APPG on Social Media appg

UK Safer Internet Centre www.saferinternet.org.uk

All- Party Parliamentary Group for Social Media Meeting

Industry Session as part of "Selfie Generation" Inquiry

Thursday 24 June 2021

Attendees:

Chris Elmore MP, Chair of APPG on Social Media
Aaron Bell MP, Secretary of APPG on Social Media
Baroness Karren Brady, Vice Chair APPG on Social Media
Lucy Cserna, Office of Sarah Champion MP
Frances Lasok, Office of Saqib Bhatti MP
Maf Haddow, Office of Simon Fell MP

Speakers:
David Miles, Facebook
Becky Foreman, Microsoft

Secretariat:

Michael Tunks, Internet Watch Foundation and UK Safer Internet Centre
Abigail Fedorovsky, Internet Watch Foundation and UK Safer Internet Centre
Emma Hardy, Internet Watch Foundation and UK Safer Internet Centre

Apologies:

Damian Hinds MP
Lisa Cameron MP
Bambos Charalambous MP
David Linden MP
Maria Miller MP
Mrs Pauline Latham MP
Robert Halfon MP
Kevin Hollinrake MP
Baroness Morgan

1. Opening remarks- Chris Elmore MP

Chris welcomed everyone to the session.

2. Opening remarks from Panel Members-

Becky Foreman- Microsoft:

Microsoft welcomes the opportunity to give evidence to the APPG for Social Media. Microsoft operates in a unique position in the global digital landscape operating across gaming, cloud, software, and search sectors.

Microsoft also sells its services through partners and do not rely on the monetisation of content on its services.

Microsoft has a long-standing commitment to protect children and many of the steps the company takes in developing its products and services go beyond what is legally required. Microsoft is also respectful of the importance of upholding both Human Rights and freedom of expression.

Eradicating CSE/A content on its platforms has long been a call to action and we know only too well the impact that the circulation of this content online has on its victims. Often this content can recirculate online for a whole generation.

Research from the Internet Watch Foundation into Self-Generated Indecent Images featuring youth has received funding support from Microsoft and that research has concluded that much of this content is harvested from its original upload locations. Microsoft works with the IWF to identify and remove these images and are seeing the number increase every year.

At Microsoft, we believe that we all have a role to play in addressing this issue – no one entity can safeguard children online. That includes the tech sector, public sector- parliament, government, law enforcement, teachers, schools, and parents. This is a societal issue, and it requires a whole societal response.

Dave Miles-Facebook:

Dave thanked the APPG for inviting him to contribute to this inquiry. Facebook's work on child safety has spanned over a decade and remains one of its most important responsibilities.

Facebook welcomed this opportunity to discuss its thinking and response to self-generated indecent images of children online. Indeed, Facebook has been at the forefront of efforts to tackle self-generated images in the UK, in collaboration with the IWF, as outlined in its submission to the inquiry in March.

Facebook takes a comprehensive approach to self-generated indecent images of children online within our broader response to child safety. It has a zero-tolerance policy against sexual exploitation and grooming on its platform and aggressively seek to prevent, detect, remove and report policy violations.

Facebook uses a combination of photo-matching technology to prevent the resharing of previously reported images, as well as artificial intelligence and machine learning to proactively detect child nudity and previously unknown and new child-exploitative content, as well as inappropriate interactions with children. Its leadership in terms of detection technology ensures we remove 99% of violating content, even before our users see it.

Facebook's recent child safety announcement in February, following the work of its data scientists have done to contextualise Facebook's Cybertip reports to NCMEC, means it is also testing Safety Alerts to tackle malicious sharing by directing them to organisations like Lucy Faithful in the UK and Troubled Desire in Germany, where they can get advice and guidance. It is also deploying pop-ups to tackle non-malicious sharing, to warn that it's against our policies and that there are legal consequences. Early indications are that this is particularly effective. Coupled with the launch in June of its "Report it. Don't Share it." video not just in the UK with the IWF, but India, Latin America and Africa, it's part of its ongoing commitment to reduce the non-malicious sharing of this content. It's important to understand that every share revictimizes the victim.

After consultation with child safety experts and organisations, Facebook have made it easier to report content for violating our child exploitation policies. To do this, it added the option to choose "involves a child " under our "nudity and sexual activity" category of reporting in more places on Facebook and Instagram. These reports are prioritized for review.

Facebook has more than 35,000 people working on security and safety at Facebook, with backgrounds in law enforcement, online safety, analytics, and forensic investigations, reviewing potentially violating content and reporting findings to the National Center for Missing and Exploited Children (NCMEC).

Cognizant that older teens may engage in non-exploitative, consensual, developmentally inappropriate behaviour, in addition to the above, it also seeks to serve young people with in-product educational messaging explaining the risk that they may be taking and the legality of the content they are engaged in.

Facebook collaborates with the IWF to help tackle child sexual abuse imagery online. Additionally, Facebook collaborated with the IWF to support the technical development and piloting of Report Remove. Report Remove, a partnership project between the IWF and NSPCC, supports a young person in reporting their sexual images or videos to enable the imagery to be removed.

Because online child exploitation is an internet problem, it demands an internet solution, and Facebook collaborates across industry through organizations like the Technology Coalition and holds leadership positions in a range of international multi-stakeholder organizations like the WePROTECT Global Alliance. Indeed, David is a member of the WePROTECT Global Alliance Threat Analysis Expert Steering Group, which will publish its findings later this year. Facebook is one of the lead signatories of the Voluntary Principles to tackle CSAM, signed in March 2020, along with the UK and a number of other governments. They lay down important principles for companies both large and small, who want to better tackle CSAM head on.

David expressed his gratitude to the Inquiry for this opportunity to share Facebook's work in tackling CSAM and self-generated images in particular.

3. Question and Answer session:

Chris Elmore: Thanked both panellists for their opening remarks and asked both Becky and David for their reflections on the recently published draft online safety bill.

Becky Foreman: We are strongly supportive of the Government tackling illegal and online harms. Microsoft is very supportive of the Duty of Care, which we believe is pragmatic and means that we can continue to make progress in innovating as a company, however, there is a need for clarity around the limitations to liability in this area.

Microsoft recognises that voluntary industry efforts have not always been enough and therefore as a company we welcome Government regulation. The Bill must also respect freedom of information and the right to privacy and freedom of expression.

It is very difficult to comment on the draft bill at present, because so much of the detail will be enacted through secondary legislation and decisions which could be taken by the Secretary of State or by Ofcom. It is quite challenging to know how services will be impacted.

It would be helpful if the Government or Ofcom could bring forward some of this detail in the coming months as there are at least 8 or 9 areas of the Bill that require further clarity.

We want to make sure we are able to be compliant with the legislation and for the larger companies we have the resources to be able to consult with lawyers, engineers, and others to be able to do this, but we do fear for smaller companies who do not have the same level of resources to prepare.

David Miles: I echo many of Becky's sentiments about the Bill, but also want to say that we broadly welcome the Government's ambitious Bill. The UK is one of 11 countries in the EMEA region looking to introduce regulation in this space and it is certainly one of the most forward-looking pieces of legislation.

We think the UK has taken a systematic approach and welcome the dialogue that new regulation brings. We are pleased that Ofcom will be appointed as the regulator. Ofcom is well-respected and we already work with them on topics like media literacy.

Tackling CSE/A is a global problem and must be looked at and considered through that lens. There is equally challenging legislation currently being developed in the EU and we must ensure that the UK's legislation also complements their approach.

Chris Elmore: We heard evidence in our previous session from Law Enforcement about how helpful Facebook had been in bringing to justice the case of David Wilson. Law Enforcement are incredibly concerned about the impact encryption of Facebook Messenger. David, would you be able to explain some of Facebook's thinking behind this policy and how you can continue the productive relationship the helped bring Wilson to justice?

David Miles: We have a policy at Facebook of not commenting on individual cases, but on the subject of encryption Facebook is committed to all-user safety and enabling people to communicate privately and safely from hackers. We are also deeply committed to being good partners to law enforcement and providing actionable referrals to law enforcement to investigate and are also responsive to their requests for available information. We have a good relationship with the National Crime Agency.

Ultimately, we want to ensure that we prevent this imagery from being exchanged in the first place via Facebook's platforms. Our CEO, Mark Zuckerberg was one of the first people to speak about this delicate balance between privacy and safety. This is a difficult and complex balance to get right.

We understand the concerns of Parliamentarians and policy makers, but it is also important to point out that 94% of the world's reports for CSAM come from Facebook, which raises questions for other companies about the detection of this content on their platforms.

We must shift the approach to prevention as this is a question of societal norms. The reporting side needs dramatic improvement. Facebook can see what many of the drivers are towards young people sharing "self-generated" imagery – they think it is very normal to share highly sexualised content.

I think we need to reframe the conversation around reports too. Through analysis from our data scientists we provided in our written submission in March we told the Inquiry that only 10% of the images and videos we acted on were unique images. We need to look at the number of reports we are placing on law enforcement who often do not have the capacity to deal with all these reports and a lot of that volume is not necessarily helpful to them.

There is a need for a more segmented, intent-based approach to the issue (for instance for minor-to-minor sharing versus adult-to-minor sharing) and we need further research into the effectiveness of law enforcement and hotline actions in this area.

Our long-term intention remains to encrypt the messenger function.

Aaron Bell: What information would Facebook make available to law enforcement following the encryption of messenger? Sorry to get technical - does that include meta data etc. but I presume that would not include the actual image or video?

David Miles: Facebook currently detects CSE/A content using both photo matching technology and classifiers. Classifiers have come a long way in the past few years and the company deploys these in both public and private areas of its platform.

We remove 300,000 images from the public spaces of WhatsApp every month and we have been able to gain a lot of insight into the normative behaviour of offenders. For example, in Messenger you can only forward messages to five people or groups at a time, which reduces virality.

We are working closely with the European Union Internet Forum to look at what the potential technical encryption solutions to deal with CSE/A content in encrypted channels. There is some loss of visibility, but there is also a huge amount Facebook can do to mitigate this.

There is also further learning to come from Europe based on the recent public debate around the e-Privacy Directive and the tensions between privacy and protection of children. This text has helped to open that debate, but we are currently awaiting the final approval of this text before we can decide whether there is sufficient legal basis for scanning to be resumed by Facebook in Europe for CSE/A content. (It remains on by default in the UK).

Karren Brady: Can I ask about the levels of investment you both have currently and what could the potential of investment in this area go to?

David Miles: Facebook's safety and security budget in 2019 was greater than the whole revenue at the time of our IPO.

We have invested millions of dollars into the National Centre for Missing and Exploited Children (NCMEC) in the US to develop its reporting system and have made significant investment in content moderators within our company.

We also share technology with smaller companies by providing open-source video and photo hashing technology through GitHub platform.

We want to understand and improve reporting which is why we have invested in research into new threats and trends that come along to then help shape interventions. The work of the IWF and NSPCC in the development of Report/Remove is world leading and a good example of how NGOs are helping to improve that process.

Becky Foreman: Microsoft has heavily invested in tools, research, and data into dealing with the issue of CSE/A online.

We launched PhotoDNA ten years ago after a member of the Toronto Police wrote to Bill Gates asking him for support with how we deal with the spread of this imagery online. In 2009, we furthered this development through Dartmouth College. We use Photo DNA to find copies of an illegal image, even if that image has been digitally altered.

We have licensed this to over 130 organisations and recently made this available on a cloud basis too. Some have called this the biggest contribution to online safety in the past decade.

Karren Brady: Is it having an impact when we still see such large volumes of content circulating online?

Becky Foreman: This is a huge task; but we have to address the reasons why we have such a large number of persistent offenders seeking to circulate this content online or demanding it online. We need a coordinated approach from different stakeholders.

Chris Elmore: My final question is on Age Verification. What are your views on its effectiveness and how it could be further improved?

Becky Foreman: Microsoft has worked closely with other companies and the UK Government to develop a new tool to identify grooming online and we launched this new detection technique in 2020, following a hackathon in the US in 2018 which considered the legal, policy and technical challenges.

The tool looks at the conversational risks and gives each conversation a risk rating which helps to flag conversations to human moderators for review. The tool is currently licensed through the US non-profit Thorn.

One of the challenges in the development of the tool has been access to data to develop the machine learning as many of these conversations are essentially contraband.

Turning to Age-Verification, on X-Box, Microsoft requires parental controls to be implemented at set up. This requires an adult to set the age of the child and the services are then appropriately set to the age of that child. These can be tweaked as applicable by parents, but it requires them to be actively engaged in the safety settings and the family safety dashboards.

Dave Miles: Facebook and Instagram policies stipulate that the services should not be used by children younger than 13. WhatsApp's policy specifies that users must be 16 and over.

We are working closely with the ICO in terms of the implementation of the Age-Appropriate Design Code but there are challenges. Within the EMEA region for example, there are no formal Government identification mechanisms to assist with age verification.

We have privacy by default for teens, but it is also important to remember that our platforms are essentially aimed at adults and there is therefore a need for pre-teen specific platforms like our Messenger Kids app. We need to look at greater segmentation of products and not treat young people as just one big cohort, but design different products for different age groups

Recent data from Ofcom highlighted that 60% of 3- to 4-year-olds already have their own devices, so we need to be having these conversations much, much earlier. Children need to be able to build resilience in a digital age.

Baroness Kidron's work in this area has been extremely important as has the work of the Australian e-safety Commissioner who has only this week launched a safety by design framework.

Lucy Cserna: Research from the City of London University has highlighted that 16-17-year-olds are increasingly using VPNs- with 50% claiming to have used them. What are your views on how this could be addressed in the online safety bill?

Becky Foreman: It is difficult to tell what could be driving this rise, but it would seem to me to suggest that this should be addressed through the application of the GDPR.

We do not yet know the intricacies of how VPNs will work with the Bill, but with many pieces of legislation, companies implement them on a global basis.

Dave Miles: VPNs are in wide use amongst teens and seem to have grown considerably over the last few years. We need to address the issue of many of these VPNs being available through free downloads. I am not sure on what is driving this trend and perhaps this is an area where further research is required to then put some guidance in place. It is important to talk to young people too.

Chris Elmore thanked Dave and Becky for their evidence.

Chris introduced Emma Hardy from the Internet Watch Foundation who briefly presented the IWF's campaign to encourage parents to have positive conversations with their children about their online safety, an empowering campaign to target children in the 11-13 age range to block, report and talk about these issues and also presented a new reporting mechanism, Report/Remove which launched this week.

Emma highlighted that the IWF saw a 77% increase in self-generated abuse from 2019 to 2020 and have so far seen a 117% increase from 2020 to 2021. Most instances show 11-13-year-old girls.

4. Closing Remarks and next steps-

This session was the final session in the inquiry, and we will now begin to draw all the evidence we have heard together with a view to producing a final report for publication prior to the recess.