Supplementary written evidence submitted by Facebook (OHC0037)

Many thanks for your letter of 14 May to Neil Potts, I am replying on his behalf. My apologies that this response is far later than I would have liked.

We greatly appreciated the opportunity to give evidence to your Committee on the progress we have made in tackling hate speech on the platform over the past eighteen months. A number of specific points were raised during the course of the evidence session, and in your subsequent letter, and our response to these is set out below.

During the session, you raised the fact that a small number of videos of the tragic Christchurch attack were still accessible on our platform. We have investigated this further and found that the videos in question were highly edited versions of the original which our automated systems were unable to detect. We have subsequently removed these from our platforms.

As we stated in the evidence session, Facebook has been working closely with the New Zealand Police as they responded to the attack and we continue to support their ongoing investigation. We removed the original Facebook Live video within minutes of the law enforcement's outreach to us and it was hashed, allowing for shares of visually similar videos to be detected and automatically removed from Facebook and Instagram. In the first 24 hours, we removed about 1.5 million videos of the attack globally. More than 1.2 million of those videos were blocked at upload. Some variants of the video, however, such as screen recordings and the examples you raised, have proven more difficult for our systems to detect. As a result we have also expanded to use additional detection systems, including the use of audio technology to ensure as much of this content is removed as possible.

As a founding member of the Global Internet Forum to Counter Terrorism (GIFCT), we have also cooperated closely with colleagues across wider industry, sharing more than 800 visually-distinct videos related to the attack via our collective database, along with URLs and context on our enforcement approaches. We have also established a $7.5 million research partnership with the University's of Maryland, Cornell, and Berkeley to look at ways to improve our image and video analysis technology.

Turning to your specific questions. You asked about links to the Daily Stormer website. We removed the 8 Daily Stormer links you provided to us, in addition to 2 other Pages we found ourselves. We will continue to remove any Daily Stormer content unless it is being shared to condemn the material or message. Finally, we have prohibited known variations of the Daily Stormer domain name from being uploaded to platform.

You asked about what oversight we have of closed groups. The key point to note is that closed groups are not closed to our systems or our rules. Our Community Standards apply in closed groups, all of the content in closed groups can be reported by group members in the usual way and our AI detection systems for content such as bullying, hate speech, terrorism or child exploitation material all apply to closed groups. We do not have figures for how many fake accounts there may be in groups because where we know about fake accounts we remove them, as our most recent transparency report showed we disabled around 2 billion fake accounts in the first 3 months of this year. I am afraid we do not have figures for how many groups contain more than 10k users.

You asked about sharing identifying data of our users with law enforcement if the UK legislated to make that a requirement. Facebook respects the law in the countries within which we operate. And there are a number of existing legal routes by which information can be shared between us and law enforcement, for example, legal requests made under the Regulation of Investigatory Powers Act in the UK and the Mutual Legal Assistance Treaty (MLAT). We publish statistics on the law enforcement requests for data we receive. In the UK last year, we received 15,000 data requests of which we complied with over 90%. We also have an extremely constructive on-going dialogue with UK law enforcement and the intelligence services about the best way to work together.

If we see evidence of a threat of imminent harm or a terror attack, we reach out to law enforcement. We have teams working across the globe around the clock who are ready to alert authorities when we become aware of an emergency involving imminent harm to someone's safety. We have a dedicated law enforcement liaison team - including former UK law enforcement officers from the National Crime Agency - who are in daily contact with police forces across the country on issues of shared concern including organised hate groups, terrorist incidents and harassment and intimidation.

We also work with law enforcement to help build cases to bring criminals to justice — including the National Crime Agency's Child Exploitation and Online Protection Command (CEOP) and the Government's Counter-Terrorism Internet Referral Unit (CTIRU).

The issue of how companies like ours can most usefully and effectively support the efforts of UK law enforcement is an issue that falls within the remit of the Government's Online Harms White Paper. We have today responsed to the White Paper and we look forward to working closely with Government, Parliament and law enforcement to identify the appropriate framework.

You asked a number of questions related to our Enforcement report. Firstly I can confirm that the numbers in the report are subject to a rigorous methodology (which is explained in the report) and so are accurate. We also worked with independent external experts in measurement, statistics, criminology and governance to gather feedback and validation of our approach and metrics we've shared in the report. We are exploring their feedback and will continue expanding and updating our report to provide meaningful and accurate metrics in the future.

Our transparency report does not include how much content is reported by users as we do not believe this is a meaningful number. We see a very high volume of reports for content which is completely trivial. As described above we work hard to ensure the metrics we use are genuinely meaningful. For example, we have recently added data on how the amount of content where our decision is appealed and what proportion of that content is restored on appeal - giving a better idea of how accurate we are.

Setting out the different types of action taken on content would be a challenging task. The consequences for violating our Community Standards vary depending on the severity of the violation and a person's history on the platform. For instance, we may warn someone for a first violation, but if they continue to violate our policies, we may restrict their ability to post on Facebook or disable their profile. There are also a wide range of actions users themselves can take to remove or hide content they do not wish to see. We're exploring other areas to

release metrics in future reports. This also applies to your question around providing a breakdown of action taken in relation to hate speech - there are a range of actions, which are informed by a number of factors including an individuals history on the platform and real world context etc which make this kind of data highly complex to gather accurately.

You asked about data on the speed with which content is actioned. We have found that time to takedown content is not a meaningful measure. This is because some content could be removed within minutes but still accumulate many views, while other content could linger for months but be seen by no one. As such time to takedown is less meaningful to measure speed of removals. See https://newsroom.fb.com/news/2019/05/measuring-prevalence/ on why we measure prevalence instead which is a measure of how many views violating content gets on the platform, which we report in our transparency report.

Finally, I understand my colleagues have been in contact regarding the specific comments made in reference to you in a group on the platform. These comments have now been deleted, as have the users who posted them, and we have disabled a number of the group's administrators. Similarly, we will be following up directly with Mr. Doughty in relation to the organisations he raised during the course of the session and which are designated as hate organisations under Facebook's policies.

I hope that the above information is useful. We look forward to continuing to work with the Committee and the Government to ensure we take every step we can to eradicate hate speech online and ensure the safety of all of our users. I would also like to reiterate our offer to host you and the Committee on a visit to our offices in London and Dublin to meet the engineers and safety and security team who are working to develop our technologies and enforce our policies in this space. Should you have any further questions, please do not hesitate to contact me directly.

Rebecca Stimson
Head of Public Policy, UK

1 July 2019