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

16th November 2018

Dear Lord Evans,

Congratulations on your appointment as Chair of the Committee on Standards in Public Life. I am writing in response to Lord Bew's letter following up on the Committee's December 2017 review on Intimidation in Public Life.

Facebook's mission is to give people the power to build community and bring the world closer together. Facilitating political debate is an important part of that. We believe Facebook is a powerful resource for MPs and those seeking political office, enabling them to engage with constituents, share information about their work and hear feedback. But intimidation of public figures online is a real challenge.

We recognise that many candidates at the 2017 UK General Election experienced such intimidation, and we understand that some continue to receive abuse. Political abuse isn't unique to Facebook as a platform, but we know we have a responsibility to tackle it. To do this, we need to hear from candidates, political parties and other experts about how we can improve our tools and policies in this area, and the Committee's review and its recommendations have been an important contribution to this work. I am setting out below a number of updates on our progress against the issues your review identified.

Taking Down Intimidatory Content

Facebook is keenly aware of the need to balance the right of constituents to criticise their politicians against our responsibilities as a platform to ensure that threatening language is removed when we become aware of it. This is always a challenging balance to get right.

That's why in April this year we published the internal guidelines we use to enforce our Community Standards, to make clear why and how we reach decisions about what content should remain on our services. They are available here: www.facebook.com/communitystandards. Our Content Policy team develops these standards in close consultation with experts around the world, and they apply across all our platforms. We believe that making these guidelines public enables our users, including those active in public life, to better hold us to account for the standards we set and how we enforce them.

Address: 10 Brock Street
Regent Place
London, England
NW13FG

Our Community Standards make clear that we have zero tolerance for any type of hate speech or credible threats of any kind. Every piece of content has a report button, and when we are made aware of instances that violate our policies, we act. This includes not only removing content, but referring cases to law enforcement when we become aware of an imminent threat, and encouraging users throughout the reporting process to contact law enforcement immediately if they or another person are in any danger.

Our Community Operations teams are based in several locations across the globe, so that when something is reported to us we have teams are available 24/7. Their numbers include experts in numerous areas, including hacked accounts, spam, hate speech, child safety, counter terrorism and many more. These teams strive to support people who need our help across a broad number of topics.

Last year we were one of the first companies to confirm how many people we had working on reviewing content reports and committed to increasing from 4000 to 7500 by the end of 2017. In the event we significantly exceeded this. Overall, in 2018 we have committed to doubling the size of the teams working on safety and security from 10,000 to 20,000. This combination of scale and expertise allows us to make decisions quickly and consistently about what should be allowed on our platform, and what should not.

We fully comply with all laws in the countries we operate, including the German Network Enforcement Act, known as 'NetzDG'. To meet our obligations under this law, we designed a new, additional reporting flow for people to report content they think is unlawful under the German criminal code. While this means we are fully compliant with NetzDG, we, like others, remain skeptical that outsourcing the determination of illegality to private companies, rather than courts, is the right approach. Some public cases have highlighted the complexity of the law and the difficulty for companies to judge the illegality of content in edge cases. German legal experts have arrived at different conclusions after considering the same set of facts and legal principles. And it is relevant to note that in the first quarter of 2018 the respective online form for complaints was only used 239 times, when the German Government had anticipated 25,000 yearly complaints.

Automated Techniques for Content Removal

In line with the Committee's recommendations, our Community Operations team are making efforts to use new technologies to help us rid our platform of intimidatory content. Our product team builds essential tools like artificial intelligence, smart automation and machine learning that help us remove much of this content with no need for prior involvement from our team members. At times we also make use of techniques that we have developed help us take action against potential phishing links, scammers, fake accounts and other types of abuse.

Address: 10 Brock Street
Regent Place
London, England
NW13FG

Our technologies are becoming increasingly sophisticated, and more and more of the content that violates our policies is taken down before it is brought to our attention. But at present there are still limits to that technology, especially in cases where context is key like hate speech and bullying. We therefore continue to rely on the thousands of content reviewers we have all over the world to enforce our rules in difficult to judge cases, and to build the library of examples that our AI tools need to get better at proactively taking down this material.

We continue to invest in technology to constantly improve our capabilities. We are, for example, experimenting with ways to filter the most obviously toxic language in comments so they are hidden from posts. As we roll out new techniques we will keep the public updated about our progress via our Newsroom at https://newsroom.fb.com/.

Transparency of Performance Data

The Committee's recommendations called for transparency about how social media companies enforce their policies online. We at Facebook are strongly committed to an effective, fair and transparent approach to the moderation of content on the platform. That's why yesterday we published our second Community Standards Enforcement Report: https://transparency.facebook.com/community-standards-enforcement.

This second report shows our enforcement efforts on our policies against adult nudity and sexual activity, fake accounts, hate speech, spam, terrorist propaganda, and violence and graphic content, for the six months from April 2018 to September 2018. We are getting better at proactively identifying violating content before anyone reports it, specifically for both hate speech and violence and graphic content. But there are still areas where we have more work to do.

- Since our last report, the amount of hate speech we detect proactively, before anyone reports it, has more than doubled from 24% to 52%. The majority of posts that we take down for hate speech are posts that we've found before anyone reported them to us. This is incredibly important work and we continue to invest heavily where our work is in the early stages and to improve our performance in less widely used languages.

- Our proactive detection rate for violence and graphic content increased 25 percentage points — from 72% to 97%

We're not only getting better at finding bad content, we're also taking more of it down. In Q3 2018, we took action on 15.4 million pieces of violent and graphic content. This included taking down content, putting a warning screen over it, disabling the offending account and/or escalating content to law enforcement. This is more than 10 times the amount we took action on in Q4 2017. This increase was due to continued improvements in our technology that allows us to automatically apply the same action on extremely similar or identical content.

Address: 10 Brock Street
Regent Place
London, England
NW13FG

The Committee will want to note that this second report also includes two new categories of data, one of which is bullying and harassment. Bullying and harassment tend to be personal and context-specific, so in many instances we need a person to report this behavior to us before we can identify or remove it. This results in a lower proactive detection rate than other types of violations. In the last quarter, we took action on 2.1 million pieces of content that violated our policies for bullying and harassment — removing 15% of it before it was reported. We proactively discovered this content while searching for other types of violations.

The fact that victims typically have to report this content before we can take action can be upsetting for them. We are determined to improve our understanding of these abuse types so we can get better at proactively detecting them.

Tools and User Options to tackle Intimidatory Content and Messages

The Committee was right to point out that the ability to report and remove intimidatory content is only one part of tackling this issue. People in public life also want to be able to prevent this content from appearing in their online profiles at all. That's why we have developed a range of tools that allow our users to moderate and filter the content that people put on their Pages.

People who help manage Facebook Pages, like those for candidates or causes, can hide or delete individual comments. They can also proactively moderate comments and posts by visitors turning on the profanity filter, or blocking specific words or lists of words that they do not want to appear on their Page. When people include a word that has been blocked in a post or comment on the user's Page, the post will be automatically marked as spam. Page admins can block different degrees of profanity from appearing on their Page. We determine what to block by using the most commonly reported words and phrases marked offensive by the community, and these blocked terms now automatically apply across a range of different languages. Page admins can also remove or ban people from their Pages using the straightforward tools available to them as adminstrators.

We are constantly iterating the tools and options available to our users to improve their ability to control their experience on Facebook. Last month, we introduced a way for people to hide or delete multiple comments at once from the options menu of their posts. We continue to talk to MPs and others in public life to understand the specific issues they face and we will take on board the concerns raised by the Committee as we develop further tools. We use our Newsroom page to publish updates on our progress: https://newsroom.fb.com/

Protecting Candidates and Election Campaigns

Facebook takes seriously our responsibility to provide advice, guidance and support to Parliamentary candidates, especially during election campaigns.

Address: 10 Brock Street
Regent Place
London, England
NW13FG

For each national election that takes place, Facebook establishes a dedicated task force of at minimum 20-30 individuals from teams across the company, to focus on protecting the integrity of that election on the platform. One of the functions of this team is to coordinate responses to reports of inappropriate content and ensure that action is taken as necessary.

As part of this process, we create a dedicated email channel which we provide to all parties to share with candidates as appropriate and which they can use as a single point of contact to report content (including but not limited to hate speech, threats of violence or real world harm). This is staffed by trained team members who are able to escalate reports as necessary and respond to emails quickly. We also have a team of content analysts with specific market expertise who are able to quickly assess content reported via this channel and take action where required.

We already employ a range of measures to support candidates on safety, security and reporting malicious behaviour during an election period. We have a dedicated team which focuses on outreach to elected officials and candidates. Ahead of scheduled elections, or even during a snap election period, this team will reach out to parties and relevant government bodies such as the Electoral Commission to provide them with advice and guidance on safety and security on the platform, which they can share with candidates.

In addition, we email all candidates who have Facebook pages with details of how to turn on two-factor authentication, a crucial way to protect accounts from being hacked, and other techniques to keep their online presence secure. We also deliver a dedicated product to the Newsfeed of all candidates containing the same information. Examples of the guides we share are available here: https://politics.fb.com/wp-content/uploads/2018/08/safety_page-admins.pdf

During the 2017 snap election, our team undertook a tour of the UK to talk to candidates about safety and security on the platform. We want to do much move ahead of future elections. Ahead of any forthcoming UK election, the team will work with the political parties to hold training events for candidates, including first-time candidates, to ensure they are aware of the tools available to them to protect themselves on the platform.

We want to make sure the right information is easily available to candidates at the places they go for help, which will often be their party contacts. That's why we will reach out to the party HQs and the right figures within the community of returning officers to make sure they are kept informed about the resources available, so that all candidates have a trusted place they can turn when they have concerns.

Outside of election periods, the team works closely with the political parties other groups to ensure that members of both Houses have access to resources and understand the range of tools available to them in order to ensure they are protected.

Address: 10 Brock Street
Regent Place
London, England
NW13FG

We agree with the Committee that there is more we can do in this area, and we are reaching out via new channels including the Parliamentary and Diplomatic Protection branch of the Metropolitan Police. We also have a publicly available website, politics.fb.com, which provides insight and advice on best practice across a range of areas, including protecting account safety and security.

I hope you find this information helpful. We look forward to continuing to work with the Committee and the Government to ensure we take every step we can to protect the integrity of UK elections and give constituents the power to hold their representatives to account.

Yours sincerely,

Karim Palant
UK Public Policy Manager, Facebook

Address: 10 Brock Street
Regent Place
London, England
NW13FG