Yvette Cooper MP
House of Commons
London, SW1A OAA

30 August 2019

Dear Ms Cooper,

Thank you for your letter of 16th July. I am sorry the Committee was not satisfied with my previous response. I also apologise that we have been unable to respond to your letter as quickly as the Committee requested. However the additional time has allowed us to, I hope, provide a fuller answer to your questions. Taking your points in turn:

You asked why we were unable to find the group you referenced during our evidence session. It is worth repeating that the comments made in reference to you in the group in question were deleted shortly after you raised them with us, as have the users who posted them, and we have disabled a number of the group's administrators. As you can imagine there is a huge volume of groups and comments on Facebook, so we usually need a link to the actual content which someone is reporting (this is how our online reporting system works) to ensure we find the exact content in question. The verbal references made to this content during the evidence session were not sufficient for our teams to locate the precise comments you were raising. This is why we encourage people to report using our reporting tools, and additionally for MPs to use the dedicated reporting channel we established for them a few months ago, as this gives us the information we need to quickly locate and review the exact piece of content.

You ask a number of questions in your letter about the approach that Facebook takes towards content in private groups. As you note in your letter, private groups can be important places for people to come together and share issues which they may not want to discuss in a public forum, like identifying as LGBTQ or discussing challenges around a health condition. But we recognise that they can also be used to share inappropriate content.

As previously explained, our Community Standards apply to private groups, every single piece of content can be reported to us and all of our proactive AI detection systems run on private groups in the usual way. The enforcement of our standards within groups does not therefore solely rely on the admins or members of those groups to report content which violates our rules. We use a variety of proactive methods, in addition to user reports, and no one moderation effort is primary or independent.

Broadly speaking the proportion of content we remove using reactive measures vs. using proactive measures (the proactive rate) is about the same between open and closed groups; a group being private does not make a material difference. This behavior is consistent across all the violation types that are reported in our enforcement report and the proactive rate in groups is very similar to what is reported in our enforcement report. We have taken action on over 600M pieces of content proactively within private groups this year. Additionally, this year we've taken action on nearly 5M pieces of content within private groups that were first reported to us by users.

Our ability to detect and remove violating content is improving all the time, as our quarterly enforcement report demonstrates. But the effectiveness of technology at finding violating content varies between different types of content. For example in Q1 2019 99.3% of the terrorist content we removed was proactively found by our systems. The percentage of hate speech found and removed before anyone reported it increased to 65%, up from 24% just over a year ago. But by contrast we currently only proactively remove 14.1% of bullying content. The reason for this is the more a piece of content relies on a nuanced understanding of context, the less well suited technology is to determining whether or not its violating content.

Threatening language, such as death threats, is an example of one such area. At the moment, automated systems are not as good at detecting comments (such as threats or bullying) which require a more sophisticated interpretation of what's being said and also often rely heavily on context which only a human could understand. For example, it is extremely challenging for AI to distinguish between a benign turn of phrase, something comedic or satirical or something malicious. This is why for this type of content we currently still need to rely more on reporting and human review. However, we continue to invest heavily in our automated capability and all user reports are used to inform the development of our classifiers over time, so it's a process of continuous improvement.

You asked for specific information related to our policy to hold group admins more accountable for content violations within a group and how many reports we have seen from admins of private groups. We have been testing this new policy in Spain and are planning on rolling it out globally shortly. At this stage we do not therefore have reliable performance data we can share with the Committee. But we would be happy to provide an update next year when we are able to gather the relevant information.

It is worth keeping in mind that the number of reports is not usually a meaningful metric and not one we tend to rely on in our own internal analysis. Firstly, as described above, user reports are only one part of the picture alongside proactive techniques to find and remove content. So they do not give a complete picture. Secondly, user reports themselves are not a reliable proxy for violating content. They can run the gamut from serious reports of, for example, hate speech, to users reporting a spoiler for a TV show. Thirdly there may be a big difference between the number of reports and number of actions taken. It is for these reasons we do not tend to use it as a metric.

Lastly, you asked a number of questions about the size and numbers of private groups which are moderated by UK users. There are just under 25k private groups with over 10k members where at least one of the admins is a UK user. Globally there are around 680k groups with over 10k members and about 620k with 5k - 10k members

I hope the information in this letter is helpful for the Committee's ongoing inquiry.

Yours sincerely,

Rebecca Stimson
Head of Public Policy, UK