Google UK
1 St Giles High Street
London
WC2H 8AG

Lord Evans of Weardale KCB DL
Chair, Committee on Standards in Public Life
Room G07
1 Horse Guards Road
London
SW1A 2HQ

Dear Lord Evans

I'm writing in response to your letter of 8th September, on the progress Google has made on the recommendations of the Committee's report on Intimidation in Public Life. The issues that your report noted are very important to Google and we welcome the opportunity for continued engagement. Given the nature of our different platforms, much of our response below relates to content on YouTube.

Since the report was published in 2017, we have continued to take action to ensure that our YouTube Community Guidelines and policies are current and that content which violates them is identified and removed from our platforms as quickly as possible. We do not want or allow content that incites hatred on our platforms. Google builds its product for all users, from all political views around the globe, and the long term success of our business is directly related to our ability to earn and maintain the trust of our users.

I have outlined the progress we have made on the specific areas noted by the Committee below. Our recent white paper on information quality and content moderation, may also be useful and be accessed here.

If you have any further questions or require any further information, please do not hesitate to contact me.

Yours sincerely

Katie O'Donovan
Head of Government Affairs and Public Policy, Google UK

Identifying violative content

The Committee's report made a number of recommendations on tackling inappropriate content. YouTube is acting on this issue, working to identify and then remove content that is either illegal or that violates our community guidelines.

For each of our products, we have a specific set of rules and guidelines on use, content and the risk of harm. These are supported by clear mechanisms to report violative content and extensive mechanisms to detect and remove it.

We use both human moderators and machine learning systems to remove harmful content - the latter is increasingly effective, allowing us to identify and remove, at speed, often before a single human user has seen it. We have invested heavily in both of these methods and updated our systems in a way which has changed the architecture of our platform:

- We've continued to grow our machine learning capability to allow us to quickly and efficiently review and remove more content, including comments, that violates our guidelines. We've invested in people. We now have over 10,000 people working to address content that might violate our policies.
- We update our policies on a very regular basis. Since January 2019 we've launched over 30 different changes on YouTube, including changes to reduce recommendations of borderline content, content that could misinform users in harmful ways and content that comes close to - but doesn't quite cross the line of - violating our Community Guidelines. Each of our policies is designed to ensure they are consistent, unbiased, well-informed, and can be applied to content from around the world. They're developed in partnership with a wide range of external industry and policy experts. Internally, we have over 100 people working on policy development, deployment, and enforcement consistency, made up of policy and enforcement experts all over the world.

In addition, YouTube has a Trusted Flaggers program which provides robust tools for individuals, government agencies, and NGOs that are particularly effective at notifying YouTube of content that violates our Community Guidelines. Our trusted flaggers can also report new trends they observe through a dedicated form.

Some examples of our trusted flaggers include Stop Hate UK, the anti-hate crime organisation, and the Counter Terrorism Internet Referral Unit (CTIRU) from the Home Office and Metropolitan Police. They have access to additional features, such as the ability to flag content that violates Community Guidelines in bulk, and the content they flag also gets prioritised for review. Once a trusted flagger has flagged content, our human reviewers will review it and decide whether it should be removed on the basis of our community guidelines. Our trusted flagger programme has enabled us to build strong working relationships with these expert organisations, helping us to have an open dialogue on potential grey areas and evolving norms.

Taking action on violative content, accounts and comments

Alongside investing in identifying content which violates our guidelines, we also have strict policies in place for accounts that host violative content.

When our reviewers determine that content violates our Community Guidelines, we remove the content and send a notice to the creator. The first time that a creator violates our Community Guidelines, the creator receives a warning. After one warning, we'll issue a Community Guidelines strike to the channel and the account will have temporary restrictions. Channels that receive three strikes within a 90-day period will be terminated. If an account is terminated, that person won't be able to access their previously posted content or allowed to create any new accounts.

Channels that are dedicated to violating our policies or that have a single case of severe abuse of the platform will bypass our strikes system and be terminated. All strikes and terminations can be appealed if the creator believes that there was an error, and our teams will re-review the decision.

Our systems are increasingly effective at removing violative content before it has been viewed. For example, between April and June 2020, we removed over 11.4 million videos from YouTube for violating our guidelines and 10.8 million of these were flagged by machines. Our transparency report (see above) from April-June 2020 shows that of the over 11.4 million videos removed, 42% had never received a single view, and more than 75% had been viewed fewer than 10 times. In the UK, we removed 156,822 videos during this time period.

We use a combination of smart detection technology and human reviewers to crack down on hateful comments which have no place on our platform. Between April and June 2020 we removed over 2 billion comments from our platforms.

We've also empowered our users on YouTube by introducing new policies to give them more control over what other people or organisations are allowed to post in comments on their channel. This includes tools to allow users to moderate comments on their videos, which have been used by over one million channels, and the introduction of penalties and removal for serious or repeated violations.

On YouTube, account holders can delete inappropriate comments and block any user they wish so they can't view videos or leave more comments. Comments can also be turned off for any video by the uploader or managed by requiring pre-approval before they are posted publicly, and any user can report comments as abusive. Users can also block comments containing certain words from appearing on their videos. YouTube accounts are penalised for community guideline violations, and serious or repeated violations will lead to account termination. If an account is terminated, that person won't be able to access their previously posted content or allowed to create any new accounts.

Enhanced transparency

One of the Committee's key recommendations focused on public reporting of content removals. In 2018, YouTube began publishing a quarterly transparency report which provides aggregate data about the flags we receive and the actions we take to remove videos and comments that violate our content policies.

We also publish a transparency report for Google services more broadly, which provides details on, for example, the number of requests we have received from Government or law enforcement

to remove or delist content.

We are committed to publication of these transparency reports which now include details on our efforts to enforce our community guidelines and terms of service on a country-by-country basis, for example, the number of YouTube videos removed per quarter in the UK, as well as globally.

Escalating illegal content

Our systems for addressing illegal content include a dedicated channel through which governmental and other authorities report illegal content, following a fast-track assessment of the notification, and multilingual moderation teams who are highly trained to review and assess content.

YouTube operates a notice-and-action system for users and governmental authorities to report content which may be unlawful in local jurisdictions. When we receive notifications to remove allegedly illegal content, we review each notification carefully. We provide a tool to help users report content that they believe should be removed from YouTube based on applicable laws, through a form that seeks appropriate information to help us resolve the matter as quickly as possible.

The option to report or "flag" content that breaches Community Guidelines is available under every YouTube video and comment, and we receive over a hundred thousand flags a day in this manner from an engaged and diverse global community. Any logged-in user can flag a video by clicking on the three dots to the bottom right of the video player and selecting "Report." Trained teams evaluate videos before taking action in order to ensure it actually violates our policies and to protect content that has an educational, documentary, scientific, or artistic purpose. The teams carefully evaluate flags 24 hours a day, seven days a week.

In the first quarter of 2020, we received 12,155,210 flags on YouTube (globally). These flags helped us to identify and take action against content that does not meet our Community Guidelines.

We have also introduced 'Reporting History Dashboards', to enable individual users to track any action taken on videos they report and are pleased to be able to provide more information and confidence to those who use our flagging system.

In addition, YouTube's Trusted Flaggers program provides robust tools for individuals, government agencies, and NGOs that are particularly effective at notifying YouTube of content that violates our Community Guidelines. Our trusted flaggers can also report new trends they observe through a dedicated form.

Election periods

As we outlined in our letter to the Committee last year, we convened a cross-functional team focused on the 2019 general election and we continued our dialogue with the Home Office, security, policing and electoral authorities throughout the campaign to ensure that safety and reporting guidance was able to reach the widest possible audience of candidates and electoral

staff.

We also published candidate safety guidance on the Internet Association's website, detailing our Hate Speech, Harassment and Cyberbullying, and Harmful or Dangerous Content policies, and provided detailed instructions on how to report content that violates these policies. Prior to the last general election, we also wrote to all sitting MPs and political party headquarters to outline the security and safety measures we made available and to offer in-person briefings.

We will take the lessons of the election period forward and look to build on our ability to help protect candidates online in future election periods.