Google UK
Central St Giles
1 St Giles High Street
London
WC2H 8AG

Lord Bew
Committee on Standards in Public Life
Room GC07
1 Horse Guards Road
London
SW1A 2HQ

Dear Lord Bew,

Thank you for your recent letter regarding progress made against the Committee on Standards in Public Life's recommendations, following its review about "intimidation in public life".

Google's mission is to organise the world's information and make it universally accessible and useful. We believe that technology's ability to promote education and engagement brings extraordinary value to our society. However, we recognise that we have a duty to ensure that our platforms are used responsibly. We want to make sure that users have the tools and knowledge they need to make responsible choices online, and that they are able to flag and report abuse, which is acted upon.

We have developed a number of initiatives to make progress in the areas outlined in the committee's report and have set out our progress against your specific recommendations below.

Developing and implementing automated techniques for content removal and takedown times

Our goal is to make it harder for policy-violating content to surface or remain online. We have taken a number of steps to protect our community on YouTube by tightening our policies on what can appear on the platform and investing in new machine learning technology to scale the efforts of our human moderators. We are also using our cutting-edge machine learning more widely to allow us to quickly and efficiently review and remove content that violates our guidelines for YouTube.

As part of this, we are now publishing monthly 'Transparency Reports', which include details of videos taken down during that quarter. The reports contain data on the flags that YouTube receives and how we enforce our policies.

Machine learning (ML) is helping our human reviewers remove nearly five times as many policy-violating videos than they were previously - it is particularly effective at highlighting violent extremist content. YouTube removed 7.7 million videos during the second quarter of 2018, 6.8 million of which were first flagged through our automated flagging system. Of those videos removed by machine learning, 76.5% had no views at the time of takedown. We're confident that our algorithms and systems will continue to improve.

Because we have seen such positive results using the technology to tackle violent extremist content, we have begun training machine-learning technology across other challenging content areas, including abuse, hate speech and child safety. We also use machine learning classifiers to identify hate, harassment and bullying comments. We know that ML is not perfect so we still use human reviewers to help us achieve tackle content that contravenes our community guidelines.

We now have more people reviewing more content, with the goal of bringing the total number of people across Google working to address content that might violate our policies to over 10,000 by the end of 2018.

With open systems such as YouTube, no system can ever be perfect and our work to keep our platforms safe will be constant as we adapt to evolving threats. Our algorithms and systems will continue to get better. Challenges to our platform are constantly evolving and changing, so our enforcement methods must and will evolve to respond to them. No matter what challenges emerge, our commitment to combat them will be sustained.

Preventing users from receiving hostile messages and tools to enhance users ability to tackle online intimidation

We are committed to our users' safety and have strong community guidelines in place to ensure this. 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.

In the last few months we've used machine learning to help human reviewers find and terminate hundreds of accounts and shut down hundreds of thousands of comments. On YouTube, account holders can also 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. Users can also block comments containing certain words from appearing on their videos.

Transparency reporting

We share the Committee's view that providing users and policy makers with relevant information on technology platforms' community guidelines, reporting procedures and subsequent action can help create a greater understanding of, and confidence in, the measures platforms take to deal with content relating to harms identified by the Government's internet safety strategy.

Since Spring 2018 we have been publishing quarterly transparency reports on YouTube's guidelines enforcement. You can find the detailed information included in each report here, they include:

- Videos removed: Number of total videos removed in the quarter and how the removals were first detected (by our automated flagging systems, users, or Trusted Flaggers).
- Automated flagging removals: Percentage of videos flagged through our automated flagging system that were removed before views/after views.
- Human flagging: Number of total unique videos flagged by human flaggers, and the number of human flags by type of flagger
- Top 10 countries by human flagging volume: The ten countries from which we have received the most human flags.
- Flagging reasons: The reasons selected by human flaggers at time of flagging.

We recently launched new '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.

We have also produced a video 'The life of a flag' which is designed to help build understanding, awareness and trust in our flagging system. This compliments our already easy to understand user guidelines which we updated and redesigned late in 2017.

Your letter also argues for a UK specific approach to transparency reporting, as does the draft transparency report proposed by the government. This contains many requests for information which we already share with our users and the public more generally, but also includes requests for information which are not relevant to our platform, not legally permissible or aren't pertinent to the way we process data. We continue to talk to DCMS about their plans for detailed transparency reports.

Escalating reports to the police

Google has a strong collaborative working relationship with the Met Police and are working proactively with the Mayor's Office for Policing and Crime and community groups on the removal of hateful content online. The Met is a trusted flagger and we have ongoing dialogue on a variety of issues to ensure we are aware of any new cases or emerging trends.

We are also members of the Online Hate Crime Hub which has allowed us to develop new Trusted Flagger partnerships and ensure we are aware and take action on incidents of hate crime flagged across the UK.

More broadly, we are expanding the network of academics, industry groups and subject matter experts who are critical partners that help us better understand emerging issues. We also work with NGOs and trusted flaggers to report illegal behaviour to the police.

Social media and election campaigns

Google provides training to candidates and political parties ahead of planned election campaigns. This includes dealing with threats or intimidation online and ensuring that political parties and candidates are aware of all of the tools available to them. We anticipate providing this training well ahead of the next planned UK general election.

Google also has a programme called ‘Protect Your Election’. This provides free tools to help provide additional online protection for candidates, as well as tools to help protect campaigns against Distributed Denial of Service (DDoS) attacks.

I do hope this information is useful and do let me know if you have any further questions.

Yours sincerely,

David Skelton