



**Written Testimony of Neal Mohan  
Chief Product Officer, YouTube and SVP, Google**

**U.S. Senate Committee on Homeland Security and Governmental Affairs  
“Social Media’s Impact on Homeland Security: Part II”  
September 14, 2022**

Chairman Peters, Ranking Member Portman, and distinguished members of the Committee; thank you for the opportunity to appear before you today. My name is Neal Mohan, and I am the Chief Product Officer for YouTube. In my role, I am responsible for YouTube’s products, user experience, and trust and safety globally. I appreciate the opportunity to provide testimony to the Senate Homeland Security and Governmental Affairs Committee.

At YouTube, our mission is to give everyone a voice and show them the world. Every day, we build and improve tools and systems that empower creators, viewers, and businesses to find and share information like never before. Billions of people come to YouTube to enjoy videos and music, upload original content, and learn new skills. Our work creates new economic opportunities for artists, creators, journalists, and small businesses to share their creativity and products in the United States and across the globe.

Openness has always been one of our guiding principles, but our commitment to openness works hand in hand with our responsibility to protect our viewers, creators, and partners from harmful content. Responsibility is our top priority at YouTube and informs every product and policy decision we make.

My testimony today will focus on (1) our approach to Responsibility at YouTube; (2) our policies against harmful content; and (3) our collaboration across industry to combat terrorist content online.

**Youtube’s Four Rs of Responsibility**

At YouTube, we have four pillars of Responsibility, described in detail below, which we refer to as “the Four Rs.” Used together, these pillars have enabled us to make substantial progress in tackling harmful content, as well as to respond quickly in the face of unprecedented events such as the COVID-19 pandemic and the war in Ukraine. Under the Four R approach, we

*Remove* content that violates our policies as quickly as possible; we *Raise up* authoritative sources when people are looking for news and information; we *Reduce* the spread of misinformation and content that brushes up against our policy lines; and we *Reward* trusted, eligible creators and artists.

**Remove Violative Content:** YouTube is an open platform but that has never meant anything goes. Our Community Guidelines provide public-facing rules of the road for content on YouTube and we remove content that violates these rules as quickly as possible. We work continuously to improve our efforts through enhanced detection and enforcement, relying on a combination of technology and people. Our automated detection systems are an area of significant investment for YouTube, and our engineering teams continuously evaluate their efficacy and make improvements. These systems help our human review teams remove content at scale, with the speed and volume that could not be achieved with people alone.

We are proud of the advancements we have made in our enforcement efforts as a result of our investment. In the first half of 2022, we removed close to 8.4 million videos for violating our Community Guidelines, more than 92% of which were first flagged by our automated systems. Approximately 67% of those first flagged by our systems received 10 or fewer views before we removed them. During these two quarters, we removed more than 130,000 videos specifically for promoting violence and violent extremism in violation of our policies. As noted above, our approach enables us to adapt quickly in the face of unprecedented events, including those that may lead to harmful disinformation on our platform. For example, since the war in Ukraine began, YouTube has removed more than 76,000 videos and 9,000 channels related to the ongoing war in Ukraine for violating our Community Guidelines and Terms of Service.

In addition to removing violative content, we track the percentage of views on YouTube that comes from this content. We refer to this metric as the Violative View Rate, or VVR. In the first half of 2022, the VVR was .09% - .11%. In other words, for every 10,000 views on YouTube, only 9-11 went to content that proved to be violative.

In order to hold ourselves accountable to our users and the public at large, we publish a quarterly Community Guidelines Enforcement Report that provides data on VVR and our enforcement across channels, videos, comments, user flags, and appeals and reinstatements.

**Raise Up Authoritative Voices:** YouTube is a source for news and information for people around the world—whether about events unfolding in local communities or more existential global issues like the COVID-19 pandemic. Not all queries are the same, however. For topics like music or entertainment, relevance, newness, and popularity are most helpful to understand what people are interested in. But for subjects such as news, science, and historical events, where accuracy and authoritativeness are key, the quality of information and context are paramount. Our search and recommendations systems are designed to raise up

authoritative voices in response to user queries that are “news-y” or related to topics prone to misinformation.

We also recognize that there are topics that may require additional context, which we provide through our information panels. These panels appear in the search results and video watch pages, with a link to an authoritative source. Topics include, for example, the holocaust, moon landing, and election-related information such as how to register to vote.

**Reduce the Spread of Borderline Content:** Equally important to raising up authoritative information is reducing the spread of borderline content, or content that does not quite cross the line of our policies for removal but that we don’t necessarily want to recommend to people. We use machine learning to reduce the recommendations of this type of content, including potentially harmful misinformation, and as a result of our efforts it represents just a fraction of what is watched on YouTube in the United States. We are able to raise up authoritative information and reduce borderline content by using classifiers to identify whether a video is “authoritative” or “borderline.” These classifications rely on human evaluators who assess the quality of information in each channel or video. These evaluators hail from around the world and are trained through a set of detailed, publicly available rating guidelines. We also rely on certified experts, such as medical doctors when content involves health information. Taken together, all of our responsibility work around recommendations has shown real impact. Watchtime of authoritative news is up dramatically and borderline viewing is down.

We use a number of signals to recommend videos to users, including a user’s language, time of day, and user satisfaction. But responsibility outweighs these considerations: if a video is identified as borderline content, it will be demoted in recommendations. As a result of our efforts, we saw a 70% drop in watchtime on this type of content in the US in 2019, and continue to keep views below 1% of total views of content on YT.

**Reward Trusted Creators:** Finally, we are proud to be a place where creative entrepreneurs can build thriving businesses, and we reward trusted creators. However only creators that meet a high bar can make money on YouTube.

More than 2 million creators around the world are part of our YouTube Partner Program (YPP). Through YPP, we share the majority of our advertising revenue with our partners, and offer a variety of ways to make money, including ads and channel memberships. Over the past three years we’ve paid out more than \$30 billion to creators, artists and media companies (as of November 2020).

In order to join YPP and earn money on YouTube, creators must meet specific eligibility requirements and comply with our monetization policies. Individual videos must follow our advertiser-friendly content guidelines to earn money.

Channels that repeatedly brush up against our monetization policies or our Community Guidelines—including hate speech, harassment, and misinformation—will be suspended from the YouTube Partner program, meaning they can't run ads on their channel or use other monetization features like Super Chat. To protect our ecosystem of creators, advertisers, and viewers, we also tightened our advertising criteria in 2017. After careful analysis and conversations with creators, we changed certain eligibility requirements for monetization, which significantly improved our ability to identify creators who contribute positively to the community, while also preventing potentially inappropriate videos from monetizing content.

As a testament to our responsibility, we were the first digital platform to receive accreditation for content level brand safety from the Media Rating Council (MRC). This means that MRC found us to be 99% effective at ensuring ads only show up where they should. We also provide metrics in all four categories of the Global Alliance for Responsible Media's cross-industry Aggregated Measurement Report, which evaluates platform safety for brands and consumers.

### **YouTube's Robust Policies To Address Violent Extremist Content**

As noted above, our Community Guidelines set forth what content is not allowed on YouTube. With respect to violent extremist content, we have a network of robust policies in place that work together to combat it.

First, our Community Guidelines prohibit terrorist organizations from using YouTube for any purpose, including recruitment, as well as content that promotes terrorism, glorifies terrorist acts, or incites violence. To complement the policies aimed at terrorist organizations, the Guidelines also include policies that address violent extremism, violent or graphic content, harmful or dangerous content, election misinformation, and hate speech. Together, these policies work to prohibit key categories of harmful content regardless of who posts it. More details about these policies is below.

- **Harmful or Dangerous Content:** YouTube prohibits content that encourages dangerous or illegal activities that risk serious physical harm or death. This policy applies to various types of conduct, including content that provides instructions to build a bomb meant to injure or kill others, and content that promotes or glorifies violent tragedies, such as school shootings.
- **Firearms:** YouTube prohibits content intended to sell firearms, instruct viewers on how to make firearms, ammunition, and certain accessories, or instruct viewers on how to install those accessories. We also do not allow live streams that show someone holding, handling, or transporting a firearm, unless they are uniformed police or military.

- **Harassment & Cyberbullying:** YouTube prohibits content that threatens individuals or targets an individual with prolonged or malicious insults based on intrinsic attributes. This policy includes, for example, direct or implied threats, or sharing someone's personally identifiable information online, known as doxxing.
- **Hate Speech:** YouTube removes content promoting violence or hatred against individuals or groups based on age, caste, disability, ethnicity, gender identity and expression, nationality, race, immigration status, religion, gender, sexual orientation, veteran status, or victims of a major violent event and their kin. This policy also 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. We also prohibit content denying that well-documented violent events, like the Holocaust or the shooting at Sandy Hook Elementary, took place.
- **Violent Criminal Organizations:** YouTube prohibits content intended to praise, promote, or aid violent criminal organizations. Examples of content that violates this policy would be videos or comments directing users to sites hosting manifestos from the perpetrators of well-documented violent events or content that is aimed at recruiting new members to violent criminal or terrorist organizations designated by the U.S. government.
- **Violent or Graphic Content:** YouTube prohibits violent or gory content intended to shock or disgust viewers, or content encouraging others to commit violent acts. This includes, but is not limited to, content that incites others to commit violent acts against individuals or a defined group of people; fights involving minors; and content where animals are encouraged or coerced to fight by humans. It also applies to footage, audio, or imagery involving road accidents, natural disasters, war aftermath, terrorist attack aftermath, street fights, physical attacks, sexual assaults, immolation, torture, corpses, protests or riots, robberies, medical procedures, or other such scenarios with the intent to shock or disgust viewers. Further, footage or imagery showing bodily fluids such as blood or vomit with the intent to shock or disgust viewers is prohibited by this policy.

Content that violates our policies against violent extremism, detailed above, includes material produced by organizations designated by the U.S. government as 'foreign terrorist organizations'. We do not permit these terrorist organizations to use YouTube for any purpose. Content produced by violent extremist groups that are not government-listed foreign terrorist organizations is subject to our policies, including those described above, regardless of who posts the video, or the group to which they belong.

We regularly review and update our policies to address new and emerging threats. For example, in June 2019, we strengthened our hate speech policy to specifically prohibit videos alleging that a group is superior to justify discrimination, segregation or exclusion based on

qualities like age, gender, race, caste, religion, sexual orientation or veteran status. In December 2019, we updated our harassment policy to take a stronger stance against threats and personal attacks, and introduced new penalties for patterns of harassing behavior. In October 2020, we expanded our hate and harassment policies to prohibit content that targets an individual or group with conspiracy theories that have been used to justify real-world violence, like QAnon.

In addition to our efforts to remove violative content produced by individuals and organizations, Google deploys efforts to counter government-based influence operations. We take the security of our users very seriously, and we have dedicated teams in place to protect against attacks from a wide range of sources. YouTube works closely with Google's Threat Analysis Group (TAG), which is dedicated to protecting users from threats posed by state-sponsored malware attacks and other advanced persistent threats. On any given day, TAG is tracking more than 270 attacker groups from more than 50 countries. When we detect attempts to conduct coordinated influence operations on our platforms, whether state-backed or otherwise, we swiftly remove offending content from our platforms and terminate these actors' accounts. We take steps to prevent possible future attempts by the same actors, and share intelligence to prevent attacks elsewhere.

This ongoing work includes the removal of Chinese threat actors seeking to conduct operations on our platforms and target our users. In the first 6 months of 2022, we terminated more than 24,000 of these as a result of investigations into coordinated influence operations linked to China. Data about YouTube channel terminations made as a result of these investigations is included in TAG's quarterly Bulletin.

Due to the evolving nature and shifting tactics of groups promoting violative content, we continuously review and adapt our policies to stay ahead of bad actors. Combating these threats is an area where we constantly look to build stronger defenses, including technological mechanisms to defend our platform at scale.

### **YouTube's Collaborative Work to Curtail Terrorist or Extremist Content**

In light of the increasingly interconnected nature of the information ecosystem, we collaborate across the industry to combat terrorist and violent extremist content on the Internet.

As noted above, in 2016, we collaborated with industry partners to develop a hash-sharing database where we share "digital fingerprints" of terrorist content to stop its spread. The shared database includes more than 320,000 unique hashes of both video and still image

media, both videos and images. Since 2017, the number of companies contributing to and benefiting from this database has grown from 4 to 13.

We also share best practices on counterterrorism with our industry peers through the Global Internet Forum to Counter Terrorism (GIFCT), which is dedicated to disrupting terrorist abuse of digital platforms. With GIFCT partners, we developed a content incident protocol to enable a swift and coordinated response to perpetrator-created content across platforms. GIFCT also collaborates with the Tech Against Terrorism initiative to hold workshops with more than 100 smaller tech companies around the world.

\* \* \*

Thank you, Mr. Chairman, for convening this important hearing. Responsibility is and will continue to be YouTube's number one priority—our business literally depends on it. We look forward to continuing to work with you to address these challenges. Thank you.