



**Written Testimony of Derek Slater
Director, Information Policy, Google LLC**

**United States Senate Committee on Commerce, Science, and Transportation
“Mass Violence, Extremism, and Digital Responsibility”
September 18, 2019**

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 US 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 US 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% of the 9 million videos we removed in the second quarter of 2019 were first flagged by our automated systems. More than 80% 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.