However, there are concerns about its ability to understand the context of text or images, and the length of time it takes to train systems with new data to recognize objectionable content.

- Can you give us an estimate of how many content moderation decisions are made by your machine learning systems? And can you provide an estimated error rate for content flagged by machine learning?

Answer. We don't have an either-or approach to reviewing content. 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 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.

- Are there instances where machine learning is more effective in flagging certain content than others? Does the error rate change significantly from one type of content to another?

Answer. AI tools lend themselves toward identifying certain content more easily than others. For example, we are better able to enforce our nudity policies with automated tools than we are hate speech, due to the linguistic and cultural nuances involved. One area in which we have made significant progress is the detection of terrorist content. We proactively detect 99 percent of the ISIS-and Al Qaeda-related content that we remove before someone reports it. And we are committed to continuing to improve our technology across different types of content.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. SHELLEY MOORE CAPITO TO NICK PICKLES

Question 1. I applaud Twitter's engagement and collaboration with the Global Internet Forum to Counter Terrorism (GIFCT). As you mentioned in your testimony, Twitter has already started partnering with smaller tech companies to share best practices and continues to partner with additional companies. What are some of the smaller community groups Twitter has been working with and what are some of the best practices you plan on sharing for combatting extremism?

Answer. Collaboration with our industry peers and civil society is critically important to addressing common threats from terrorism globally. In June 2017, we launched the Global Internet Forum to Counter Terrorism (the "GIFCT"), a partnership among Twitter, YouTube, Facebook, and Microsoft.

The GIFCT facilitates, among other things: information sharing; technical cooperation; and, research collaboration, including with academic institutions. In September 2017, the members of the GIFCT announced a significant financial commitment to support research on terrorist abuse of the Internet and how governments, tech companies, and civil society can respond effectively. Our goal is to establish a network of experts that can develop platform-agnostic research questions and analysis that consider a range of geopolitical contexts.

Technological collaboration is a key part of GIFCT's work. In the first two years of GIFCT, two projects have provided technical resources to support the work of members and smaller companies to remove terrorist content.

First, the shared industry database of "hashes"—unique digital "fingerprints"—for violent terrorist propaganda now has more than 100,000 hashes. The database allows a company that discovers terrorist content on one of its sites to create a digital fingerprint and share it with the other companies in the forum, who can then use those hashes to identify such content on their services or platforms, review against their respective policies and individual rules, and remove matching content as appropriate or block extremist content before it is posted.

Second, a year ago, Twitter began working with a small group of companies to test a new collaborative system. Because Twitter does not allow files other than photos or short videos to be uploaded, one of the behaviors we saw from those seeking to promote terrorism was to post links to other services where people could access files, longer videos, PDFs, and other materials. Our pilot system allows us to alert other companies when we removed an account or Tweet that linked to material that promoted terrorism hosted on their service. This information sharing ensures the hosting companies can monitor and track similar behavior, taking enforcement action pursuant with their individual policies. This is not a high-tech approach, but it is simple and effective, recognizing the resource constraints of smaller companies.

Based on positive feedback, the partnership has now expanded to 12 companies and we have shared more than 12,000 unique URLs with these services. Every time a piece of content is removed at source, it means any link to that source—wherever it is posted—will no longer be operational.

We are eager to partner with additional companies to expand this project, and we look forward to building on our existing partnerships in the future.

Separately, Twitter provides training to civil society groups around the globe that work on preventing and combating violent extremism in their communities. These trainings aim to help credible organizations amplify their voices using Twitter tools and cover a wide range of best practices, which are summarized in our NGO training handbook. Twitter does not advise on the specifics of the message, as these partners are best placed to craft their own authentic content. We are happy to provide a copy of the handbook upon request.

In addition, Twitter has helped amplify the voices and reach of these organizations through in-kind assistance in the form of donated advertising credit, both on Twitter and offline. Recently, for example, Twitter donated advertising space in New York City to Parents for Peace, an NGO founded and run by former extremists and families impacted by extremism which aims to prevent radicalization.

Question 2. This Committee has held a number of hearings on the rise and importance of artificial intelligence (AI) in today's digital economy. AI has been invaluable in collecting and sorting massive amounts of data. In the case of today's hearing, AI has become critical in order to identify radicalization and terrorist threats. Each company has identified key tools each company uses in identifying bad actors on your platforms, but machine learning being one of the most critical. What factors are given priority when determining radicalized or terrorist content?

a. You also mention the importance of human expertise in determining more nuanced cases. When does human expertise step in after AI has identified or flags content?

b. After content has been flagged for law enforcement involvement, what is the process that takes place afterward? Does that content get sent to the FBI and then disseminated to state law enforcement?

Answer. Twitter's philosophy is to take a behavior-led approach, utilizing a combination of machine learning and human review to prioritize reports and improve the health of the public conversation. That is to say, we increasingly look at how accounts behave before we look at the content they are posting. This is how we seek to scale our efforts globally and leverage technology even where the language used is highly context specific. Twitter employs content detection technology to identify potentially abusive content on the service, along with allowing users to report content to us either as an individual or as a bystander.

We suspended more than 1.5 million accounts for violations related to the promotion of terrorism between August 1, 2015, and December 31, 2018. In 2018, a total of 371,669 accounts were suspended for violations related to promotion of terrorism. We continue to see more than 90 percent of these accounts suspended through proactive measures.

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 allows 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.

The long term challenge for industry is the availability and sharing of training data for AI and machine learning models. Good progress has been made in cross-industry collaboration on a number of fronts, but this is an area where more can be done.

We have well-established relationships with law enforcement agencies, and we look forward to continued cooperation with them on these issues, as often only they have access to information critical to our joint efforts to stop bad faith actors. The threat we face requires extensive partnership and collaboration with our government partners and industry peers. We have continuous internal coverage to address requests from law enforcement around the world and have a portal to swiftly handle law enforcement requests rendered by appropriate legal process.

If we have a good faith belief that there is an imminent threat of death or serious physical harm to an identifiable person or group, and we have information that we believe is relevant to mitigating that threat, we share such information with law enforcement. We become aware of such threats through reports to our content moderation team, or through an Emergency Request submitted by law enforcement.

Twitter does not have a role in how information is shared between the FBI and other law enforcement entities.

Response to Written Questions Submitted by Hon. Richard Blumenthal to Nick Pickles

NO HATE Act and Reporting

I have introduced legislation, the Jabara-Heyer NO HATE Act, which would help states implement and train officers in the National Incident-Based Reporting Systems. The NO HATE Act would also provide grants to states to better address hate crimes by training law enforcement, establish specialized units, create community relations programs, and run hate crime hotlines.

Question. Do you support the Jabara-Heyer NO HATE Act?

Answer. Twitter believes that successfully combating violent extremism requires a whole of society approach, including at the grassroots and community-based level. Twitter has a positive working relationship with law enforcement agencies, that play a key role in preventing and addressing violent extremism. Twitter supports efforts that seek to strengthen the ability of both community groups and law enforcement agencies to more effectively address violent extremism.

I understand that it has taken some time for Google and Facebook to establish reliable and timely channels to report threats made on your platform to the proper authorities. Mr. Slater testified that Google now has a strong relationship with the Northern California Regional Intelligence Center, who has been effective at quickly getting reports of threats into the right hands.

Question 1. Would you support adding measures to the Jabara-Heyer NO HATE Act to expand the NCRIC model of integrated threat reporting nationwide?

Answer. Twitter supports efforts that seek to strengthen the ability of law enforcement agencies to more effectively address violent extremism within appropriate legal frameworks.

Question 2. What steps would improve communications channels with law enforcement to make sure the right information gets into the right hands quickly?

Answer. Twitter has well-established relationships with law enforcement agencies, and we look forward to continued cooperation with them on these issues, as often only they have access to information critical to our joint efforts to stop bad faith actors. The threat we face requires extensive partnership and collaboration with our government partners and industry peers. We have continuous internal coverage to address reports from law enforcement around the world and have a portal to swiftly handle law enforcement requests rendered by appropriate legal process.

We encourage law enforcement and other government agencies to consider whether it is possible to move more quickly to declassify information that may be useful to industry in assessing and reacting to the changing nature of bad actors.

Amplification of 8Chan and Other Hate Sites

We have seen over this year that fringe sites are a breeding ground for racist and violent hate communities. However, extremists then use mainstream platforms to recruit and amplify their hate and ideologies to a larger audience. In particular, the site 8chan has had a repeated role in multiple mass shootings this year. The perpetrators of Christchurch mosque shootings, Poway synagogue shooting, and El Paso massacre each posted manifestos to 8chan before their attacks. It is also sites such as 8chan that facilitate campaigns of harassment and terrorism that target the victims of mass shootings, such as the Sandy Hook families. 8chan is currently offline after webhosting providers finally cut their ties after the El Paso shootings. However, 8chan's owner has said that he plans to revive the site as soon as this week.

Question 1. Has your company taken any steps to limit the spread of 8chan content, including the communities that hosted the manifestos of shooters, on your platforms?

Answer. Twitter is committed to improving our ability to stop the rapid spread of violent extremist content. We are implementing a number of actions from lessons we learned from the Christchurch attack.

For example, the distribution of media immediately after the Christchurch attack was manifestly different from how ISIS and other terrorist groups had historically operated. These changes in the wider threat environment require a renewed approach and a focus on immediate crisis response.

In the immediate hours after the Christchurch shootings, an array of individuals sought continuously to re-upload the content created by the attacker, both the video and his manifesto, including this same content hosted on third party services.

The Twitter rules make clear that we do not allow material to be shared that threatens violence against an individual or group of people, the glorification of violence or the promotion of terrorism. We regard manifestos as falling under this rule.

As such, we will take action to limit the ability of individuals to share materials, including raw video files and manifestos, wherever they are posted, including on third party services.

Question 2. Please describe the specific steps you to restrict the amplification of 8chan and other violent sites on your platforms, including what sites you have taken action to restrict.

Answer. We are taking a number of steps, many in collaboration with our GIFCT peers, to tackle the challenge of violent content that may spread to our platform from another location on the Internet, including but not limited to 8chan.

In addition to our commitment to the Christchurch Call, Twitter and other leading websites recently voluntarily committed to the following five distinct actions:

1. Terms of Service. First, we committed to updating our terms of use, community standards, codes of conduct, and acceptable use policies to expressly prohibit the distribution of terrorist and violent extremist content. We believe this is important to establish baseline expectations for users and to articulate a clear basis for removal of this content from our platforms and services and suspension or closure of accounts distributing such content.
2. User Reporting of Terrorist and Violent Extremist Content. Second, we committed to establishing one or more methods within our online platforms and services for users to report or flag inappropriate content, including terrorist and violent extremist content. We will ensure that the reporting mechanisms are clear, conspicuous, and easy to use, and provide enough categorical granu-

larity to allow us to prioritize and act promptly upon notification of terrorist or violent extremist content.

3. Enhancing Technology. Third, we committed to continuing to invest in technology that improves our capability to detect and remove terrorist and violent extremist content online, including the extension or development of digital fingerprinting and AI-based technology solutions.
4. Livestreaming. Fourth, we committed to identifying appropriate checks on livestreaming, aimed at reducing the risk of disseminating terrorist and violent extremist content online. These may include enhanced vetting measures (such as streamer ratings or scores, account activity, or validation processes) and moderation of certain livestreaming events where appropriate. Checks on livestreaming will necessarily be tailored to the context of specific livestreaming services, including the type of audience, the nature or character of the livestreaming service, and the likelihood of exploitation.
5. Transparency Reports. Finally, we committed to publishing on a regular basis transparency reports regarding detection and removal of terrorist or violent extremist content on our online platforms and services and ensuring that the data is supported by a reasonable and explainable methodology.

In addition, all members of the GIFCT committed to the following four collaborative actions:

1. Share Technology Development. We committed to working collaboratively across industry, governments, educational institutions, and NGOs to develop a shared understanding of the contexts in which terrorist and violent extremist content is published and to improve technology to detect and remove terrorist and violent extremist content more effectively and efficiently. This will include:
    - Work to create robust shared data sets to accelerate machine learning and AI and sharing insights and learnings from the data.
    - Development of open source or other shared tools to detect and remove terrorist or violent extremist content.
    - Enablement of all companies, large and small, to contribute to the collective effort and to better address detection and removal of this content on their platforms and services.
2. Crisis Protocols. We also committed to working collaboratively across industry, governments, and NGOs to create a protocol for responding to emerging or active events, on an urgent basis, so relevant information can be quickly and efficiently shared, processed, and acted upon by all stakeholders with minimal delay. This includes the establishment of incident management teams that coordinate actions and broadly distribute information that is in the public interest.
3. Education. Third, we committed to working collaboratively across industry, governments, educational institutions, and NGOs to help understand and educate the public about terrorist and extremist violent content online. This education includes reminding users about how to report or otherwise not contribute to the spread of this content online.
4. Combating Hate and Bigotry. Finally, we committed to working collaboratively across industry to attack the root causes of extremism and hate online. This includes providing greater support for relevant research—with an emphasis on the impact of online hate on offline discrimination and violence—and supporting capacity and capability of NGOs working to challenge hate and promote pluralism and respect online.

Testing of Consumer Platforms

Question 1. Please describe the process you use to test and evaluate new consumer facing products, including algorithms designed to promote forms of engagement. What methods are employed to assess the impact of these products on individuals and groups, both for an immediate and medium term response?

Answer. We want Twitter to provide a useful, relevant experience to all people using our service. With hundreds of millions of Tweets per day on Twitter, we have invested heavily in building systems that organize content on Twitter alongside tools for individuals to control their own experience. At the core of the Twitter service is the individual's choice of which accounts to follow, and thus, are shown in their home timeline. We want to help our customers to have an informative and enjoyable experience on Twitter by doing some of the work to surface content of interest.

With 335 million people using Twitter every month, in dozens of languages and countless cultural contexts, we rely upon machine-learning algorithms to help us organize content. Twitter uses a range of algorithms and behavioral signals to determine how Tweets are organized and presented in the home timeline, conversations, and search based on relevance to individuals. Individuals can control themselves whether to see their home timeline without any algorithmic processing or instead with our suggested ranking. We are constantly iterating our product to provide the best possible experience to all people using our service.

A wide range of teams are involved in assessing potential product and policy changes, from a broad range of perspectives. This includes work to understand how algorithms are functioning.

Question 2. Do you ever identify unintended consequences of such proposed products and then revise them or decide not to launch?

Answer. Yes. We are constantly iterating our product to provide the best possible experience to all people using our service. Our teams implement rigorous processes to think through all aspects of potential product changes to ensure they respect consumer privacy and further our goal of fostering healthy public conversation.

Question 3. What testing and measurement methodologies are routinely used and how are the product evaluation teams selected? Please submit any criteria you have developed for new or revised data driven products or applications, including their intended impact, demographic reach, and revenue potential.

Answer. We use a range of criteria to evaluate the success of our work. For example, in April 2019, we published a range of metrics that demonstrate the different ways we seek to measure our progress.

- 38 percent of abusive content that's enforced is surfaced proactively to our internal teams for review instead of relying on external reports from people on Twitter.
- 16 percent fewer abuse reports after an interaction from an account the external reporter doesn't follow.
- 100,000 accounts suspended for creating new accounts after a previous suspension during January–March 2019—a 45 percent increase from the same time last year.
- 60 percent faster response to appeals requests with our new in-app appeal process.
- 3 times more abusive accounts suspended within 24 hours after a report compared to the same time last year.
- 2.5 times more private information removed with a new, easier reporting process.

We continue to work with outside partners to develop a framework for measuring healthy conversation, following an international call for proposals. This will not be a quick or simple process, but we are investing in the long-term health of the public conversation online.

Response to Written Questions Submitted by Hon. Tom Udall to Nick Pickles

Question 1. We cannot talk about mass violence without talking about the social and political climate that is dividing America. Most recently, content that demonizes and spreads hate against immigrant communities is proliferating across social media. This content is too often indistinguishable from social media posts from some elected representatives. How does your company define hate speech?

Answer. Twitter has a policy against hateful conduct. Under this policy, people on Twitter are not permitted to promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. We also do not allow accounts whose primary purpose is inciting harm toward others on the basis of these categories.

We do not allow individuals to use hateful images or symbols in their profile image or profile header. Individuals on the platform are not allowed to use the username, display name, or profile bio to engage in abusive behavior, such as targeted harassment or expressing hate toward a person, group, or protected category.

Under this policy, we take action against behavior that targets individuals or an entire protected category with hateful conduct.

When determining the penalty for violating our hateful conduct policy, we consider a number of factors including, but not limited to the severity of the violation and an individual's previous record of rule violations. For example, we may ask someone to remove the violating content and serve a period of time in read-only mode before they can Tweet again. Subsequent violations will lead to longer read-only periods and may eventually result in permanent account suspension. If an account is engaging primarily in abusive behavior, or is deemed to have shared a violent threat, we will permanently suspend the account upon initial review.

Question 1a. And how do you address situations when content that meets that definition comes from a political leader?

Answer. When it comes to the actions of world leaders on Twitter, we recognize that this is largely new ground and with important implications. We understand the desire for our decisions to be "yes/no" binaries, but it's not that simple. The actions we take and policies we develop will help set precedents around online speech and we owe it to the people we serve to be deliberate and considered in what we do.

Twitter's mission is to provide a forum that enables people to be informed and to engage their leaders directly. We also have a responsibility to the people who use Twitter to better explain why we make the decisions we make, which we will do here.

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.

Direct interactions with fellow public figures, comments on political issues of the day, or foreign policy saber rattling on economic or military issues are generally not in violation of the Twitter Rules. However, if a Tweet from a world leader does violate the Twitter Rules but there is a clear public interest value to keeping the Tweet on the service, we may place it behind a notice that provides context about the violation and allows people to click through should they wish to see the content. We announced this in June 2019.

Our goal is to enforce our rules judiciously and impartially. In doing so, we aim to provide more insight into our enforcement decision-making, to serve public conversation, and protect the public's right to hear from their leaders and to hold these same leaders to account.

Question 2. Knowing that the problem of extremism and mass violence extends beyond the screen, I would like you to describe your partnerships with communities and organizations around the country to fight against extremism and hate. What are you doing to promote their voices on your platforms? And what makes them effective?

Answer. Twitter works around the globe to support civil society voices and promote positive messages. Twitter provides regular trainings to local, credible groups on five continents on how to amplify their content using our tools. In addition, we have provided pro-bono advertising to groups to enable their messages to reach millions of people. When we at Twitter talk about the health of the public conversation, we see the principles of civility, empathy, and mutual respect as foundational to our work. We will not solve problems by removing content alone. We should not underestimate the power of open conversation to change minds, perspectives, and behaviors.

Question 3. We are entering another election year and we know that foreign actors have amplified divisive rhetoric on social media and, in some cases, orchestrated actual protests. What specific actions are you taking to prepare for 2020 to prevent Russia and other foreign actors from trying to inflame racial and political tensions through social media?

Answer. The public conversation occurring on Twitter is never more important than during elections, the cornerstone of democracy. Any attempts to undermine the integrity of our service is antithetical to our fundamental values and undermines the core tenets of freedom of expression.

We remain vigilant about malicious foreign efforts to manipulate and divide people in the United States and throughout the world, including through the use of foreign disinformation campaigns that rely in certain instances upon the use of deepfakes. In April 2019, we issued a new Twitter policy regarding election integrity governing different categories of manipulative behavior and content related to elections. First, an individual cannot share false or misleading information about how to participate in an election. This includes but is not limited to misleading information about how to vote or register to vote, requirements for voting, including identification requirements, and the official announced date or time of an election. Second, an individual cannot share false or misleading information intended to intimidate or dissuade voters from participating in an election. This includes but is not limited to misleading claims that polling places are closed, that polling has ended, or other misleading information relating to votes not being counted.

Third, we do not allow misleading claims about police or law enforcement activity related to polling places or elections, long lines, equipment problems, voting procedures or techniques that could dissuade voters from participating in an election, and threats regarding voting locations. Finally, we do not allow the creation of fake accounts which misrepresent their affiliation, or share content that falsely represents its affiliation to a candidate, elected official, political party, electoral authority, or government entity.

If we see the use of any manipulated content to spread misinformation in violation of our policies governing election integrity, we will remove that content.

Additionally, we make available a unique comprehensive archive of removed Tweets and media associated with suspected state-backed information operations. Our industry peers, academics and policymakers can leverage the range of signals we publish including links, media, and account indicators. The data sets we have published so far include more than 30 million Tweets and more than one terabyte of media.

Further, 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 well-established relationships with law enforcement agencies active in this arena, including the U.S. Federal Bureau of Investigation's Foreign Influence Task Force and the U.S. Department of Homeland Security's Election Security Task Force. We look forward to continued cooperation with federal, state, and local government agencies on election integrity issues because in certain circumstances only they have access to information critical to our joint efforts to stop bad faith actors.

On Election Day in the 2018 U.S. midterms, Twitter participated virtually in an operations center convened by the U.S. Department of Homeland Security. The operations center also convened officials from the U.S. Department of Justice, the Federal Bureau of Investigation, and the Office of the Director of National Intelligence, in addition to federal, state, local, and private sector partners. In the lead up to Election Day, and throughout the course of the day itself, Twitter remained in constant contact with officials throughout all levels of government. We plan to do the same in the 2020 U.S. election period.

We also worked in close collaboration with the National Association of Secretaries of State (NASS) and the National Association of State Election Directors (NASED). Founded in 1904, NASS is the Nation's oldest, nonpartisan professional organization for public officials, and is open to secretaries of states and lieutenant governors in the 50 states, D.C. and territories. In February 2019, Twitter participated in a panel discussion convened by NASS on the Role of Social Media in Democracy and their New Voters Forum, broadcast on C-Span.

Question 4. Regarding the shared industry database of hashes linked to content that promotes terrorism; I would like to understand the thresholds for including certain content in the database. Who makes the decision to include content in that database and how is that decision made? What percent of that database concerns white nationalist or other domestic extremist content?

Answer. Collaboration with our industry peers and civil society is critically important to addressing common threats from terrorism globally. In June 2017, we launched the Global Internet Forum to Counter Terrorism (the "GIFCT"), a partnership among Twitter, YouTube, Facebook, and Microsoft.

The GIFCT facilitates, among other things: information sharing; technical cooperation; and, research collaboration, including with academic institutions. In September 2017, the members of the GIFCT announced a significant financial commitment to support research on terrorist abuse of the Internet and how governments, tech companies, and civil society can respond effectively. Our goal is to establish a network of experts that can develop platform-agnostic research questions and analysis that consider a range of geopolitical contexts.

Technological collaboration is a key part of GIFCT's work. In the first two years of GIFCT, two projects have provided technical resources to support the work of members and smaller companies to remove terrorist content.

As reported in GIFCT's first transparency report, published in July 2019, the GIFCT Hash Sharing Consortium has reached over 200,000 unique pieces of terrorist content. Companies often have slightly different definitions on "terrorism" and "terrorist content." The taxonomy includes the following labels that are applied to the content when a company ads hashes to the shared database.

- Imminent Credible Threat (ICT): A public posting of a specific, imminent, credible threat of violence toward non-combatants and/or civilian infrastructure.
- Graphic Violence Against Defenseless People: The murder, execution, rape, torture, or infliction of serious bodily harm on defenseless people (prisoner exploitation, obvious non-combatants being targeted).
- Glorification of Terrorist Acts (GTA): Content that glorifies, praises, condones, or celebrates attacks after the fact.
- Recruitment and Instruction (R&I): Materials that seek to recruit followers, give guidance, or instruct them operationally.
- New Zealand Perpetrator Content: The GIFCT set a new precedent in the wake of the New Zealand terrorist attack. Due to the virality and cross-platform spread of the attacker's manifesto and attack video, and because New Zealand authorities deemed all manifesto and attack video content illegal, the GIFCT created a crisis bank to mitigate the spread of this content.

The following shows the breakdown of how much content has been ingested into the shared database of hashes based on the above taxonomy.

- Imminent Credible Threat: 0.4 percent
- Graphic Violence Against Defenseless People: 4.8 percent
- Glorification of Terrorist Acts: 85.5 percent
- Radicalization, Recruitment, Instruction: 9.1 percent
- New Zealand Perpetrator Content: 0.6 percent

More information can be found here: *https://gifct.org/transparency/*

In addition to these efforts by the GIFCT, in 2018 Twitter began working with a small group of companies to test a new collaborative system. Because Twitter does not allow files other than photos or short videos to be uploaded, one of the behaviors we saw from those seeking to promote terrorism was to post links to other services where people could access files, longer videos, PDFs, and other materials. Our pilot system allows us to alert other companies when we removed an account or Tweet that linked to material that promoted terrorism hosted on their service. This information sharing ensures the hosting companies can monitor and track similar behavior, taking enforcement action pursuant with their individual policies. This is not a high-tech approach, but it is simple and effective and recognizes the resource constraints of smaller companies.

Based on positive feedback, the partnership has now expanded to 12 companies with which we have shared more than 12,000 unique URLs. Every time a piece of content is removed at source, it means any link to that source—wherever it is posted—will no longer be operational.

We are eager to partner with additional companies to expand this project, and we look forward to building on our existing partnerships in the future.

Response to Written Questions Submitted by Hon. Jacky Rosen to Nick Pickles

Question 1. The challenge for social media platforms prohibiting certain types of behavior on their sites is creating clear and concise rules for users to comply. Offensive conduct isn't a static issue, and as technology has evolved, so have our definitions of what constitutes abusive behavior such as cyberbullying and misinformation campaigns.

- Can you explain to us how your companies come up with rules regarding hateful speech and how those rules have evolved? What are your guidelines for determining when charged rhetoric crosses the line into becoming hate speech? For example, how do you determine if rhetoric is anti-Semitic?
- How closely do you work with outside groups, researchers, and users to come up with definitions of what constitutes hate and abusive speech and policies to deal with ambiguous cases? For instance, have you worked with the Anti-Defamation League or other groups combating hate when determining guidelines?

Answer. We draft and enforce the Twitter Rules to keep people safe on our service, and to protect the health of the public conversation. The Twitter Rules apply to everyone. In general, we create our rules with a rigorous policy development process; it involves in-depth research, analysis of the behavior of individuals on Twitter, historical violation patterns, and immersion in academic material.

We appreciate these issues are complex and we value the input of external voices in developing our approach. As part of our internal development process, we consult with a wide range of stakeholders and we focus on the risk of gaming, subverting, or otherwise abusing our policies and product changes. We supplement this work with conversations with outside experts and organizations where appropriate.

For example, many scholars have examined the relationship between dehumanization and violence. In September 2018, we tried something new by asking the public for feedback on a policy before it became part of the Twitter Rules. Our goal was to test a new format for policy development whereby the individuals who use Twitter have a role in directly shaping our efforts to protect them. We wanted to expand our hateful conduct policy to include content that dehumanizes others based on their membership in an identifiable group, even when the material does not include a direct target.

We asked for feedback to ensure we considered a wide range of perspectives and to hear directly from different communities and cultures who use Twitter around the globe. In two weeks, we received more than 8,000 responses from people located in more than 30 countries.

Following our review of public comments, in July 2019, we expanded our rules against hateful conduct to include language that dehumanizes others on the basis of religion.

We also work with outside groups, including those represented on the Twitter Trust and Safety Council, of which the Anti-Defamation League is a member. These groups are able to provide input on a range of policy and product approaches, both as part of the council and in direct conversations with teams at Twitter.

Question 2. With almost three and a half billion social media users worldwide—and one million users joining every day—social media platforms have turned to a mix of machine learning and human moderators to detect and take down hate speech, terrorist propaganda, cyber-bullying, and disinformation. Machine learning can be a useful tool in identifying objectionable content quickly, preventing it from spreading. However, there are concerns about its ability to understand the context of text or images, and the length of time it takes to train systems with new data to recognize objectionable content.

- Can you give us an estimate of how many content moderation decisions are made by your machine learning systems? And can you provide an estimated error rate for content flagged by machine learning?
- Are there instances where machine learning is more effective in flagging certain content than others? Does the error rate change significantly from one type of content to another?

Answer. Twitter's philosophy is to take a behavior-led approach, utilizing a combination of machine learning and human review to prioritize reports and improve the health of the public conversation. That is to say, we increasingly look at how accounts behave before we look at the content they are posting. This is how we can scale our efforts globally and leverage technology even where the language used is highly context specific. Twitter employs extensive content detection technology to identify potentially abusive content on the service, along with allowing users to report content to us either as an individual or as a bystander.

For abuse, this strategy has allowed us to take three times the amount of enforcement actions on abuse within 24 hours than this time last year. We now proactively surface over 50 percent of abusive content we remove using our technology compared to only 20 percent a year ago. This reduces the burden on individuals to report content to us. Since we started using machine learning three years ago to reduce the visibility on abusive content:

- 80 percent of all replies that are removed were already less visible;
- Abuse reports themselves have been reduced by 7.6 percent;
- The most visible replies receive 45 percent less abuse reports;
- 100,000 accounts were suspended for creating new accounts after a previous suspension during January through March 2019—a 45 percent increase from the same time last year;
- 60 percent faster response to appeals requests with our new in-app appeal process;
- 3 times more abusive accounts suspended within 24 hours after a report compared to the same time last year; and
- 2.5 times more private information removed with a new, easier reporting process.

Machine learning plays an important role across a multitude of our product surface areas. Making Twitter healthier also requires making the way we employ machine learning more fair, accountable, and transparent.

In many areas, machine learning is not sufficiently accurate to utilize in content removal decisions. For example, machine learning is not well suited to address sarcasm, innuendo, satire, or distinguish news coverage from propaganda broadcasts. A human role in these content decisions is essential to protect vulnerable groups, public debate, and free expression.

As machine learning evolves, there are some challenges that are more difficult than others. Often this is tied to the availability of training data for models. Rare events create fewer opportunities to obtain training data that hinder future efforts to identity similar incidents, while those that happen frequently offer greater data to train models. This is an area where further industry collaboration is essential, as the availability of training data is fundamental to the ability of companies to develop machine learning models that are better able to identify and remove different types of problematic content.

We welcome efforts to increase collaboration in this area, both with industry and governments. Increased efforts to foster technical collaboration will enable us to build upon work already done, and policymakers can support these efforts with greater legal protections for companies sharing content of this nature.

Response to Written Questions Submitted by Hon. Shelley Moore Capito to Derek Slater

Question 1. I represent a state that has one of—if not—the highest rates of drug overdoses deaths, the vast majority of which are due to opioids. I appreciate Google's recent announcement to launch new tools to help connect individuals recovering from opioid addiction with treatment resources. Does Google plan on doing similar initiatives to combat radicalization, like providing resources to mental health services?

Answer. We hope that Google's growing, multi-faceted efforts to address the opioids epidemic can help West Virginia and families nationwide grapple with substance use issues. These efforts include prevention, like making it easier for people to find medication disposal sites in their communities, to the treatment resources you mentioned, and providing uplifting resources for people in long term recovery for substance abuse.

Like substance use, mental health treatment generally is complex and stigmatized, with 50 percent of individuals not receiving needed treatment for depression and 1 in 5 not receiving needed treatment for PTSD. In partnership with the National Alliance on Mental Illness (NAMI), Google enabled people searching for information on mental health conditions including depression and PTSD to understand the likelihood of having these conditions by taking brief, clinically validated surveys (PHQ–9 for depression and PC–PTSC–5 for PTSD). We also help users to find resources to take action toward recovery including directing them to instant access to the National Suicide Hotline by phone or chat if they are having suicidal thoughts.

Question 2. This Committee has held a number of hearings on the rise and importance of artificial intelligence (AI) in today's digital economy. AI has been invaluable in collecting and sorting massive amounts of data. In the case of today's hearing, AI has become critical in order to identify radicalization and terrorist threats. Each company has identified key tools each company uses in identifying bad actors on your platforms, but machine learning being one of the most critical. What factors are given priority when determining radicalized or terrorist content?

a. You also mention the importance of human expertise in determining more nuanced cases. When does human expertise step in after AI has identified or flags content?

b. After content has been flagged for law enforcement involvement, what is the process that takes place afterward? Does that content get sent to the FBI and then disseminated to state law enforcement?

Answer. We use a mix of people and technology to address terrorist and violent extremist content on our platforms. We apply our most advanced machine learning research to train new "content classifiers" to help us more quickly identify and remove extremist and terrorism-related content. This can be challenging: a video of a terrorist attack may be informative reporting by a news agency, or glorification of violence if uploaded in a different context by a different user. Human reviewers play a key role in making nuanced decisions about the line between violent propaganda and newsworthy speech. Our efforts to address this content have also included consultation with dozens of experts in subjects like terrorism, violent extremism, civil rights, and free speech.

The Stored Communications Act allows Google and other service providers to voluntarily disclose user data to governmental entities in emergency circumstances where the provider has a good faith belief that disclosing the information will prevent loss of life or serious physical injury to a person. When we have a good faith belief that there is a threat to life or serious bodily harm made on our platform in the United States, the Google CyberCrime Investigation Group (CCIG) will report it to the Northern California Regional Intelligence Center (NCRIC). In turn, NCRIC quickly gets the report into the hands of officers to respond. Our team is staffed on a 24/7/365 basis to respond to these emergency disclosure requests (EDRs).

In other cases, law enforcement agencies at the Federal and state levels make emergency requests to Google for user data in situations involving danger of death or serious physical injury to any person. As illustrated in our transparency report covering government requests for user data, the number of EDRs submitted from agencies in the U.S. almost doubled from 2017 to 2018. We have grown our teams to accommodate this growing volume and to ensure we can quickly respond to emergency situations that implicate public safety.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. RICHARD BLUMENTHAL TO DEREK SLATER

NO HATE Act and Reporting

I have introduced legislation, the Jabara-Heyer NO HATE Act, which would help states implement and train officers in the National Incident-Based Reporting Systems. The NO HATE Act would also provide grants to states to better address hate crimes by training law enforcement, establish specialized units, create community relations programs, and run hate crime hotlines.

Question 1. Do you support the Jabara-Heyer NO HATE Act? I understand that it has taken some time for Google and Facebook to establish reliable and timely channels to report threats made on your platform to the proper authorities. Mr. Slater, you testified that you now have a strong relationship with the Northern California Regional Intelligence Center, who has been effective at quickly getting reports of threats into the right hands.

Question 2. Would you support adding measures to the Jabara-Heyer NO HATE Act to expand the NCRIC model of integrated threat-reporting nationwide?

Question 3. What steps would improve communications channels with law enforcement to make sure the right information gets into the right hands quickly?

Answer. Addressing questions 1–3: We appreciate your work in this area and share your interest in getting threat information to law enforcement so they can take immediate action. We have worked with law enforcement to create efficient processes with NCRIC and think we have all made important strides together. While we have not yet taken a position the Jabara-Heyer NO HATE Act, we support the intent and we are interested in learning more about how it could be amended to include an expansion of the NCRIC model. Google is supportive of efforts to expand the approach spearheaded by NCRIC to allow for greater geographic coverage, handle overflow work, and to make the process more robust. It is also important that NCRIC continue to receive the necessary funding to continue building its capacity and effectiveness.

Amplification of 8Chan and Other Hate Sites

We have seen over this year that fringe sites are a breeding ground for racist and violent hate communities. However, extremists then use mainstream platforms to recruit and amplify their hate and ideologies to a larger audience. In particular, the site 8chan has had a repeated role in multiple mass shootings this year. The perpetrators of Christchurch mosque shootings, Poway synagogue shooting, and El Paso massacre each posted manifestos to 8chan before their attacks. It is also sites such as 8chan that facilitate campaigns of harassment and terrorism that target the victims of mass shootings, such as the Sandy Hook families. 8chan is currently offline after webhosting providers finally cut their ties after the El Paso shootings. However, 8chan's owner has said that he plans to revive the site as soon as this week.

Question 4. Has your company taken any steps to limit the spread of 8chan content, including the communities that hosted the manifestos of shooters, on your platforms?

Question 5. Please describe the specific steps you to restrict the amplification of 8chan and other violent sites on your platforms, including what sites you have taken action to restrict.

Answer. Answering questions 4–5: We take a number of steps to address harmful content across our platforms, regardless of the source, including the following:

> *Removing content from hosted platforms:* Hate speech is not allowed on YouTube and other Google hosted platforms, and we are bringing significant attention to detection and removal of hateful content on our platforms. We have a number of policies that work together to disallow hateful content—our hate speech policy, our harassment policy which disallows malicious attacks against individuals, and our general policy that disallows incitement to violence. Our policies would be applicable to "manifestos" from those who commit violent acts.
>
> *Removing financial incentives:* In addition, our longstanding policies prevent ads from running on violative content, including hate speech. On YouTube, channels that have shown a history of brushing up against our hate speech policies (even if they haven't crossed the line), will be suspended from our YouTube Partner program.
>
> *Reducing recommendations of borderline content:* Besides removal of content on YouTube, we also take other steps to curb potentially harmful content. Several months ago, we began reducing visibility of borderline content (which comes close to but doesn't quite violate our rules) or content that can misinform users in harmful ways. This will be a gradual change, but this approach is already starting to bear fruit. In the U.S. we've seen a 50 percent drop of views from recommendations to this type of content, meaning quality content has more of a chance to shine.

COPPA Settlement and Children's Privacy

Question 6. How will Google implement its new promises under its COPPA-related consent decree regarding data collection on content in which it has a direct financial and curatorial relationship, including Google Preferred?

Answer. We are making a number of changes to how we treat data on children's content on YouTube to address the concerns reflected in the FTC's investigation.

In order to identify this category of content, we will be requiring creators to tell us when their content is made for kids. We will also use machine learning to find videos that clearly target young audiences, for example those that have an emphasis on kids characters, themes, toys, or games and use it as a signal that will help us define the child directed content at YouTube.

We will treat data from anyone watching made for kids content on YouTube as coming from a child, regardless of the age of the user. This means that we will limit data collection and use on videos made for kids only to what is needed to support the operation of the service. We will stop serving personalized ads on this content entirely, and some features will no longer be available on this type of content, like comments and notifications.

Advertising inventory on content made for kids will be available for certain reservation ad buys, such as custom packs and sponsorships, but we currently don't have plans to include it in reservation buys like the Google Preferred Lineup and Breakout Video packages.

Question 7[1]. When Google says it will only collect data on videos made for kids for "what is needed to support the operation of the service," what specifically will it gather and how will it be used?

Answer. On content that is identified as made for kids, we are putting in place limitations on the data we collect and use as described above. This means:

- Limiting the collection of personal information like name, address, or contact information. We will collect user activity information such as when users watch a video or click on an advertisement, and information about their device such as IP address.
- Disabling features including comments, sharing features, notification requests, and add-to-playlist features. Actions such as Subscribe or Like may still be enabled for users logged-in to their Google accounts, but would have limited functionality.
- Prohibiting the serving of personalized advertising or remarketing ads.

Using data only to support the operation of the service, which includes: performing actions you request (*e.g.,* playing a video); respecting your settings (*e.g.,* preferred language/country which allows us to surface *e.g.,* videos in French for users in France); preventing fraud and abuse; personalizing users' experience on YouTube (*e.g.,* recommending relevant content based on watch history), depending on their account settings; and serving contextual advertisements.

Question 7[2]. What role, if any, will the Google Marketing Platform play in this regard?

Answer. The changes that we're implementing to how we collect and use data on content identified as made for kids, as described above, will apply across our ad products that may be used to serve advertising on YouTube.

Question 8. Will Google expand its new kid privacy safeguards to its other child directed services, such as the Play Store?

Answer. We have already made a number of improvements this year as part of the revamp of the Designed for Families Program, which *launched* in May 2019. These policy changes build on our existing efforts to help ensure that apps for children have appropriate content, show suitable ads, and handle personally identifiable information correctly; they also reduce the chance that apps not intended for children could unintentionally attract them. Developers who have children as part of their target audience must meet stringent policy requirements in their apps concerning both content safety, ads appropriateness, and privacy protections. We will also be double checking apps to make sure that they are not seeking to attract children but attempting to avoid these requirements. We take action when we identify developers who do not fulfill these policy requirements and remove their content from the Play store when appropriate.

Question 9. Will the new fund for children's content creators principally fund non-commercial and ad free content aimed for kids, families and for education?

Answer. We are currently determining the criteria for distributing the funds, working with children's media experts to ensure we are funding high-quality content with a global reach. We will look to fund enriching content similar to the type of content we featured in our *Creating for Families* Field Guide. When complete, the content will appear on our existing platforms, which are supported through contextual advertising so that they can remain free and accessible for all families, regardless of ability to pay.

Question 10. How will Google deal with influencer and unboxing videos aimed at children?

Answer. We do not currently allow paid promotional content on YouTube Kids. See *here* and *here* for more details because there is not yet an industry consensus on what an appropriate disclosure for such an audience would require. On the main YouTube service, however, which is intended for a wider audience, paid promotional videos are generally permitted. Such promotions must be disclosed as paid promotions and must abide by all YouTube ads policies (including restrictions around targeting kids under 13). These policies are outlined in our Help Center *here.* We are also seeking guidance among industry partners and regulators to determine what an appropriate disclosure for kid appealing content might look like in this context.

There is a wide range of content that is sometimes called 'unboxing'—ranging from videos of kids playing with toys (or indeed even cardboard boxes, making them into pirate ships or castles) to an adult showing off a Chewbacca mask. Therefore not all unboxing is necessarily commercial, not all of it is compensated, and the audiences are diverse. When creators of unboxing videos are compensated in any way by an advertiser, Google would consider the video to be a paid promotion and the creator would be required to mark it as such under our policies. Unboxing videos that are not motivated by any consideration or connection with an advertiser are not commercial, and are outside of the scope of such policies. We are currently working with experts to understand how to treat this content and whether we need to update our policies or practices.

Question 11. Are the YouTube changes all global?

Answer. Yes, the changes we are making consistent with the FTC settlement will be implemented globally.

Testing of Consumer Platforms

Question 12. Please describe the process you use to test and evaluate new consumer facing products, including algorithms designed to promote forms of engagement. What methods are employed to assess the impact of these products on individuals and groups, both for an immediate and medium term response?

Answer. Google's evaluation and testing processes reflect the diversity of our consumer-facing products and offerings, ranging from laptops to operating systems or search engines–typically involving multiple rounds of testing, experiments, and reviews with product, engineering, trust-and-safety, legal, policy, and privacy experts.

Those reviews are designed to verify that the product functions as expected, to explore unintended consequences, and to address possible risks.

For instance, to help 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. Search Quality Raters follow strict guidelines that define our goals for Search algorithms and are publicly available for anyone to see. The ratings provided by Search Quality Raters help us benchmark the quality of our results so that we can meet a high bar for users of Google Search all around the world.

In addition to the Search quality tests, we conduct live traffic experiments to see how real people interact with a feature, before launching it to everyone. Results from these experiments undergo a review by experienced engineers and search analysts, as well as other legal and privacy experts, who then determine whether the change is approved to launch. In 2018, we ran over 654,680 experiments, with trained external Search Raters and live tests, resulting in more than 3,234 improvements to Search. For more information on this process and our methods, please refer to *www.google.com/search/howsearchworks.*

YouTube also conducts robust evaluation and testing processes ahead of launching new features or policies. Given our scale, it's important that we roll out new offerings and product changes incrementally so we can monitor performance and feedback from users. Creators can also submit feedback directly through YouTube Studio.

YouTube's development of platform policies provides another example. At YouTube, we have developed robust "Community Guidelines" that set the rules of the road for what we don't allow. We are constantly evaluating these policies and their enforcement, incorporating feedback from experts and trends we see on the platform; we made 30 updates to our policies in the last year alone. For instance, we strengthened our hate speech policy in June by specifically prohibiting videos alleging that a group is superior in order to justify discrimination, segregation or exclusion based on qualities like age, gender, race, caste, religion, sexual orientation or veteran status. When evaluating our approach towards hateful content we consulted with dozens of experts in subjects like violent extremism, civil rights, and free speech.

In addition to our teams focused on policy development and enforcement, we've also established an intel desk to help us detect emerging trends in how people and organizations may try to misuse our platform. It identifies new trends in harmful content by synthesizing leads from third party intel vendors, internal trend data, social listening, and other relevant inputs.

Question 13. Do you ever identify unintended consequences of such proposed products and then revise them or decide not to launch?

Answer. Yes. Where our reviews identify significant unintended consequences for users or society that we cannot adequately resolve in time for our planned launch date, we may postpone temporarily or indefinitely. This is a normal part of doing business for each of the products and services that we operate.

Question 14. What testing and measurement methodologies are routinely used and how are the product evaluation teams selected? Please submit any criteria you have developed for new or revised data driven products or applications, including their intended impact, demographic reach, and revenue potential.

Answer. Our evaluation criteria for new products or services are a direct function of their intended goals. For instance, our hardware product launch process will involve in-depth reviews for the quality, durability, and resilience of each individual component, in line with the best industry standards, whereas the same processes would not make sense for software, storage, or computing services. When it comes to our ranking algorithms, as mentioned above, we use a number of methods including side-by-side experiments with trained raters and live experiments to test whether a given change to our products represents a tangible improvement for our users. Our criteria are both quantitative and qualitative, aimed at measuring not just the changes in user behavior, but also whether a given change advances the goals outlined in our Search Quality Rater Guidelines.

Response to Written Questions Submitted by Hon. Tom Udall to Derek Slater

Question 1. We cannot talk about mass violence without talking about the social and political climate that is dividing America. Most recently, content that demonizes and spreads hate against immigrant communities is proliferating across social media. This content is too often indistinguishable from social media posts from some elected representatives. How does your company define hate speech? And how do you address situations when content that meets that definition comes from a political leader?

Answer. Hate speech is not allowed on YouTube and other Google hosted platforms, and we are bringing significant attention to detection and removal of hateful content on our platforms. We enforce those policies regardless of a speaker's political persuasion.

We have a number of policies that work together to disallow hateful content—our hate speech policy, our harassment policy which disallows malicious attacks against individuals, and our general policy that disallows incitement to violence. For instance, our hate speech policy on YouTube specifically prohibits: "Content that encourages or glorifies violence against individuals or groups, or whose primary purpose is to incite hate against individual or group based on attributes including age, ethnicity, disability, gender, nationality, race, immigration status, religion, sex, sexual orientation, and veteran status." It's important to note that YouTube takes action against hateful content, not based on speakers.

We don't allow content that dehumanizes individuals or groups with these attributes, claims they are physically or mentally inferior, or praises or glorifies violence against them. We also don't allow use of stereotypes that incite or promote hatred based on these attributes, or racial, ethnic, religious, or other slurs where the primary purpose is to promote hatred. Our policy prohibits content that alleges the superiority of a group over those with any of the attributes noted above to justify violence, discrimination, segregation, or exclusion. We also do not allow content that denies that a well-documented, violent event took place.

In enforcing our hate speech policy, we consider the purpose of the video. If users are posting educational, documentary, scientific, or artistic content related to hate speech, we encourage them to be mindful to provide enough information so viewers understand the context, such as through an introduction, voiceover commentary, or text overlays, as well as through a clear title and description. We give users tips and tools for adding context on YouTube.

Question 2. Knowing that the problem of extremism and mass violence extends beyond the screen, I would like you to describe your partnerships with communities and organizations around the country to fight against extremism and hate. What are you doing to promote their voices on your platforms? And what makes them effective?

Answer. In 2016, we launched YouTube Creators for Change, an initiative dedicated to amplifying the voices of role models who are tackling difficult social issues with their channels. From combating hate speech, to countering xenophobia and extremism, to simply making the case for greater tolerance and empathy toward others, these creators are helping to foster productive conversations around tough issues and make a positive impact on the world.

As part of their commitment to the program, Creators for Change Ambassadors and Fellows receive mentorship and promotional support to aid the creation of their Impact Projects—films that tackle a wide range of topics, from self-acceptance and showing kindness to others, to celebrating cultures and advocating for global empathy.

Creators for Change is a global program that thrives through its many local chapters. From providing education on the dangers of fake news, to helping create safe spaces for making content that addresses hate speech, these chapters empower thousands of young people to drive positive social change across Europe, the Middle East and the Asia-Pacific region.

We have produced annual reports detailing information about the Creator Ambassadors and the billions of views their content has generated, see *https://www.youtube.com/creators-for-change/*.

In addition, because technology alone is not a silver bullet, we have greatly increased the number of independent experts in YouTube's Trusted Flagger program. Machines can help identify problematic videos, but human experts still play a role in nuanced decisions about the line between violent propaganda and religious or newsworthy speech. While many user flags can be inaccurate, Trusted Flagger reports are accurate over 90 percent of the time and help us scale our efforts and identify emerging areas of concern. We will expand this program by adding 50 expert NGOs to the 63 organizations who are already part of the program, and we will support them with operational grants. This allows us to benefit from the expertise of specialized organizations working on issues like hate speech, self-harm, and terrorism. We will also expand our work with counter-extremist groups to help identify content that may be being used to radicalize and recruit extremists.

Finally, we would also note that Jigsaw, a project of Google's parent company Alphabet, created the Redirect Method—a way to use AdWords targeting tools and curated YouTube videos uploaded by people all around the world to confront online radicalization. For example, it focuses on the slice of ISIS' audience that is most susceptible to its messaging, and redirects them towards curated YouTube videos debunking ISIS recruiting themes. This open methodology was developed from interviews with ISIS defectors, respects users' privacy and can be deployed to tackle other types of violent recruiting discourses online.

Question 3. We are entering another election year and we know that foreign actors have amplified divisive rhetoric on social media and, in some cases, orchestrated actual protests. What specific actions are you taking to prepare for 2020 to prevent Russia and other foreign actors from trying to inflame racial and political tensions through social media?

Answer. Although we found limited activity on our platforms in 2016 and during the 2018 midterms, we understand the existence of this threat, and take the integrity of our elections very seriously. We have a team dedicated to ensuring the integrity of election-related content and ads across our platforms, including combating potential foreign influence.

We've taken various key steps to combat election interference. For example, we have policies that prohibit misrepresentation and other forms of abuse, and we have devoted significant resources to enforcing our policies, and we have conducted vulnerability testing across key products and made several changes to safeguard our products from being used to confuse voters, such as through manipulation of search features (*e.g.,* WebAnswers, Knowledge Panels). We also worked closely with others in industry and government election integrity task forces to be able to identify threats and respond quickly.

We are approaching the 2020 election with vigilance and commitment. We expect to once again establish a war room with dedicated full-time employees to provide 24/7 monitoring and rapid escalation of any issues in the days before and after the elections. We continue to provide regular updates on our work to that end.

This is in addition to our broader efforts to ensure the integrity of our elections. For example, we've trained 1,000 campaign professionals last year about online security, and we've released "*Protect Your Election,"* a suite of digital tools designed to help election websites and political campaigns protect themselves from digital attacks.

Question 4. Regarding the shared industry database of hashes linked to content that promotes terrorism; I would like to understand the thresholds for including certain content in the database. Who makes the decision to include content in that database and how is that decision made? What percent of that database concerns white nationalist or other domestic extremist content?

Answer. In working together to build technological solutions that will prevent and disrupt the spread of terrorist content online, the largest cross-platform advancement supported by the Global Internet Forum to Counter Terrorism (GIFCT) has been the creation of a Hash Sharing Consortium. The consortium shares "hashes" (or digital fingerprints) of known terrorist images and videos. The image or video is "hashed" in its raw form and is not linked to any source original platform or user data. Hashes appear as a numerical representation of the original content and can't be reverse engineered to create the image and/or video. A platform needs to find a match with a given hash on their platform in order to see what the hash corresponds with.

It is up to each consortium member how they utilize the database and how they contribute to it, depending on their own terms of service, how their platform operates, and how they utilize technical and human capacities.

Companies often have slightly different definitions for "terrorism" and "terrorist content". For the purposes of the hash sharing database, and to find an agreed upon common ground, founding companies in 2017 decided to define terrorist content based on content relating to organizations on the UN Terrorist Sanctions lists. Companies also agreed upon a basic taxonomy around the type of content ingested relating to these listed organizations. The taxonomy includes labels that are applied to the content when a company adds hashes to the shared database.

GIFCT released its first transparency report in 2019; it includes the specific taxonomy used by the Hash Sharing Consortium and the respective percentages of each category, available at: *gifct.org/transparency.*

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. JACKY ROSEN TO DEREK SLATER

Question 1. The challenge for social media platforms prohibiting certain types of behavior on their sites is creating clear and concise rules for users to comply. Offensive conduct isn't a static issue, and as technology has evolved, so have our definitions of what constitutes abusive behavior such as cyberbullying and misinformation campaigns.

- Can you explain to us how your companies come up with rules regarding hateful speech and how those rules have evolved? What are your guidelines for determining when charged rhetoric crosses the line into becoming hate speech? For example, how do you determine if rhetoric is anti-Semitic?

Answer. We are investing in the policies, resources and products needed to live up to our responsibility and protect the YouTube community from harmful content. Over the past few years, this work has focused on four pillars: removing *violative content,* raising up *authoritative content,* reducing the spread of *borderline content* and *rewarding trusted creators.* Thanks to these investments, videos that violate our policies are removed faster than ever and users are seeing less borderline content and harmful misinformation. As we do this, we're *partnering closely* with lawmakers and civil society around the globe to limit the spread of violent extremist content online.

We review our policies on an ongoing basis to make sure we are drawing the line in the right place: in 2018 alone, we made more than 30 policy updates. One of the most complex and constantly evolving areas we deal with is hate speech.

YouTube has always had rules of the road, including a longstanding policy against hate speech, but we've been taking a close look at our approach towards hateful content in consultation with dozens of experts in subjects like violent extremism, supremacism, civil rights, and free speech. Based on those learnings, we made several updates.

In June 2019, we updated YouTube's hate speech policy by specifically prohibiting videos alleging that a group is superior in order to justify discrimination, segregation or exclusion based on qualities like age, gender, race, caste, religion, sexual orientation or veteran status. This would include, for example, videos that promote or glorify Nazi ideology, which is inherently discriminatory. Additionally, we will remove content denying that well-documented violent events, like the Holocaust or the shooting at Sandy Hook Elementary, took place.

Human reviewers remain essential to both removing content and training machine learning systems because human judgment is critical to making contextualized decisions on content. The total number of people across Google working to address content that might violate our policies is over 10,000. Our trust and safety teams manually review millions of videos, helping train our machine-learning technology to identify similar videos in the future.

- How closely do you work with outside groups, researchers, and users to come up with definitions of what constitutes hate and abusive speech and policies to deal with ambiguous cases? For instance, have you worked with the Anti-Defamation League or other groups combating hate when determining guidelines?

Answer. Regarding the definition of hate speech, we operate in 190 countries, and hate speech laws vary by country. We respect the law as required in each country, and will block illegal hate speech content in a given country to comply with its applicable local laws. In addition, our hate speech policy is part of the YouTube Community Guidelines, which we enforce globally. That policy prohibits content that promotes violence against individuals or groups based on certain attributes, such as race, religion, disability, gender, age, or veteran status. We enforce those policies regardless of a speaker's political persuasion.

The YouTube Trusted Flagger program is an important part of our work with external experts. The program was developed by YouTube to help provide robust tools for individuals, government agencies, and non-governmental organizations (NGOs) that are particularly effective at notifying YouTube of content that violates our *Community Guidelines.* Trusted flaggers have expertise in at least one policy vertical, flag content frequently with a high rate of accuracy, and are open to ongoing discussion and feedback with YouTube about various content areas.

We have made improvements to YouTube's flagging tools, based on feedback from our Trusted Flagger network. In addition to our bespoke tools for Trusted Flaggers, we designed a dashboard that allows any user to check the status of flags they have submitted. The dashboard tells users if the content they flagged is active, removed, or restricted.

Consultation with external experts is a key aspect of how we develop our approach to tough issues and how we evaluate our guidelines and enforcement mechanisms.

Question 2. With almost three and a half billion social media users worldwide—and one million users joining every day—social media platforms have turned to a mix of machine learning and human moderators to detect and take down hate speech, terrorist propaganda, cyber-bullying, and disinformation. Machine learning can be a useful tool in identifying objectionable content quickly, preventing it from spreading. However, there are concerns about its ability to understand the context of text or images, and the length of time it takes to train systems with new data to recognize objectionable content.

- Can you give us an estimate of how many content moderation decisions are made by your machine learning systems? And can you provide an estimated error rate for content flagged by machine learning?
- Are there instances where machine learning is more effective in flagging certain content than others? Does the error rate change significantly from one type of content to another?

Answer. As you might imagine, it takes a combination of both machine learning and human review to effectively review content and we actively monitor the success of both efforts. With human review, we check to see what decisions are being made by reviewers and update our guidelines if they are not clear or not meeting expectations. And we are constantly working to improve machine learning.

The profound impact of YouTube's updated hate speech policy update is already evident in the data released in YouTube's Q2-2019 transparency report: the number of individual video removals for hate speech saw a 5x spike to over 100,000, the number of channel terminations for hate speech also saw a 5x spike to 17,000, and the total comment removals nearly doubled in Q2 to over 500 million due in part to a large increase in hate speech removals. And because of our ability to remove this content quickly, videos that violate our policies generate a fraction of a percent of the views on YouTube. For example, the nearly 30,000 videos we removed for hate speech over the last month generated just 3 percent of the views that knitting videos did over the same time period.

Response to Written Questions Submitted by Hon. Richard Blumenthal to George Selim

NO HATE Act and Reporting

I have introduced legislation, the Jabara-Heyer NO HATE Act, which would help states implement and train officers in the National Incident-Based Reporting Systems. The NO HATE Act would also provide grants to states to better address hate crimes by training law enforcement, establish specialized units, create community relations programs, and run hate crime hotlines.

Question 1. Do you support the Jabara-Heyer NO HATE Act?

Answer. ADL strongly supports the Jabara-Heyer NO HATE Act.

Since 1990, the FBI has been collecting and reporting hate crime data, required by the Hate Crime Statistics Act of 1990 (HCSA). While the FBI HCSA data provides the best national snapshot of bias-motivated criminal activity in America, it is clearly incomplete. For example, in 2017, the most recent data available, only 2,040 of the 16,149 reporting law enforcement agencies—less than 13 percent—reported one or more hate crimes to the FBI. The remaining 87 percent of participating agencies affirmatively reported zero hate crimes to the FBI, including 92 cities with populations over 100,000. And more than 1,000 law enforcement agencies did not report any data to the FBI, including 9 cities over 100,000. The entire state of Mississippi reported one hate crime in 2017, Alabama reported 9, and Arkansas reported 7. By contrast, two cities that have focused on effective hate crime response, Boston and Seattle, reported 140 hate crimes and 234, respectively.

Studies have shown that more comprehensive, complete hate crime reporting can deter hate violence.[1] Better data will assist in proper allocation of police resources and personnel—preventing crimes and reassuring victims. And better data will advance police-community relations. Improved data collection will necessarily require outreach and expanded networking and communication with targeted communities, as well as more training for law enforcement personnel in how to identify, report, and respond to hate violence.

I understand that it has taken some time for Google and Facebook to establish reliable and timely channels to report threats made on your platform to the proper authorities. Mr. Slater testified that Google now has a strong relationship with the

[1] See "Investigation of Hate Crimes; Model Policy; Concepts & Issues Paper; Need to Know. . . ." from the IACP Law Enforcement Policy Center, September 2016, *https://www.theiacp.org/sites/default/files/2018–08/HateCrimesBinder2016v2.pdf.*

Northern California Regional Intelligence Center, who has been effective at quickly getting reports of threats into the right hands.

Question 2. Would you support adding measures to the Jabara-Heyer NO HATE Act to expand the NCRIC model of integrated threat reporting nationwide?

Answer. While law enforcement collaboration should be part of any conversation on improving responsiveness to hate crimes, I am not prepared to comment specifically on the NCRIC model. Overall, it is important that the Jabara-Heyer NO HATE Act should remain focused on one thing—improving reported hate crime data. Threats may or may not be criminal activity.

Question 3. What steps would improve communications channels with law enforcement to make sure the right information gets into the right hands quickly?

Answer. Improving hate crime data requires at least two efforts—law enforcement authorities ready and willing to collect the data, and members of the targeted communities ready and willing to contact the police to report that they have been the victims of bias-motivated violence. The Department of Justice should incentivize state and local law enforcement to more comprehensively collect and report hate crimes data to the FBI, with special attention devoted to large underreporting law enforcement agencies that either have not participated in the FBI HCSA program at all or have affirmatively and not credibly reported zero hate crimes.

If marginalized or targeted community members—including immigrants, people with disabilities, LGBTQ community members, Muslims, Arabs, Middle Easterners, South Asians and people with limited language proficiency—cannot report, or do not feel safe reporting hate crimes, law enforcement cannot effectively address these crimes, thereby jeopardizing the safety of all. Such efforts could be supported through the promotion of model policies and best practices and the passage of legislation designed to improve hate crime data collection and reporting legislation, such as the Jabara-Heyer NO HATE Act.

Incentives can encourage police departments to report their hate crime data, and help overcome negative publicity that can accompany hate crime reporting. Police departments need to have the support of the community when their hate crime numbers increase; an increase may well indicate improved police-community relations, increased trust in police, public confidence that they will respond seriously to hate crime reports.

Lastly, law enforcement must be encouraged to create relationships with community members who may be privy to threat information.

This past December, in Monroe, Washington, a clearly troubled young man made a series of anti-Semitic rants and violent posts online. He bragged about planning to “shoot up an (expletive) school” in a video while armed with an AR–15-style weapon, and on Facebook posted that he was “shooting for 30 Jews.” Fortunately, these posts came to the attention of the Anti-Defamation League, which was able to tip off the FBI. The ADL’s vigilance prevented another Parkland or Tree of Life attack.

Question 4. I take it that the ADL does not report every terrible or obscene comment on the Internet to the FBI. Can you tell me about the process and criteria that your organization uses to identify threats, such as in Monroe?

Answer. Investigative researchers at ADL encounter hundreds, if not thousands, of posts daily by individuals who make extreme and threatening comments on various online platforms. When we find such threats, we delve deeper into that person’s online footprint. If that person displays photos showing his/her weapons and expresses a desire to use those weapons against a community and we can identify either the person or where he/she lives, we will report that person’s comments to law enforcement. In addition, if we see that this individual is citing literature that promotes violence or previous violent acts as inspiration, we are more likely to report the person. These individuals often post their comments online hoping to receive support and encouragement to carry out acts of violence. Individuals we report to law enforcement are not just talking about hating a group of people—they want to take action. For example, we reported Dakota Reed in Washington State for making threats to carry out a mass killing of Jews, and also reported the comments of Corbin Kauffman of Leighton, Pennsylvania, in March 2019. In Kauffman’s case, in addition to posting violent comments and pictures of weapons, he posted a photo of himself carrying out an act of anti-Semitic vandalism. Law enforcement was able to identify him because of the various clues we provided about his identity.

Response to Written Questions Submitted by Hon. Tom Udall to George Selim

Question 1. Mr. Selim, in response to the recent spate of mass shootings, this administration has floated the dangerous idea of monitoring persons with mental illness to try to predict gun violence. This policy, however, would only serve to further stigmatize people with mental illness—who are more likely to be victims of crimes than perpetrators. It is also a disingenuous way of talking about gun violence without having to talk about guns.

What do you think about this proposal? In addition, can you describe what indicators of mass violence or extremism actually look like, based on the best research?

Answer. It is wrong to assert that people with mental health disabilities, including those with perceived mental health disabilities, are inherently dangerous and the cause of our Nation's gun violence problem or that targeting them will solve our country's gun violence problem is wrong. In fact, many studies have demonstrated that people with disabilities, including mental health disabilities, are far more likely to be victims of gun violence than perpetrators. Blaming persons with mental health disabilities is counterproductive, a distraction from the real problem, and can result in stigmatizing people with mental health disabilities and the disability community as a whole.

There is no one path to extremism, and there are no specific indicators that can act as predictors of extremist actions. Some academic sources have explored whether a combination of factors may indicate an over-arching risk of radicalization, and I direct you to those sources, such as this NIJ overview: *https://www.ncjrs.gov/pdffiles1/nij/251789.pdf*

Question 2. You were the director of the Office for Community Partnerships, and led the Countering Violent Extremism Task Force under the Department of Homeland Security. These offices were responsible for providing grants to anti-extremist groups and combatting domestic terrorism through interagency partnerships. Unfortunately, it appears these offices have been gutted by the current administration. We cannot fight against white supremacy and violent domestic extremism without partnering with communities, civil society, and federal, state, and local governments.

In your opinion, should funding be reinstated to support these initiatives, and why?

Answer. Yes, funding should be reinstated and scaled much higher. In light of how domestic terrorism laws differ from those of international terrorism, there are fewer law enforcement resources at the government's disposal, and prevention therefore is a key undertaking for the government. However, in light of the current administration's inadequacies and singular focus on Islamist-motivated forms of extremism, entities outside government must take the lead in preventing extremist violence. A public-private effort—with Congress funding research universities, technology companies, non-profit expert organizations, and state and local government partners—could provide the critical boost that prevention efforts need while also avoiding misgivings many have about the implications of an overly-federalized effort.

Response to Written Questions Submitted by Hon. Jacky Rosen to George Selim

As we discuss the spread of hate online, I want to turn our focus to combating anti-Semitism in the digital sphere. Last year we saw the deadliest attack on the Jewish community in American history, when eleven people were killed at the Tree of Life Synagogue in Pittsburgh. Perhaps unsurprisingly, the shooter was linked to numerous anti-Semitic postings on a fringe social networking site called Gab. And hours after a gunman opened fire at a synagogue in Poway, California, a violently anti-Semitic letter from the shooter appear on 8chan and Facebook, with links to the letter later showing up on Twitter and other social media sites, spreading his hateful ideas across the world.

Mr. Selim, online forums such as 8chan and Gab do very little to police their site from hateful and violent speech.

Question 1. What role do these sites play in perpetuating mass violence and domestic terrorism in our country, AND

Answer. Fringe web communities play a critical role in the dissemination of hate and extremist content. Perhaps the most important contributor to the subculture of the alt right is the so-called "imageboard," a type of online discussion forum originally created to share images. One of the most important is 4chan, a 15-year-old imageboard whose influence extends far beyond the alt right, as a key source of Internet memes. Its/pol subforum is a dark place, an anarchic collection of posts that range from relatively innocuous to highly offensive.

Over time, 4chan has become home to many racists and open white supremacists. Some of its imitators, such as 8chan, lean even more towards racism and white supremacy. Parts of Reddit, a popular website that contains a massive collection of subject-oriented discussion threads, also share the chan subculture, as do parts of Tumblr.

In April 2019, ADL released a report, a collaboration between Network Contagion Research Institute and ADL's Center on Extremism (COE), analyzing the similar ideological motivations and online activity of the perpetrators of the Pittsburgh and Christchurch massacres. Both killers announced their violent plans to their preferred Internet forums, Gab and 8chan, and were consumed by the white supremacist conspiracy theory of "white genocide," which is frequently referenced on both sites.

Both Gab and 8chan are rife with white supremacist, hateful, anti-Semitic bigotry. Image boards such as 4chan are totally anonymous, without user names, allowing participants to say or post whatever they want, no matter how offensive, without fear of being exposed. Many take full advantage to engage in some of the most crude and blatant offensive language online, taking aim at many targets. The chan subculture has a strong tendency to portray all such content as a joke, even when not intended to be, resulting in a strong "jkbnr" ("just kidding but not really") atmosphere. The alt right has also absorbed an even darker aspect of chan subculture: online harassment campaigns against people who have angered them.

Question 2. In the immediate aftermath of deadly attacks motivated by hate, how should mainstream social networks such as Facebook and Twitter interact with these fringe sites to stop the spread of manifestos, letters, and other hateful writings?

Answer. While the most extreme forms of online content thrive on websites like 8chan, Gab, and 4chan, larger social media platforms like Facebook, Twitter, and YouTube need to remain vigilant. Extremists leverage larger mainstream platforms to ensure that the hateful philosophies that begin to germinate on message boards like Gab and 8chan find a new and much larger audience. Twitter's 300 million users and Facebook's 2.4 billion dwarf the hundreds of thousands of users on 8chan and Gab. Extremists make use of mainstream platforms in specific and strategic ways to exponentially increase their audience while avoiding content moderation activity that Facebook and Twitter use to remove hateful content. These include creating private pages and events, sharing links that directly lead users to extreme content on websites like 8chan, as well as using coded language called dog whistles to imply and spread hateful ideology.

To address this, mainstream platforms should limit the ways they are spreading hateful messages from smaller platforms. Fringe platforms like Gab and 8chan openly cater to users interested in spreading hate and conspiracies, and a considerable amount of their content would violate mainstream platforms' terms of service. As a result, mainstream platforms must aim to decrease cross-users' ability to recruit and spread hate and should increase the friction for users between their platforms and fringe platforms.

Beyond their community guidelines and content moderation policies, features available on social media platforms need to be designed with anti-hate principles in mind. Companies need to conduct a thoughtful design process that puts their users first and incorporates society's concerns before, and not after, tragedy strikes. Today, the most popular method of developing technology tools is through a Software Prototyping approach: an industry-wide standard that prompts companies to quickly release a product or feature and iterate on it over time. This approach completely devalues the impact of unintended design consequences. For example, the Christchurch shooter used Facebook's livestreaming feature to share his attack with the world. The feature could have been designed to limit or lock audiences for new or first-time streamers or prevent easy recording of the video.

Question 3. Earlier this year, ADL's Center on Technology and Society called on technology companies, including several of those testifying here before us, to release "transparency reports" providing details on how they define and identify hate speech, how they moderate hateful content, and the efficacy of these techniques.

- Mr. Selim, can you discuss why such reports are useful?

Answer. Knowledge on the efficacy of platforms' content moderation efforts at dealing with the problem of white supremacist activity remains extremely limited. Meaningful transparency will allow stakeholders to answer questions such as: "How significant is the problem of white supremacy on this platform?" "Is this platform safe for people who belong to my community?" "Have the actions taken by this tech company to improve the problem of hate and extremism on their platform had the desired impact?"

We can conduct external research to evaluate their efforts, but companies often do not share user data, limiting opportunities to collect and use data for research. Alternatively, we can review transparency reports on content moderation efforts published by technology companies, but these too offer very limited information.

Mainstream social media platforms have a few potentially relevant metrics to the issue of extremism, especially white supremacist extremism, that they share in their regular transparency reports. Though each platform provides its own metrics on extremist activity, the metrics published are limited across the board, they are self-reported by the companies, and we have no real way of knowing what content has been put into which category outside of the brief descriptions given by the platforms as part of their reporting.

In order to truly assess the problem of hate on social platforms, technology companies must provide meaningful transparency with metrics that are agreed upon and verified by trusted third parties and that give actionable information to users, civil society, government and other stakeholders. Until technology platforms are willing to actively engage external parties and meaningfully address their concerns through greater transparency efforts, our ability to understand the extent of the problem of hate and extremism online, or how to meaningfully and systematically address it, will be extremely limited.

- Have such reports been released?

Answer. Mainstream social media platforms publish transparency reports. These include Facebook, Twitter, and YouTube.

If we look at the published metrics characterized as being related to terrorism (Facebook reported 6.4 million pieces of content related to terrorist propaganda removed from January to March 2019), this may seem relevant. However, typically, social platform platforms define terrorism in terms of Al Qaeda and ISIS-related activity and do not include white supremacist violence or activity as part of the terrorism classification. White supremacist extremist content could be categorized as hate speech or violent content on a platform, but at the same time, so could a wide variety of other types of content not associated with extremism or white supremacy.

Moreover, when Facebook claims in their transparency report that they took action on four million pieces of hate speech from January to March 2019, we still have no sense of how that compares to the level of hate speech reported to them, what communities are impacted by those pieces of content or whether any of that content is connected with extremist activity on their platform.

YouTube provides more granularity, sharing a number of different categories of content reported by users as well as the amount of content in each category that YouTube actioned. That being said, the names of the categories actioned by YouTube differs from those reported by users, making a comparison between what is reported and actioned impossible, and providing in the end the same level of opaqueness as Facebook's report.

Twitter's transparency report on the other hand provides both the users reported to the platform and users actioned by the platform in identical categories, but does not provide any information on the amount of content reported versus amount actioned, making the scale of their activity similarly opaque.

With almost three and a half billion social media users worldwide—and one million users joining every day—social media platforms have turned to a mix of machine learning and human moderators to detect and take down hate speech, terrorist propaganda, cyber-bullying, and disinformation. Machine learning can be a useful tool in identifying objectionable content quickly, preventing it from spreading. However, there are concerns about its ability to understand the context of text or images, and the length of time it takes to train systems with new data to recognize objectionable content.

Question 3. Mr. Selim, earlier this year the ADL announced a partnership with the Network Contagion Research Institute (NCRI) to research how extremism and hate speech spread on social media. NCRI uses machine learning to expose hate on digital platforms. Can you talk tell us about your findings?

Answer. On October 27, 2018, Robert Bowers perpetrated the deadliest attack against Jews in American history when he stormed a Pittsburgh synagogue armed with an assault rifle and three handguns. Shouting "All Jews must die," Bowers killed eleven people in their place of worship. Within months, Brenton Tarrant perpetrated the deadliest attack against Muslims in New Zealand's history when he slaughtered 50 people gathered for prayer at two mosques. In the wake of these horrific crimes, Jewish and Muslim communities worldwide and concerned citizens across the globe began searching for clues about attacks that seemed to come out of nowhere.

In hindsight, though, these killings should not have been surprising. Both attackers were enmeshed in online communities that exposed them to content designed to make them hateful and potentially violent. Bowers was a member of a fringe online community called Gab, which, like similar online forums, is a bastion of hatred and bigotry. Gab has seen a surge in racist and anti-Semitic postings since the 2016 presidential election. Tarrant, too, was part of a fringe online community called 8chan, one of the most notoriously hateful online communities on the internet.

Platforms like these force us to reassess our understanding of how violence may be inspired by such hateful echo chambers. Even more broadly, as we have recently reported, mainstream platforms can sometimes push such individuals from an open community, such as Twitter, into fringe environments like Gab that foster acceptability of dangerous views.

In September 2018, the Network Contagion Research Institute and its partners published a study, also detailed in a Washington Post article, which indicates that the state of online echo chambers of hate is far worse than many may imagine. Analyzing more than 100 million comments and tens of millions of images posted between July 2016 and January 2018 to Gab and 4chan's "politically incorrect" message board (/pol/), the NCRI performed the largest quantitative study to date regarding the rise of anti-Semitism and white nationalism on these popular white supremacist web communities. The study shows that anti-Semitic slurs and content doubled on these platforms after the election of President Donald Trump. During the same timeframe, these web communities also showed a dramatic surge in the expression of racism, including a substantial increase in the use of the n-word slur.

NCRI's research also shows that these web communities influence the spread of hateful memes and images to more mainstream networks like Twitter and Reddit. This research (along with other studies) shows an uptick in hateful rhetoric on fringe web communities in the wake of significant political events or highly publicized extremist violence. Relatedly, some studies have similarly demonstrated that ethnic hate expressed on social media can cause surges in real-life hate crimes. The implications of this online-offline dynamic are highly concerning.

On Gab, Bowers demonstrated how online propaganda can feed acts of violent terror. On 8chan, Tarrant showed how violent terror can itself create online propaganda. In both cases, the shooters strongly signaled back to their fringe web communities with their criminal acts, as though they were including them as knowing coconspirators. In both cases, the participation of these fringe web communities proves to be key to the scope, sensationalism, and ideological thrust of the act. Moreover, both shooters claim the same twisted notion of "white genocide"—or the imminent destruction of the white race by Jews and people of color—as the motive behind their terrorist acts, suggesting a shared ideological motivation. In fringe online communities, many members indoctrinate other users based on the conspiracy propaganda of a "white genocide" not online violent extremists of other ideologies spreading a grievance used to justify their malign views.

The evidence overwhelmingly suggests that such platforms can serve to spread modern violent extremism in ways that could not have been predicted from the early days of social media. Gab and 8chan fan the flames of bigotry and hatred and organize violent fantasies in online communities even as they fuel them in the real world.

There is no telling who else on Gab or 8chan may take cues from Bowers and Tarrant and act on the violent ideologies they derive from these online communities. In essence, these platforms serve as round-the-clock white supremacist rallies, amplifying and fulfilling their vitriolic fantasies.