

APPG on  
Social Media



**All Party Parliamentary Group for Social Media meeting with Facebook to discuss preventing abuse of public figures**

**Attendees:**

Chris Elmore MP, Chair of the APPG for Social Media  
Damian Hinds MP, Vice Chair of the APPG for Social Media  
Sir Peter Bottomley MP  
Lord Lipsey  
Alexandra Landes, Office of Chris Elmore MP  
John Maier, Office of Chris Elmore MP  
Nansi Morgan, Office of David Linden MP  
Ryan Pratt, Office of Lord Taylor of Warwick

**Apologies:**

Lord Clement-Jones  
Carla Lockhart MP  
Lisa Cameron MP  
Lord Taylor of Warwick

**Secretariat:**

Michael Tunks, Internet Watch Foundation/ UK Safer Internet Centre  
Abi Fedorovsky, Internet Watch Foundation/ UK Safer Internet Centre

**Representatives from Facebook:**

Richard Earley  
Megan Thomas  
Ella Fallows  
Kathy Lee

**1. Introductions-**

Chris Elmore MP opened the session and welcomed Members of the APPG and representatives from Facebook, before passing over to the Facebook team to take the group through the presentation.

## 2. Presentation from Facebook-

Megan explained Facebook's mission was to create strong communities by enabling people to connect online.

The company now has 35,000 people working on safety and security in the UK and Facebook takes a five-step approach to online safety including: Partnerships, Policies, Tools, Resources and Feedback.

The purpose is to promote a safe and healthy online environment but one that also enables people to express their views online.

Facebook is investing heavily in new technology and tools which help the platform achieve that objective as well as providing help and resources to its community.

Facebook were keen to run this session to provide parliamentarians with insight into how their platform works ahead of the Online Safety Bill. It is crucial that Parliamentarians scrutinise the legislation carefully with a good understanding of platforms.

Facebook shares the same objective to make its platform as safe as possible. A recent survey in Germany showed that when people see bad content online, it makes them less likely to want to use it.

Kathy Lee explained Facebook's community standards apply globally. This is challenging as what maybe illegal in one jurisdiction, is not illegal in another, and therefore Facebook consults widely to consider different perspectives when developing its community standards and policies.

When developing policies, Facebook considers and prioritises certain principles: voice, privacy, safety, authenticity, and dignity.

Facebook takes three steps to develop its community standards, they must be principled, operable, and explicable. Kathy stressed the policies must work for the platforms 2.7 billion active users and the community guidelines are not static and are responsive to the needs of its users.

Richard Earley explained the companies hate speech policy. He explained hate speech was not tolerated on Facebook and that it was defined by protected characteristics.

Ella Follows explained that there was then a need to define what would constitute an attack. Calls to violence or dehumanising speech, statements of inferiority and calls for segregation are all defined as attacks under the policy.

Attacks on institutions are defined slightly differently. For example, it would not be tolerated to say I hate Danes, but it would be tolerated to say I hate Denmark (the policies of its Government).

Anything that incites violence or incitement to violence- death, serious injury or physical harm to an individual is not tolerated on the platform. Facebook also works with local law enforcement in these cases.

Richard explained that the company uses classifier technology to detect content before it is uploaded to the platform. The classifiers consider whether something violates community standards, it removes the content and tags the removed content.

97% of hate speech is detected before it was uploaded, and the platform does have an appeals process so users can appeal any decisions made by classifiers inaccurately.

Damian Hinds MP asked representatives to clarify the difference in the company's policy between public figures and private persons.

Facebook has clear guidelines which explain their approach in this area on the following link: <https://www.facebook.com/communitystandards/bullying>

The policy states:

'We distinguish between public figures and private individuals because we want to allow discussion, which often includes critical commentary of people who are featured in the news or who have a large public audience. For public figures, we remove attacks that are severe as well as certain attacks where the public figure is directly tagged in the post or comment. For private individuals, our protection goes further: we remove content that's meant to degrade or shame, including, for example, claims about someone's sexual activity. We recognise that bullying and harassment can have more of an emotional impact on minors, which is why our policies provide heightened protection for users between the ages of 13 and 18.'

Lord Lipsey asked about the accuracy of the detection rate and how does the company know what it isn't detecting?

Richard explained the company is investing in technology to push up what it is finding proactively. In recent transparency reporting the company is trying to develop a new figure around prevalence. This means if you open Facebook 10,000 times, we are trying to estimate the percentage chance of someone encountering hate speech online to give a more accurate representation of the problem.

Lord Lipsey asked about how Facebook's performance compared with other platforms?

Richard responded by stating there is a need for regulatory oversight which would help to draw some of these comparisons.

Sir Peter Bottomley stated that the numbers on Facebook appeared to be low, but what proportion of the overall problem was on Facebook's network?

Richard responded by stating it was Facebook's intention to make this a difficult place to conduct hate speech and harm. They work with organisations like Hope Not Hate and GIFCT to share information and give away technology to small providers to help improve their response to the problem.

The concern, however, is that in improving the response of the larger platforms, that these harms could be displaced to smaller platforms.

Damian Hinds asked about content moderators being based in the UK in relation to understanding racist terminology directed at footballers?

Richard stated it would depend upon the nature of the violation. To moderate nudity for example, that could be done from anywhere in the world. Other reports requiring more context would be

categorised and then depending on the categorisation could be reviewed by UK moderators, who are then supported by a pyramid system of management review.

Ella explained how the platform empowers public figures to control who can comment on your posts, both on your profiles and posts. This has only been in place for four weeks and at this stage there is no detail on how effective this has been, but Facebook would be happy to share with APPG Members once an evaluation has been carried out.

Facebook provides guidance on how to make sure you have page moderation tools turned on, which is [here](#), and they have introduced new controls on who can comment on posts, [here](#).

They also flagged the dedicated reporting form for politicians which is [ukpol@fb.com](mailto:ukpol@fb.com). You and your team can use it to report content you are concerned about from your page or elsewhere on the platform.

Facebook has three classification standards-

Illegal- goes against laws

Harmful- goes against community standards

Inappropriate content- subject to the individual public account holder.

Facebook provides several tools to assist those with public profiles including a profanity filter and comment controls. These tools can be used to hide language terms from appearing and can be further moderated by blocking emojis or keywords as chosen by you.

Comment controls mean you can determine who comments on your posts and who replies.

On Instagram, it is possible to filter offensive words on direct messages. This is because Facebook does not moderate its private communications channels in the same way as its public platform.

It is also possible to moderate the page using multiple comment controls, allowing for large numbers of comments to be removed at the same time.

By reporting on the platform, you can explain why you are reporting something, and you can also report individual users' comments.

A safety guide for public figures is being developed and can be shared more widely and Ella explained that Facebook does share monthly information bulletins with MPs and Peers and has also conducted training on these processes with the political parties and at drop-in sessions with Parliamentarians.

### **3. Closing remarks-**

Chris thanked Facebook for their contribution and explained that there needs to be much greater awareness amongst politicians and that he would give some thought about how all of this could be communicated to wider groups of Parliamentarians.