Supplementary written evidence submitted by Google

Letter from Peter Barron, VP, Communications and Public Affairs Google EMEA, to the Chair of the Committee, 30 March 2017

Thank you for giving us the opportunity to provide evidence to your Committee as part of your inquiry into ‘hate crime and its violent consequences’ on 14th March 2017.

We wanted to respond to the questions raised by the Committee and set out our plans to address your concerns. These are complex issues and we are deeply committed to playing our role and using our resources - both in terms of expertise and technology - to tackle them. We already have thousands of people working on trust and safety issues across the company and have invested hundreds of millions of pounds in tackling abuse of all kinds on our platforms.

This is, however, an evolving challenge and we know we must do more. We work closely with Government, industry, community groups and our users to develop effective responses to inappropriate and illegal content on our services and are committed to continuing our efforts and investment in this area.

Content policies and enforcement on YouTube

Our starting point for the content we allow on our platforms is of course the law, and we act quickly to block content when notified of anything illegal, such as terrorism, hate crime and child sexual abuse imagery.

In addition to our legal removals process, YouTube has established community guidelines that, for example, prohibit ‘predatory behavior, stalking, threats and harassment’ or ‘violent or gory content that is primarily intended to be shocking, sensational, or disrespectful’. We want people to use the platform to share ideas and express their views but we strive to find the right balance between free speech - exposing to debate views that we might not agree with - and firm action against hateful content wherever it crosses the line. We recognise that not everyone will agree with where we draw that line. The highly offensive David Duke video that we discussed in the hearing is a good example. Many thought this video should be removed on grounds of anti-semitism but others - including the editor of the Jewish Chronicle - argued it should stay up to be exposed and challenged. We have highly trained teams around the world making these difficult judgments daily and these decisions are not taken lightly. They are discussed and debated by our legal and policy teams, and are often escalated to our senior leadership team. We recognise that we do not always get it right and are always striving to improve.

When considering our removal decisions and content policies, it is worth noting that we give careful thought to context. The same piece of content - uploaded by different people, in different contexts, with different intents - may be handled differently. Take National Action for example. As a proscribed terrorist organisation, we blocked their channel as soon as we were notified, and we block videos that invite support for National Action, including the ones flagged to us in your recent letter. However, news or documentaries, or content condemning National Action may still be permitted - for example a recent BBC documentary on radicalisation remains available although it includes an interview with a spokesman for National Action.

We have one of the largest teams of policy and content experts in the world and have built out the industry’s first large scale flagging system - a model that has since spread across our industry. We have also built up a network of expert trusted flaggers - such as the Counter-Terrorism Internet Referral Unit - who proactively monitor our platforms and have become valuable enforcement partners. To further scale
our efforts we’ll be announcing plans to significantly extend our Trusted Flagger programme, including by providing financial support for additional enforcement activities. Our hope is that ultimately this network will operate not just across our platforms but across social media more broadly.

Our systems are constantly evolving to deal with the scale and challenges online. On YouTube:

- We receive over 200,000 flags a day
- We action 98% of flagged content within 24 hours.
- We removed 92 million videos in 2015, up from 14 million in 2014, through a mix of user flags and technology.
- We’ve received user flags from across 196 countries.
- We terminated 2.8 million YouTube channels globally in 2016 for posting content that violates our policies or the law.
- We comply with the EU Code of Conduct on countering illegal hate speech online, "reviewing the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, where necessary."

We are proud of the progress made over recent years but we recognise we need to continue to develop, invest in and improve our processes. We recently announced that we are taking a hard look at the content we allow on our platforms and have started a review of our community guidelines on hate speech, which will benefit from conversations with many of the NGOs we work with. We will keep the committee informed of any future changes we make.

Measures to proactively detect harmful content

Whilst we harness the power of our diverse community of over a billion users to flag inappropriate content, we do not rely solely on our flagging system. We use technology to help us proactively detect hateful and illegal content in a number of different ways:

- **Tackling hateful comments** - In February, Google and Jigsaw (formerly Google Ideas) launched a new tool for publishers to identify potentially toxic and hateful comments on their websites. Last year, we launched similar technology on YouTube, where we provided an AI-classifier to creators so they can moderate potentially offensive comments. This technology has been in development for several years and will continue to improve as the classifiers (predictive machine learning models) develop over time.

- **Fingerprinting** - We have used matching technology to help prevent the re-upload of content that violates our policies since the early days of YouTube. To help address child sexual abuse imagery (CSAI), we developed ground-breaking video fingerprinting technology and created a service to be shared across the industry to combat CSAI. Recently, YouTube announced a partnership with Facebook, Microsoft and Twitter to share hashes across our systems to enable us to scan proactively for egregious terrorist content and enable us to terminate associated infringing accounts.

- **Fighting abusive accounts** - We use signals and patterns to prevent abuse of our products at scale by bad actors who intend to harm our users. We are constantly improving the technology that helps us detect and prevent bad actors from creating accounts to spam and abuse users across our products.

We are committed to further research and development to identify new ways in which technology - and particularly machine learning - can help us in the fight against hate speech online.

Our guidelines for advertising
We have a strict set of policies for monetising content on Google and YouTube and a robust enforcement system. Last year, we removed over 1.7 billion ads that breached our guidelines and ejected 100,000 publishers from our AdSense programme. We also prevented ads from serving on over 300 million videos inappropriate for advertising. But we know there is room for improvement and that we need to do more. We have apologised to advertisers for recent failures in our system which resulted in brands appearing against inappropriate content, and announced changes in three areas:

- **Policies** - We will update and improve our policies to raise the bar on videos and sites allowed for advertising, taking a tougher stance on hateful, offensive and derogatory content. This includes removing ads from content that demeans people based on their race, religion, gender or similar categories.

- **Advertiser controls** - We already have a wide variety of tools to give advertisers and agencies control over where their ads appear, such as topic exclusions and site category exclusions. It is clear that we need to do more. It is our job to make the controls easier and better understood. We will add new controls that make it simpler for brands to control where ads appear and introduce safer default settings for all advertisers on both YouTube and the Google Display Network.

- **Enforcement** - We have been increasing, and will further increase, our investment in enforcing our advertising guidelines, both in terms of technology and people. We will use AI and machine learning to increase our capacity to detect questionable content and block it from advertising.

You can find more details of our expanded safeguards for advertisers in this blogpost, published last week.

As you will be aware, we met the Home Secretary earlier today - alongside other tech companies - to discuss ways in which we can improve our collective response to tackling extremism online. We would be happy to meet you in the coming weeks to update you on our plans and would like to invite the entire Committee to visit our Trust and Safety team in Dublin, where we can provide additional information on our processes on a confidential basis. In the meantime, I hope these clarifications and updates will be helpful.

Yours sincerely,

Peter Barron
VP, Communications and Public Affairs Google EMEA