Written supplementary evidence submitted by Google (OSB0218)

Thank you very much for your invitation for Google to participate in the joint committee's evidence session on the Draft Online Safety Bill, dated 28 October.

We are keen to participate fully in Parliamentary committees, answering questions and providing detailed clarity on our approach, policies, implementation and further areas of importance. Markham and Leslie were pleased to be able to share with the committee our work on protecting users (especially children) on YouTube, Search and Ads. As ever, we are very pleased to follow up in writing regarding some of the questions posed during the session.

1. **YouTube approach to videos that have been recommended before removal and the "Plandemic" film**

In April this year, YouTube announced the release of a new metric called Violative View Rate (VVR) as part of our publicly available Community Guidelines Enforcement Report. VVR helps us estimate the percentage of the views on YouTube that come from violative content. Our data science teams have spent more than two years refining this metric, which we consider to be our North Star in measuring the effectiveness of our efforts to fight and reduce abuse on YouTube.

Between April and June 2021, our overall VVR was 0.19%–0.21%, meaning that for every 10,000 views of content on YouTube, 19-21 of those are of violative content. For more information on our methodology, please see the third-party analysis that was recently conducted by MIT Professor of Statistics, Arnold Barnett, available here.

In relation to your specific question about 'Plandemic', we quickly remove flagged content that violates our Community Guidelines, including content that features medically unsubstantiated diagnostic advice for COVID-19 and re-uploads of the original clip if they contain segments that we deem to be violative of YouTube's Community Guidelines. From the very beginning of the pandemic, we've had clear policies against COVID-19 misinformation and are committed to continue providing timely and helpful information at this critical time.

Between May 4 and May 18, 2020, we removed thousands of violative Plandemic videos; over 30% of those removed had zero views, and 89% had 100 or fewer views.

2. **YouTube research into the harmful content**

Our Violative View Rate (VVR), which we've mentioned above, helps us determine what percentage of views on YouTube comes from content that violates our policies. Our teams started tracking this back in 2017, and across the company it's the primary metric used to measure our responsibility work and assess user exposure to harmful content. We have added historical data for this metric to our Transparency Report, showing that, since Q4 of 2017, we

have seen a 70% drop in VVR. This reduction is due in large part to our investments in machine learning to identify potentially violative content. VVR does not take into account any individual user characteristics and is based on a sample of views.

When it comes to recommendations, we recently published a blog post that provides more details about how our recommendations system works, including the signals that are taken into account when we recommend content: clicks, watchtime, survey responses, sharing, likes, and dislikes. As the post discusses, we work to actively reduce recommendations of content that may come close to violating our policies but does not cross the line. In January 2019, we announced changes to our recommendations systems to limit the spread of this type of content. These changes resulted in a significant drop in watchtime on non-subscribed, recommended borderline content that year. While algorithmic changes take time to ramp up and consumption of borderline content can go up and down, our goal is to have views of non-subscribed, recommended borderline content below 0.5%. We seek to drive this number to zero, but no system is perfect; in fact, measures intended to take this number lower can have unintended, negative consequences, leading legitimate speech to not be recommended. As such, our goal is to stay below the 0.5% threshold, and we strive to continually improve over time.

3. **Hate speech identification and removal on YouTube**

Hate speech is not allowed on YouTube. Our hate speech policy specifically prohibits content that encourages or glorifies violence against individuals or groups, or whose primary purpose is to incite hate against an individual or group based on protected attributes. Neither do we allow denialism, trivialization or minimization of well-documented violent events.

Detecting hate speech content and enforcing our policies on a global scale does have its challenges. For hate speech content, the context can be a very important factor. This is why we rely on a combination of machines to detect content at scale and human reviewers to make more complex decisions.

We publish information about our removals for violations of our hate speech policies in our publicly available Community Guidelines Enforcement Report. Between April and June this year we removed over 6.2 million videos for violating our Community Guidelines globally. Of these, over 5.9 million were first detected by our automated systems. During the same quarter, we removed over 87,000 videos globally and more than 3,000 videos uploaded in the UK for violating our Community Guidelines relating to hate speech.

4. **Google ads, misinformation and Newsguard study**

We have several long standing policies in place to prevent ads from running alongside unreliable and harmful claims and content promoting hate or violence. We sometimes reshare guidance, particularly when we see worrisome trends or areas of potential confusion. Recently,

we've seen an uptick in content that violates certain policies and reminded publishers of our requirements and their obligations.

Our policies against misrepresentative content prohibit (among others) harmful health claims, or claims that relate to a current, major health crisis and contradict authoritative scientific consensus. This includes misinformation related to the COVID-19 pandemic, such as claims that the virus is not real, false claims about vaccines, and content promoting unsubstantiated remedies or treatments. When we find content that violates this policy, we stop ads from serving. Our enforcement can be as targeted as removing ads from an individual violating page. We escalate our enforcement to the site-level when violations are persistent and egregious.

We reviewed this report and removed ads from several URLs as well as entire sites that we found to be in violation of our publisher policies.

It's also important to note that all publisher policies apply both to publisher-generated content, such as news articles, as well as user-generated content, such as comment sections. All publishers monetizing on our platforms have to comply with our policies, regardless of political affiliation, and we enforce our policies consistently without exceptions or bias. Our policies are put in place for many reasons, including ensuring good experience for people viewing our ads, to prevent user harm, and to help make sure that ads follow applicable laws in the countries where they appear. They are also there to instill trust in our advertiser partners that their ads are running against content that is appropriate.

Additionally, Google also provides advertisers with robust tools that lets them decide where their ads appear. These controls allow advertisers to exclude specific websites and URLs as well as entire topics that they wish to avoid when running ads. This year, we announced a new feature that allows advertisers to upload dynamic exclusion lists that can be updated and maintained by trusted third parties or by the advertisers themselves. Once advertisers upload these exclusion lists to their accounts, they can schedule automatic updates as their third party partners add new sites, ensuring that their exclusion lists are comprehensive and always up-to-date. This new feature allows advertisers to fully leverage the resources and expertise of trusted organizations to better protect their brands and strengthen the impact of their campaigns.

5. Impact of turning off autoplay on YouTube for kids

The changes we announced this summer included a number of features to better protect the wellbeing of YouTube users, including launching the autoplay toggle on YouTube Kids and setting that toggle to default off. The goal of the Autoplay toggle is to provide users with choices so they can make the right decision for their families. This and a number of other steps we introduced aimed at encouraging more active choices by users about how they want to spend their time online were implemented in September 2021 and so, having been live for only a short period of time, we are continuing to gather and analyze the data on the impact of these

changes. In the coming months, we'll also be launching additional parental controls in the YouTube Kids app, including the ability for a parent to choose a "locked" default autoplay setting.

6. **Confirmation or clarification of the figure that 70% of YouTube videos that are watched are recommended to users.** (Source)

When YouTube's recommendations are at their best, they connect billions of people around the world to content that uniquely inspires, teaches, and entertains. Our recommendation system is built on the simple principle of helping people find the videos they want to watch and that will give them value.

The 70% statistic cited above was one calculated in 2018, prior to the significant efforts we undertook to reduce recommendations of borderline content (explained in detail above). Our system doesn't follow a set formula, but develops dynamically as user viewing habits change. Today, recommendations do drive a majority of the overall watchtime on YouTube, even more than channel subscriptions or search, but the precise percentage of watchtime that comes from recommendations fluctuates. However, we do know that consumption of borderline content as a result of recommendations is significantly below 1%, and our goal is to have views of borderline content from recommendations below 0.5% of overall views on YouTube.

Thank you again for the opportunity to share these details about how our systems protect users. Please do not hesitate to get in touch if we can help with anything else.

9 November 2021