



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

House Committee on Homeland Security
“Examining Social Media Companies' Efforts to Counter Online Terror Content
and Misinformation”

June 26, 2019

Chairman Thompson, Ranking Member Rogers, and distinguished members of the Committee: Thank you for the opportunity to appear before you today. I appreciate your leadership on the important issues of radicalization and misinformation online, and welcome the opportunity to discuss Google’s work in these areas.

My name is Derek Slater, and I am the Global Director of Information Policy at Google. In my role, I lead a team that advises the company on public policy frameworks for online content -- including hate speech, terrorism, and misinformation. 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.

At Google, we believe that the Internet has been a force for creativity, learning, and access to information. Supporting this free flow of ideas is core to our mission to organize and make the world’s information universally accessible and useful. We build tools that empower users to access, create, and share information like never before — giving them more choice, opportunity, and exposure to a diversity of opinions. Products like YouTube, for example, have expanded economic opportunity for small

businesses to market and sell their goods; have given artists, creators, and journalists a platform to share their work, connect with an audience, and enrich civic discourse; and have enabled billions to benefit from a bigger, broader understanding of the world.

While the free flow of information and ideas has important social, cultural and economic benefits, there have always been legitimate limits, even where laws strongly protect free expression. This is true both online and off, especially when it comes to issues of terrorism, hate speech, and misinformation. We are deeply troubled by the increase in hate and violence in the world, particularly by the acts of terrorism and violent extremism in New Zealand. We take these issues seriously and want to be a part of the solution.

This is why, in addition to being guided by local law, we have Community Guidelines our users have to follow. We also work closely with government, industry, and civil society to address these challenges in partnership within the United States and around the world. In my testimony today, I will focus on two key areas where we are making progress to help protect our users: (i) on the enforcement of our policies around terrorism and hate speech; and (ii) in combatting misinformation broadly.

Enforcement on YouTube for Terrorism and Hate Speech

We have rigorous policies and programs to defend against the use of our platform to spread hate or incite violence. This includes: terrorist recruitment, violent extremism, incitement to violence, glorification of violence, and videos that teach people how to commit terrorist attacks. 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.

Tough policies have to be coupled with tough enforcement. Over the past two years, we have invested heavily in machines 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 it is somewhat similar to videos that 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. In the first quarter of 2019, over 75% of the more than 8 million videos removed were first flagged by a machine, the majority of which were removed before a single view was received.
- 2) Machine learning technology is what helps us find this content and enforce our policies at scale. But hate and violent extremism are nuanced and constantly evolving, which is why we also rely on experts to find videos the algorithm might be missing. Some of these experts sit at our intel desk, which proactively looks for new trends in content that might violate our policies. We also allow 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) Finally, we go beyond enforcing our policies by creating programs to promote counterspeech on our platforms to present narratives and elevate the voices that are most credible in speaking out against hate, violence, and terrorism.
 - a) 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. There have been 59 million views of 2018 Creators for Change videos so far; 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.

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

This broad and cross-sectional work has led to tangible results. In Q1 2019, YouTube manually reviewed over 1M suspected terrorist videos and found that only fewer than 10% (90K videos) violated our terrorism policy. Even though the amount of content we remove for terrorism is low compared to the overall amount our users and algorithms flag, we invest in reviewing all of it out of an abundance of caution. As comparison point, we typically remove between 7 and 9 million videos per quarter—a fraction of a percent of YouTube's total views during this time period. Most of these videos were first flagged for review by our automated systems. Over 90% of violent extremist videos that were uploaded and removed in the past 6 months (Q4 '18 & Q1 '19) were removed before receiving a single human flag, and of those, 88% had fewer than ten views.

Our efforts do not end there. We are constantly taking input and reacting to new situations. 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. This would include, for example, videos that promote or glorify Nazi ideology, which is inherently discriminatory. It also prohibits content denying that well-documented violent events, like the Holocaust or the shooting at Sandy Hook Elementary, took place. We began enforcing the updated policy the day it launched; however, it will take time for our

systems to fully ramp up and we'll be gradually expanding coverage over the next several months.

Similarly, the recent tragic events in Christchurch presented some unprecedented challenges and we had to take some unprecedented steps to address the unprecedented volume of new videos related to the events--tens of thousands, exponentially larger than we had ever seen before, at times coming in as fast as one per second. In response, we took more drastic measures, such as automatically rejecting new uploads of clips of the video without waiting for human review to check if it was news content. We are now reexamining our crisis protocols, and we've been giving a lot of thought to what additional steps we can take to further protect our platforms against misuse. Google and YouTube also signed the Christchurch Call to Action, a series of commitments to quickly and responsibly address terrorist content online. The effort was spearheaded by New Zealand's prime minister to ensure another misuse of online platforms like this cannot happen again.

Finally, we are 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. During Google's chairmanship of the Global Internet Forum to Counter Terrorism over the last year and a half, the Forum sought to expand its membership and to reach out to a wide variety of stakeholders to ensure we are responsibly addressing terrorist content online. For example, we hosted a summit in Sunnyvale so G7 security ministers could hear the concerns of smaller platforms. We have also convened workshops with activists and civil society organizations to find ways to support their online counter-extremism campaigns, and sponsored workshops around the world to share good practices with other tech companies and platforms.

Combating Misinformation

We have a natural, long-term incentive to prevent anyone from interfering with the integrity of our products. We also recognize that it is critically important to combat misinformation in the context of democratic elections, when our users seek accurate, trusted information that will help them make critical decisions. We have worked hard to curb misinformation in our products. Our efforts include designing better ranking algorithms, implementing tougher policies against monetization of misrepresentative content, and deploying multiple teams that identify and take action against malicious actors. At the same time, we have to be mindful that our platforms reflect a broad array of sources and information and there are important free-speech considerations. There is no silver bullet, but we will continue to work to get it right, and we rely on a diverse set of tools, strategies, and transparency efforts to achieve our goals.

We make quality count in our ranking systems in order to deliver quality information, especially in contexts that are prone to rumors and the propagation of false information (such as breaking news events). The ranking algorithms we develop to that end are geared toward ensuring the usefulness of our services, as measured by user testing. The systems are not designed to rank content based on its political perspective.

Since the early days of Google and YouTube, some content creators have tried to deceive our ranking systems in order to increase their visibility, a set of practices we view as a form of spam. To prevent spam and other improper activity during elections, we have multiple internal teams that identify malicious actors wherever they originate, disable their accounts, and share threat information with other companies and law enforcement officials. We will continue to invest resources to address this issue and to work with law enforcement, Congress, and other companies.

In addition to tackling spam, we invest in trust and safety efforts and automated tools to tackle a broad set of malicious behaviors. Our policies across Google Search,

Google News, YouTube, and our advertising products clearly outline behaviors that are prohibited, such as misrepresentation of one's ownership or primary purpose on Google News and our advertising products, or impersonation of other channels or individuals on YouTube. We make these rules of the road clear to users and content creators, while being mindful not to disclose so much information about our systems and policies as to make it easier for malicious actors to circumvent our defenses.

Finally, we strive to provide users with easy access to context and a diverse set of perspectives, which are key to providing users with the information they need to form their own views. Our products and services expose users to numerous links or videos from different sources in response to their searches, which maximizes exposure to diverse perspectives or viewpoints before deciding what to explore in depth. In addition, we develop many tools and features to provide additional information to users about their searches, such as knowledge or information panels in Google Search and YouTube.

Conclusion

We want to do everything we can to ensure users are not exposed to content that promotes or glorifies acts of terrorism. Similarly, we also recognize that it is critically important to combat misinformation in the context of democratic elections, when our users seek accurate, trusted information that will help them make critical decisions. Efforts to undermine the free-flow of information is antithetical to our mission. We understand these are difficult issues of serious interest to the Committee. We take them seriously and want to be responsible actors who are a part of the solution.

We know that our users will value our services only so long as they continue to trust them to work well and provide them with the most relevant and useful information. We

believe we have developed a responsible approach to address the evolving and complex issues that manifest on our platform.

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.