CarnegieUK TRUST

CHANGING MINDS • CHANGING LIVES

CARNEGIE UK TRUST RESPONSE TO THE APPG ON SOCIAL MEDIA INQUIRY: "SELFIE GENERATION": WHATS BEHIND THE RISE OF SELF-GENERATED INDECENT IMAGES OF CHILDREN ONLINE

January 2021

1. We welcome the APPG's inquiry on this topic and the opportunities we have had to date to engage with the Chair and the Secretariat on our work. Our response focuses on how the Government's proposals to regulate Online Harms can – if Ministers are willing – be designed in such a way so as to address the harm caused by the generation and spread of self-generated indecent content or material concerning children.

Background to our work

2. In early 2018, Professor Lorna Woods (Professor of Internet Law at the University of Essex and member of the Human Rights Centre there) and former civil servant William Perrin started work to develop a model to reduce online harms through a statutory duty of care, enforced by a regulator. The proposals were published in a series of blogs and publications for Carnegie and developed further in evidence to Parliamentary Committees[1]. In April 2019, the government's Online Harms White Paper[2] proposed a statutory duty of care enforced by a regulator in a variant of the Carnegie model and we welcome the fact that the final response to the White Paper consultation (published on 15th December[3]) continues to draw on our work, though we need to analyse the detail of that further. In December 2019, we published a draft bill[4] to implement a statutory duty of care regime, based upon our full policy document of the previous April[5]. Professor Woods also published a comprehensive paper on the statutory duty of care and fundamental freedoms, including freedom of expression.

3. The regulatory approach first set out by Carnegie in 2018 is to regulate social media companies' systems and processes for safety. This begs key questions for the regulator to ask the services that dominate the selfie market: what are the reasonably foreseeable risks of harm to people that arise from the operation of your services? What steps have you taken to mitigate these risks? Are these steps working? And in answering these questions companies should show evidence. At present, as the APPG might know, social media companies are opaque in respect of the above. Should the APPG take evidence from companies it might wish to ask these questions and follow through. The issues contained therein will underpin OFCOM's implementation of the government's proposals. A competent, responsible company would have no problem answering these points rapidly and in detail.

1 Our work can be found here: https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/

2 https://www.gov.uk/government/consultations/online-harms-white-paper

3 https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper

4 https://www.carnegieuktrust.org.uk/publications/draft-online-harm-bill/

5 https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2019/04/08091652/Online-harm-reduction-a-statutory-duty-of-care-and-regulator.pdf

The Online Safety Bill and self-generated indecent content relating to children

4. Our proposals have focused on developing a regulatory framework that takes account of the right of all users, but particularly vulnerable groups such as children, to be protected from a range of harms as well as to protect their health and wellbeing. We welcome the Government's publication of its full response to the Online Harms White Paper and support, broadly, its direction of travel and focus on a regulatory framework that is systemic and built upon the obligation for platforms to undertake a range of risk assessments of their services. While we have concerns about the boundary issues between the platforms' obligations in respect of criminal and "legal but harmful" content, which we will address in our forthcoming detailed analysis of their proposals, the priority given to protecting children within the regulatory framework provides significant opportunity for action on the concerns under investigation in the APPG's inquiry.

5. The focus of our work has been on the need for urgent legislative measures to ensure that social media and other online providers take action at a system level ("by design") to ensure that their services are designed in such a way to protect their users so far as possible from reasonably foreseeable harm, so allowing children to access the opportunities afforded by digital services concomitant with their individual competence and developmental stage. In this regard, we have very much supported the work led by Baroness Kidron to address the fundamental platform design issues that can leave children unprotected from a wide range of harms, for example through her influence on the development of the Age-Appropriate Design Code[6] and 5 Rights' recent "risky by design" framework.[7]

6. We welcome the fact that the Government's response includes harmful content affecting children (eg porn, violence) as a priority category for action and that companies in both category 1 and category 2 have a duty to "address illegal content and activity and protect children" (Para 28). We also welcome the explanation in broad terms, of how service design decisions can contribute to the risk of online harms (Box 5, page 28), particularly to children or vulnerable users and specifically flags as an example "higher risk" services that would allow children to be contacted by unknown adult users, or that allow all users to livestream themselves. This section also sets out how the regulators' risk assessment process will work in this regard. We also welcome that the response clearly sets out that companies will be required to assess the likelihood of children accessing their services and, if so, to provide additional protections for them (para 2.15). A focus on such design factors will be important in reducing the risk of significant harm to children from producing or sharing self-generated indecent imagery, whether that imagery meets the threshold for criminal content or not.

In summary: a statutory duty of care for social media

7. Our focus in developing a statutory of duty of care has been on systems, design and rights. Our proposal sets out how a statutory duty of care would require most companies that provide social media or online messaging services to protect people from reasonably foreseeable harms that might arise from use of those services. This approach is risk-based and outcomes-orientated. A regulator would ensure that companies delivered on their statutory duty of care and it would have sufficient powers to drive compliance.

6 https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/age-appropriate-design-a-code-of-practice-for-online-services/

7 https://www.riskyby.design/introduction

8. Social media service providers should each be seen as responsible for a public space they have created, much as property owners or operators are in the physical world. Everything that happens on a social media service is a result of corporate decisions: about the terms of service, the software deployed and the resources put into enforcing the terms of service.

9. In the physical world, it is accepted in many jurisdictions that those owning or operating buildings and spaces are responsible for them; in the UK, Parliament has long imposed statutory duties of care upon property owners or occupiers in respect of people using their places, as well as on employers in respect of their employees. Variants of statutory duties of care also exist in other sectors where harm can occur to users or the public. A statutory duty of care is simple, broadly based and largely future proof. It identifies the objective – harm reduction – and leaves the detail of the means to those best placed to come up with solutions in context: the companies who are subject to the duty of care.

10. Harms should be identified by the legislature. These should include harms to children arising from criminal activity directed towards them (eg grooming); accessing age-inappropriate content as well as emotional harm arising, eg, from cyber-bullying (especially where there is a racial, religious, gendered element). There are also wider harms that arise from the way in which social media platforms can create pressure on individuals relating to their image or popularity; for example, to portray a "perfect" image of oneself and upload selfies or other images that will boost one's self-esteem; the impact of chasing after likes and other metrics of being popular; the risk of body dysmorphia (generally from eg instagram but specifically arising out of the use of filters- and also plastic surgery filters); algorithms pushing pro-ana/NSSI content (and the creation of groups around these subjects which although potentially acting as a support might also act as reinforcement); and the personalisation of the pushing and re-pushing of content (eg that around suicide) once a user has expressed an interest.

11. The regime would cover reasonably foreseeable harm that occurs to people who are users of a service and reasonably foreseeable harm to people who are not users of a service. This is particularly important in relation to children, where risks to children (for example in the spread of self-generated images that may have been posted on other platforms) might take place on platforms where the individual that is being harmed is not themselves a user. The practice of "capping" – where child abusers and groomers take screenshots during explicit video calls and live streams with minors, and then circulate them widely amongst abuser networks, or use them to blackmail or coerce children – would fall into this category. Platforms have responsibility both where the content originated and also on the platforms or services where groups of sex offenders congregate.

12. Central to the duty of care is the idea of risk. If a service provider targets or is used by a vulnerable group of users (e.g. children), the risk of harm is greater and service provider should have more safeguard mechanisms in place than a service which is, for example, aimed at adults and has community rules agreed by the users themselves. In that instance, however, the platform operator should ensure that it has rigorous safeguards in place to ensure that its service is not used by children; we would see these forming part of the child risk assessment proposed by the Government. We argue throughout our work that content takedown or indeed age verification is not the only tool in the box – indeed, on its own it is inadequate for the scale of the problem.

13. The government's proposals emphasise the importance of "proportionate" regulation. We agree that it is important to be mindful of the size of the company concerned and the risk its activities present and that small or low-risk companies should not face an undue burden from the proposed regulation. However, some groups are sufficiently vulnerable (e.g. children) that any business aiming a service at them should take an appropriate level of care, no matter what its size or newness to market. Again, taking a "by design" approach means that baking in harm reduction to the design of services from the outset reduces uncertainty and minimises costs later in a company's growth. It also future proofs the legislation against new or emerging threats. We hope, in crafting its legislation, the government will take this into account.

How should Online Safety regulation work?

14. The Government has confirmed that it will appoint Ofcom as the regulator. In undertaking its role, Ofcom must be independent and is evidence-based in its decision-making and be given substantial freedom in its approach so as to remain relevant and flexible over time. Risk assessments and testing will be crucial to its approach: Ofcom needs to agree tests for harm (in this instance, the harm that is caused to children when they are coerced into sharing indecent images and when those images are spread without consent), run the tests, the company responsible for harm invests to reduce the tested level, test again to see if investment has worked and repeat if necessary. If the level of harm does not fall or if a company does not co-operate then the regulator will have sanctions.

15. In a model process, the regulator would work with civil society, users, victims and the companies to determine the tests and discuss both companies' harm reduction plans and their outcomes. The regulator would have the power to request information from regulated companies as well as having its own research function. The regulator is there to tackle systemic issues in companies and, in this proposal, individuals would not have a right of action to the regulator or the courts under the statutory duty of care.

16. The industry develops fast and urgent action is needed. We have concerns over potential delays to the introduction of the Online Safety Bill ("later this year" is the current Government commitment) and the lack of clarity as to whether pre-legislative scrutiny will be the first step. This creates a tension with a traditional deliberative process of forming legislation and then regulation. We are urging the UK government to find a route to act quickly and bring a duty of care to bear on the companies as fast as possible. Ofcom has recently signalled that it will only take on its powers "once the Bill has received Royal Assent"; if the Government delays further, then this could be as late as end-2022. In this situation, we would urge Parliamentarians to focus on pushing the Government to enable Ofcom to operate a form of "practice regulation".

Concerns with the government proposals

17. As described in the Government's Online Harms full response, we are not clear whether the duty of care refers to the totality of the obligations for the companies in scope or just to the regulatory responsibilities set out in the framework. This throws up significant boundary issues in relation to children/young adults: what would be the difference between the levels of protections that, for example, 19-year olds would have compared to children who are the priority for action in both category 1 and 2 companies? In addition, protecting individuals from "significant harm" is a high threshold when we consider children: how do we ensure that children aren't just protected at the same level as adults, where the harm may be assumed to occur more easily?

18. We are also concerned that there is too much emphasis in the Governments' framing on causality ("companies to take action to prevent user-generated content or activity on their services causing significant physical or psychological harm to individuals"), rather than risk detection and prevention.

19. The differentiation between category 1 and category 2 services also needs to be more than quantitative (eg the size of the platform) but to have clear thresholds for functionality and clarity over where services aimed at children would sit. For example, a small UK service aimed at children might quickly become high-risk if it is badly designed and badly run.

20. Finally, we would encourage Parliamentarians to press the Government, through all channels, to seek further detail on the proposals for the "child risk assessment": this will be crucial to determining the types of service design features and operations that will minimise the risk of harm to children and young people from the generation and spread of self-generated imagery. At a bare minimum, companies need to know whether children are accessing their services (or indeed, whether adults are accessing their services pretending that they are children) but it will be important that the regulator does not take the companies' risk assessments at face value. It will be important for assumptions that inform those risk assessments are tested independently so that the regulator can assure themselves of their robustness.

Carnegie UK Trust
January 2021
Contact: maeve.walsh@carnegieuk.org