CarnegieUK TRUST

CHANGING MINDS • CHANGING LIVES

Response to The Online Harms White Paper

June 2019

Background: Carnegie UK Trust

1. Carnegie UK Trust was established in 1913 by Scottish-American industrialist and philanthropist Andrew Carnegie to seek:

   Improvement of the well-being of the masses of the people of Great Britain and Ireland by such means as are embraced within the meaning of the word "charitable" and which the Trustees may from time to time select as best fitted from age to age for securing these purposes, remembering that new needs are constantly arising as the masses advance.

2. In 2018-2019, Professor Lorna Woods (Professor of Internet Law in the School of Law at the University of Essex) and William Perrin (a Carnegie UK Trustee and former UK government Civil Servant) developed a public policy proposal to improve the safety of some users of internet services in the United Kingdom through a statutory duty of care enforced by a regulator. Woods and Perrin's work under the aegis of Carnegie UK Trust took the form of many blog posts, presentations and seminars.

3. A full reference paper[1] drawing together their work on a statutory duty of care was published in April 2019, just prior to the publication of the Online Harms White Paper. It can be viewed, along with all the other material relating to this proposal, on the Carnegie UK Trust website: https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/

4. Our work has influenced the recommendations of a number of bodies, including: the House of Commons Science and Technology Committee, the Lords Communications Committee, the NSPCC, the Children's Commissioner, the UK Chief Medical Officers, the APPG on Social Media and Young People and the Labour Party.[2] Most recently, though it did not refer to our work, a report to the French Ministry of Digital Affairs referenced a "duty of care" as the proposed basis for social media regulation.[3]

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

2 https://www.nspcc.org.uk/globalassets/ documents/news/taming-the-wild-west-web-regulate-social-networks.pdf; https://www.childrenscommissioner.gov.uk/2019/02/06/childrens-commissioner-publishes-astatutory-duty-of-care-for-online-service-providers/; https://www.gov. uk/government/publications/uk-cmo-commentary-on-screen-time-and-social-media-map-ofreviews/; https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/82202.htm; https://labour.org.uk/press/tom-watson-speech-fixing-distorted-digital-market/; https://www.parliament.uk/business/committees/committees-a-z/lords-select/ communications-committee/inquiries/parliament-2017/the-internet-to-regulate-or-not-toregulate/; https://www.rsph.org.uk/our-work/policy/wellbeing/new-filters.html

3 http://www.iicom.org/images/iic/themes/news/Reports/French-social-media-framework---May-2019.pdf

5. We have already published a summary response to the Online Harms White Paper which draws attention to the main differences between the government's proposals and our work[4]. This formal response is in two parts: the first sets out our overarching response to the White Paper, and addresses some of the views expressed by other stakeholders since its publication; the second part responds to the specific questions posed by the White Paper consultation.

Part one: overall response

6. The Online Harms White Paper is a significant step in attempts to improve the online environment. Given the White Paper's breadth of scope, the complexity of the issues tackled as well as the challenging political environment within which it was drafted, we commend DCMS and Home Office officials and Ministers for its production and appreciate that all these factors have made it more of a "green" consultation document than might have otherwise been the case in different circumstances.

7. As the authors of detailed work on a proposal for a statutory duty of care, our principal concern relates to the meaning of the statutory duty of care put forward by the Government. We are particularly concerned that the White Paper puts forward an emphasis on detail in the codes of practice – perhaps to satisfy strong voices – without basing them on a sufficiently clear model of the duty of care.

Spelling out a systemic duty of care: the model of platform responsibility

8. The White Paper says in paragraph 3.1 that:

> *The government will establish a new statutory duty of care on relevant companies to take reasonable steps to keep their users safe and tackle illegal and harmful activity on their services … This statutory duty of care will require companies to take reasonable steps to keep users safe and prevent other persons coming to harm as a direct consequence of activity on their services.*[5]

To some extent, this may not appear that different from our proposal, borrowing from the language of the Health and Safety at Work Act 1974. What is less clear in the White Paper is the reason for the platform's responsibility in this context and consequently the sorts of steps that they might be required to take. The design choices made by the companies in constructing these platforms are not neutral; they have an impact on content and how it is shared. Every pixel a user sees on an online service is there as a result of decisions taken by the company that operates it: decisions about the terms of service, the software that operates the service and decisions about the resources put into enforcing the terms of service and maintaining the software. This can be best seen in the difference in content and user behaviour between services – they are different because they are designed and operated to be so. Companies have to own responsibility for reasonably foreseeable matters that arise from operation of their service.

9. By contrast, older regulatory models (e.g. the e-Commerce Directive) have not expressly recognised the role of company design and maintenance in contributing to the creation of the problem, but instead limit the role of the platforms to take-down and other ex post content creation mechanisms

4 https://www.carnegieuktrust.org.uk/blog/online-harms-response-cukt/

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

(e.g. moderation). It is here that the White Paper is not clear about the types of steps being required of companies. The White Paper, for instance, does not mention that companies should perform a thorough risk assessment of their operations – that is, an assessment to determine the risk of the harms that are specified under the duty of care arising – from which their actions to mitigate those identified risks should flow. Companies will not be unfamiliar with this process from, for example, data protection assessment requirements. Their risk assessment should be shared with the regulator who can critique it. From the risk assessment should flow a risk mitigation/reduction action plan, for the highest risk companies, this would be agreed with the regulator. An important part of a systemic approach is that it is to some extent forward looking. For instance, companies making risk assessments of the impact of changes to software on harms and acting on indicative evidence that has arisen from a framework such as the precautionary principle[6].

10. Based on the evidence given by the Secretary of State to the DCMS Select Committee on 8 May 2019,[7] and on subsequent reassurances from DCMS officials in meetings with ourselves and other stakeholders, we accept that the White Paper intended to describe a more systemic approach. We acknowledge that it is also possible to point to elements in the White Paper that reflect a recognition that companies contribute to the development of the problem and therefore need to take steps earlier on in the system design process. For example, the White Paper at paragraph 3.16 notes that the aim of increased transparency is to get companies to "take responsibility for the impacts of their platforms and products on their users" – though this could be referring to the negative impacts from the content of others, rather than the impact of platform design on communications choices users make. Perhaps more tellingly, the White Paper refers to safety by design, which is described (but not elaborated) in paragraph 8.1[8]. A number of the Codes refer to this principle, as well as other requirements that could mitigate against the creation of the problem in the first place. They are not, however, linked to a clear description of the responsibility model and so the point rather gets lost.

11. We feel it is valuable here to refer back to our proposal for a systemic statutory duty of care; the summary below is taken from our full reference paper:

At the heart of the new regime would be a 'duty of care' set out by Parliament in statute. This statutory duty of care would require most companies that provide social media or online messaging services used in the UK to protect people in the UK from reasonably foreseeable harms that might arise from use of those services. This approach is risk-based and outcomes-focused. A regulator would have sufficient powers to ensure that companies delivered on their statutory duty of care.

Social media and messaging 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 or messaging 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 and maintaining the software. These design choices are not neutral: they may encourage or discourage certain behaviours by the users of the service. In the physical world, Parliament has long imposed

6 United Kingdom Interdepartmental Liaison Group on Risk Assessment (UK-ILGRA): "The Precautionary Principle: Policy and Application", HSE website: http://www.hse.gov.uk/aboutus/meetings/committees/ilgra/pppa.htm We discuss the precautionary principle further in part two, question 5.

7 http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/the-work-of-the-department-for-digital-culture-media-and-sport/oral/101924.html

8 "Technology can play a crucial role in keeping users safe online. By designing safer and more secure online products and services, the tech sector can equip all companies and users with better tools to tackle online harms. We want the UK to be a world-leader in the development of online safety technology and to ensure companies of all sizes have access to, and adopt, innovative solutions to improve the safety of their users."

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. A statutory duty of care is simple, broadly based and largely future-proof. For instance, the duties of care in the Health and Safety at Work Act 19741 still work well today, enforced and with their application kept up to date by a competent regulator.

A statutory duty of care focuses on 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 d[u]ty of care. A statutory duty of care returns the cost of harms to those responsible for them, an application of the micro-economically efficient 'polluter pays' principle. The E-Commerce Directive, permits duties of care introduced by Member States9; the Audiovisual Media Services Directive (as amended in 2018) requires Member States to take some form of regulatory action in relation to a sub-set of social media platforms – video-sharing platforms.10

The continual evolution of online services, where software is updated almost continuously makes traditional evidence gathering such as long-term randomised control trials problematic. New services. adopted rapidly that potentially cause harm illustrate long standing tensions between science and public policy. For decades scientists and politicians have wrestled with commercial actions for which there is emergent evidence of harms: genetically modified foods, human fertilisation and embryology, mammalian cloning, nanotechnologies, mobile phone electromagnetic radiation, pesticides, bovine spongiform encephalopathy. In 2002, risk management specialists reached a balanced definition of the precautionary principle that allows economic development to proceed at risk in areas where there is emergent evidence of harms but scientific certainty is lacking within the time frame for decision making.

Emergent evidence of harm caused by online services poses many questions: whether bullying of children is widespread or whether such behaviour harms the victim; whether rape and death threats to women in public life has any real impact on them, or society; or whether the use of devices with screens in itself causes problems. The precautionary principle provides the basis for policymaking in this field, where evidence of harm may be evident, but not conclusive of causation. Companies should embrace the precautionary principle as it protects them from requirements to ban particular types of content or speakers by politicians who may over-react in the face of moral panic. Parliament should guide the regulator with a non-exclusive list of harms for it to focus upon. Parliament has created regulators before that have had few problems in arbitrating complex social issues; these harms should not be beyond the capacity of a competent and independent regulator. Some companies would welcome the guidance.

12. In our view, had the government described the statutory duty of care in those terms in the White Paper – which we believe was the intention – then many of the points we make below would no longer apply. Moreover, the sentence highlighted in bold is particularly apposite in relation to many of the criticisms of the White Paper from freedom of expression groups, who have interpreted the proposals as describing a regime primarily focused on notice and takedown of harmful content and which could lead to social media companies becoming arbiters of free speech and/or being overly cautious in removal of legitimate content in a bid to comply with the regulation.[11] **We recommend**

9 We note that the e-Commerce Directive is likely to be reopened by the new European Commission later this year.

10 We note the government's current consultation on the implementation of the AVMSD in the UK and will respond to that shortly.

11 For example: Open Rights Group: https://www.openrightsgroup.org/blog/2019/the-dcms-online-harms-strategy-must-design-in-fundamental-rights; Article 19: https://www.article19.org/blog/resources/uk-online-harms-proposals-will-impact-press-freedom/

that the DCMS Secretary of State moves swiftly, after the consultation has closed, to clarify the systemic nature of the proposed duty of care in a published statement or speech so as to address these concerns directly, rather than waiting for the publication of the full government response later in the year.

The Codes of Practice

13. The White Paper envisages the existence of codes of practice. Codes of practice should flow from the risk assessment process – both by the regulator and the companies themselves – of particular operations and systems, as we describe above, and the likelihood of harm identified under a precautionary principle framework. Whereas in this case the government has produced illustrative outlines of codes without a risk assessment process, it needs to be clearer that these illustrations are disposable so as not to implicitly fetter the regulator's discretion.

14. Codes can be a useful mechanism for dealing with technical issues and allow the system to keep up to date in a swiftly changing environment. The difficulties arise from the way the proposed codes are drafted and structured. The government has placed heavy textual emphasis on codes of practice to implement a duty of care. These are organised by reference to eleven different types of content, each with different specified actions that must be taken into account by the relevant operators. It may be that this approach was adopted as a vehicle to demonstrate to lobby groups that particular concerns would be met. In the codes that the government chose to elaborate, there is undue emphasis on notice and take down processes with the unfortunate consequence that the government appears to prioritise these over the safety by design features inherent in a systemic statutory duty of care. Even if compliance with these codes is not in itself enough to satisfy the statutory duty of care, it is clear that in the view of the government that the codes set out important steps. The Secretary of State spelled this out in his evidence to the DCMS Select Committee

    We are not saying to online companies, "You have a duty to comply with the codes of practice". We are saying, "You have a duty to comply with that duty of care. The regulator will hold you to account for whether you do so…." […] But it is the duty of care that is the significant care, and that has to be there because the guiding principle for me is that… [q. 367][12]

15. The focus in the draft codes is on different types of content; while it allows for differentiation in terms of the intensity of action required by the operators, it has the unfortunate side effect that platform operators will need to understand the boundaries between these different types of content in order to apply the appropriate code. In our view cross-cutting codes which focus on process (such as risk assessment and harm reduction) and the routes to likely harm would be more appropriate; and such an approach appears to be well within the bounds of what the Secretary of State envisages.

16. As noted, the actions that the White Paper envisages the operators taking tend towards take-down and moderation rather than on mechanisms that incentivise certain forms of content. Worryingly, there are references to proactive action in relation to a number of forms of content (and not just the very severe child sexual abuse and exploitation and terrorist content) which in the light of

12 http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/the-work-of-the-department-for-digital-culture-media-and-sport/oral/101924.html

the emphasis in the codes could be taken to mean a requirement for upload filtering and general monitoring to support that. While the White Paper acknowledges that Article 15 e-Commerce Directive prohibits general monitoring, a provision that can be seen as privacy protecting, it is weak in its explanation as to how it resolves the conflict between these two positions.

17. Some codes the government proposes to draft itself. We cannot support the prospect of a Home Secretary drafting and approving codes of practice in relation to speech, even the most extreme and harmful speech. This is not a role that should be fulfilled by the executive. At the most it could be the responsibility of Parliament or, more likely, an independent regulator to do this, after consultation with relevant bodies including, for example, the police and security services, the Crown Prosecution Service and perhaps also the Home Secretary. We are concerned that two interim draft codes are due to be published soon after the consultation has finished and that, in advance of a legislative framework or the appointment of an independent regulator, that companies will be expected to adopt these codes and comply with them. This, we feel, will unnecessarily pre-determine the overall approach of the statutory duty of care – delivered through narrow codes focused on specific content, rather than through companies' overall risk-assessment of the design and operation of their services – and tie the regulator's hands in terms of the wider application of the duty. This undermines the principle of the systemic approach highlighted by the Secretary of State in his Select Committee evidence.

Shadow Regulator

18. The level of detail in the draft codes, as well as the possibility of executive control in this area, also raise questions about the independence of the regulator. A better route would be for the government to appoint a (shadow) regulator to take this aspect of the work forward as soon as possible, working in consultation with relevant parties: civil society, users and companies involved in the sector. Such an approach would give the parties a sense of practical and emotional investment in a long-term work programme as well as supporting the independence of that process. The outcome would be likely to be more workable in practice too. In sum, we are concerned that the government's framing of its proposals in the White Paper has significantly reduced that space for public debate and consensus building – led by the appointed regulator and at arms' length from the legislature and the executive – and will instead delay the introduction of the enabling primary legislation as various groups fight over the second-order detail.

19. There is precedent for shadow regulation in controversial technology-driven areas. In regulation of human fertilisation and embryology, there was a seven-year gap between Dame Mary Warnock's 1984 report,[13] describing a system of licensing and regulation, and the establishment of the Human Fertilisation and Embryology Authority as a licence-issuing regulator in 1991. During that period practitioners, funders and professional bodies operated a voluntary shadow regulatory system along the lines of Warnock's recommendations, as the Commons Science and Technology Committee describes[14]:

13 Warnock Report https://drive.google.com/open?id=1S43xolk3VM3wyDdhTnfDocNHWIyMjctE

14 Commons Science and Technology Select Committee, 5th Report 2004-2005 session para 9-10 https://publications.parliament.uk/pa/cm200405/cmselect/cmsctech/7/704.htm

VOLUNTARY LICENSING AUTHORITY/INTERIM LICENSING AUTHORITY

In Mach 1985, the Medical Research Council (MRC) and Royal College of Obstetricians and Gynaecologists (RCOG), recognising that the introduction of a statutory body would take time, founded the Voluntary Licensing Authority for Human in vitro Fertilisation and Embryology (VLA) under the Chairmanship of Dame Mary Donaldson. The VLA consisted of people drawn from both the scientific and medical professions but was balanced by the inclusion of lay people. The VLA comprised members who carried out the licence inspections and issued licences to centres as appropriate and a secretariat. All potential centres had to make a written application to the VLA describing the particulars of the treatment services or research that they wished to undertake or were already providing.

Following a consultation, in 1987 the government published a White Paper, Human Fertilisation and Embryology: A Framework for Legislation, in which it committed itself to legislation.[10] In April 1989 the VLA decided to emphasise the temporary nature of its existence by changing its name to the Interim Licensing Authority for Human in vitro Fertilisation and Embryology.

20. There are of course considerable differences between the medical and technology sectors, particularly in professional self-governance. But the operation of a shadow scheme could provide valuable intelligence to the future regulator and legislators.

The Regulator

21. The government has not given sufficient reasons to justify consulting over whether the regulator should be other than OFCOM. If the government is serious about the urgency of tackling harms, it should allocate the role and resources to support to an existing regulator immediately. The Commons Science and Technology Committee recommended OFCOM be in place and with powers by October 2019[15]. While this suggestion may have been ambitious in terms of timescale, OFCOM has a track record of effective engagement with some of the world's biggest media groups. A new body, with no track record, would take years to earn sufficient reputation to be taken seriously.

22. We set out more detail on this subject in part 2 of this response (questions 10 and 11). But it is important to reiterate here that the independence of the regulator – not only from government but also from business – is vital for it to make decisions based on objective evidence (and not under pressure from other interests) and be viewed as a credible regulator by the public. This is particularly important given the fundamental human rights that are in issue in the context of social media. Independence means that the regulator must have sufficient resources, as well as relevant expertise.

23. Given the political uncertainty that lies ahead, a "shadow" process led by a nominated independent regulator that informs legislation and engages the parties in solving problems outside of government is vital to the regime working – and working quickly, as a shadow process would start to tackle problems now. The failure to name the preferred regulator, or to be clear on the timescale in which the legislation will be brought forward and the identified regulator set to work is a significant weakness.

15 Commons Science and Technology Committee, "Impact of Social Media and Screen Use on Young People's Health": https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/82202.htm

Scope of Coverage

Services

24. The White Paper would have been more "white" and less "green" if it had been clearer as to which services are in or out. Although paragraph 4.1 contains a broad definition of the services in scope (backed up by a non-exhaustive list in paragraph 4.3), there are still areas of concern and confusion. We note that the very breadth of the proposed regime may give rise to issues in understanding how the duty of care applies in each case (a problem compounded, as noted, by the content-focussed approach to the draft codes).

25. We welcome the explicit inclusion of messaging, given that some private messaging groups may run into the hundreds if not thousands of users. The ability to navigate some messaging services by non-person factors such as hashtags, geography or group title points to the wider use of these services than small group messaging. Some private anonymous messaging services are navigable by third party username finders. We acknowledge the government's willingness to engage on the difficult issue of where the boundary with private communications lies. It is important to remember however that communications are protected by constitutional and international law-based privacy guarantees and any state intrusion into that space must be limited, clearly justified and subject to safeguards. We set out more detail in part two (question 7).

26. There are a number of areas that we explicitly raised in our work that are not drawn out in the document. As we understand it, the search industry (i.e. Google) is in and messaging and user-generated content in games is included. Whilst there are some issues with harm in these areas, we feel that the government should take an overall risk-managed approach and create a practical pipeline of work for the new regulator that enables a focus on the bigger risks of harm in new social media in the first instance. We also reiterate that the existing media, including the below-the-line comments, should not be dealt with by this mechanism.

Harms

27. The scope of harm covered is not clear. We appreciate that a non-exhaustive approach may be necessary in a changing field. There are, however, some internal tensions regarding the nature of the harms that are definitely in scope (e.g., the White Paper is not clear on the nature of public or societal harms, for example, caused by disinformation and misinformation[16]) and we feel there are two notable omissions: threats to democracy and consumer harms, which we explore further below.

i) Harms to democracy

28. We have co-signed a statement relating to harms to democracy with a number of other organisations[17] and summarise our position here. Along with terrorism, extremism, and the exploitation and abuse of children, the White Paper highlights several threats which can be grouped

16 The impact of the spread of anti-vaccination campaigns and misinformation via social media and the rise in unvaccinated children and outbreaks of disease is one such example: see RSPH "Moving the Needle": https://www.rsph.org.uk/uploads/assets/uploaded/3b82db00-a7ef-494c-85451e78ce18a779.pdf

17 Statement co-signed by Carnegie UK Trust with Demos, doteveryone, Fawcett Society, Glitch, Institute for Strategic Dialogue and Jo Cox Foundation: https://www.carnegieuktrust.org.uk/publications/online-harms-white-paper-the-duty-of-care-in-our-democracy/

together as threats to democracy in the UK: viral disinformation, manipulation of the information environment, and online abuse. These are daily attacks on individuals' fundamental rights to freedom of expression, privacy, and association, which are fundamental to our democracy. They inhibit democratic engagement, corrode civil responsibility, corrupt political discussion and put at risk those who take part in public debate. In the context of the duty of care, it is crucial to note that these harms to our democracy are all exacerbated by system design decisions that the companies make.

29. Technology companies – like all others, and like all citizens – have rights and responsibilities in our democracy. If applied appropriately and embracing the UN Guiding Principles on Business and Human Rights[18], the systemic, overarching approach of the duty of care could help to ensure that UK democracy is strengthened by the participation of a vibrant tech sector.[19]

30. Online threats to democracy include:

- **Disinformation**: as noted in the White Paper, the prevalence and impact of disinformation is mediated by the tech platforms themselves. Choices aimed at maximising user engagement and/or revenue may have unfortunate side effects in terms of content prioritised. This threatens to distort electoral outcomes, remove transparency from political debate and undermine the public's faith in rational and accountable political decision-making.

- **Manipulation of the information environment**: as the White Paper rightly sets out, "[a] combination of personal data collection, AI based algorithms and false or misleading information could be used to manipulate the public with unprecedented effectiveness." Over time, the progressive subdivision of the public into ever more precisely defined target audiences traps people in "filter bubbles" to whom the platforms' algorithms then feed a steady diet of similar, or progressively more polarising or extreme, content that reaffirm and entrench pre-existing beliefs. To hold the attention of these groups (so they can be shown more ads and share more content), platform company algorithms help to generate a climate of outrage and sensationalism, normalising what were once extreme views. (We set out further thoughts on algorithms in part 2 (question 1).

- **Abuse and intimidation of public figures, especially women**. As the Prime Minister said last year[20], this constitutes a threat to the healthy public debate that is essential for our democracy. Such abuse has a direct impact on individuals' rights to expression and participation in democratic processes, and has what Amnesty UK have identified as a "silencing effect"[21], particularly on marginalised groups like women and girls, deterring others from participating in public debate or influencing them to self-censor themselves online to reduce the likelihood of abuse.

31. Damian Collins, the Chair of the DCMS Select Committee, set out the scale of the democratic threat to democracy from disinformation when he launched the Committee's report on the subject earlier this year:

18 https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

19 While the Electoral Commission will also have a role within any defined electoral period, all the harms to democracy identified in the White Paper play out on an ongoing basis, and so would be outside of the Electoral Commission's mandate.

20 https://www.gov.uk/government/speeches/pm-speech-on-standards-in-public-life-6-february-2018

21 https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-5/

Democracy is at risk from the malicious and relentless targeting of citizens with disinformation and personalised 'dark adverts' from unidentifiable sources, delivered through the major social media platforms we use every day. Much of this is directed from agencies working in foreign countries, including Russia. The big tech companies are failing in the duty of care they owe to their users to act against harmful content, and to respect their data privacy rights. Companies like Facebook exercise massive market power which enables them to make money by bullying the smaller technology companies and developers who rely on this platform to reach their customers. These are issues that the major tech companies are well aware of, yet continually fail to address.[22]

32. While the government's response to the DCMS Select Committee's report makes reference to some of the White Paper's actions to address some of the activities that contribute to the spread of disinformation (for example, political advertising) it is too reliant on the use of the codes of practice to bring this into full effect. We believe that the White Paper's failure to specifically address the broader range of harms to democracy through a systemic duty of care is unlikely to make the necessary impact on the scale of harmful activity identified by the Committee.[23]

33. The Council of Ministers of the Council of Europe made a declaration on the manipulative capabilities of algorithmic processes in February 2019[24]:

Moreover, data-driven technologies and systems are designed to continuously achieve optimum solutions within the given parameters specified by their developers. When operating at scale, such optimisation processes inevitably prioritise certain values over others, thereby shaping the contexts and environments in which individuals, users and non-users alike, process information and make their decisions. This reconfiguration of environments may be beneficial for some individuals and groups while detrimental to others, which raises serious questions about the resulting distributional outcomes.... The Council of Ministers encourages member States to assume their responsibility to address this threat by [inter alia] considering the need for additional protective frameworks related to data that go beyond current notions of personal data protection and privacy and address the significant impacts of the targeted use of data on societies and on the exercise of human rights more broadly.... taking appropriate and proportionate measures to ensure that effective legal guarantees are in place against such forms of illegitimate interference.

Whilst not legally, binding such declarations are sometimes used as a guide to interpretation by the European Court of Human Rights.

34. A recent example of the silencing effect was the statement made by Professor Alice Roberts a professional science communicator[25]:

22 https://www.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/news/fake-news-report-published-17-19/

23 https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/2184/218402.htm

24 Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes (Adopted by the Committee of Ministers on 13 February 2019 Decl (13/02/2019)1 https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b

25 Alice Roberts Professor of Public Engagement in Science at University of Birmingham - Twitter – 19 June 2019 https://twitter.com/theAliceRoberts/status/1141460396223733761

I'm taking a holiday from Twitter for a while. I've argued for reason, compassion and empathy in discussions about sex and gender. That's opened me up to more hate, bile and even misogyny than I've experienced before. I'm sad and shocked. Humans can be so much better than this.

ii) Consumer harms

35. We also note the explicit exclusion of economic harms from the scope set out in the White Paper (para 2.4):

Harms to organisations, such as companies, as opposed to harms to individuals. This excludes harms relating to most aspects of competition law, most cases of intellectual property violation and the organisational response to many cases of fraudulent activity.

In our work we suggested that harms went beyond those that attract great media attention. Economic harms cover a range of activities where people seek to use online networks as a vector or medium to rip others off. The range of activity is very broad and at an exceptionally high level. The National Trading Standards eCrime team maintains a rolling list of the latest online scams[26]. Work by Anderson et al as long ago as 2012 estimated that: "cybercrime is now the typical volume property crime in the UK".[27]

36. The 2017 Public Accounts Committee report into online fraud concluded that: "Online fraud is now the most prevalent crime in England and Wales, impacting victims not only financially but also causing untold distress to those affected. The cost of the crime is estimated at £10 billion, with around 2 million cyber-related fraud incidents last year, however the true extent of the problem remains unknown. Only around 20% of fraud is actually reported to police."[28]

37. Economic harm also encompasses intellectual property crime. The UK government's IP crime and enforcement report 2017/18 reports social media as the second highest location for IP crime tackled by Trading Standards.[29]

38. Harms to organisations usually result in harms to people. A systemic approach such as the government intends should tackle harms to business before they are amplified to become harms to people or worse still harms to society. A systemic approach such as the government intends should tackle harms to business before they are amplified to become harms to people or worse still harms to wider society[30].

26 National Trading Standards eCrime Team Alerts (March 2019) http://www.tradingstandardsecrime.org.uk/alerts/

27 Measuring the Cost of Cybercrime – Anderson et al WEIS Conference Paper · January 2012 https://www.econinfosec.org/archive/weis2012/program.html

28 Public Accounts Committee 'The growing threat of online fraud' December 2017 https:// publications.parliament.uk/pa/cm201719/cmselect/cmpubacc/399/39903.htm#_idTextAnchor004

29 Trading standards successes IP crime and enforcement report 2017 to 2018 https://www.gov.uk/ government/publications/annual-ip-crime-and-enforcement-report-2017-to-2018

30 https://www.independent.co.uk/news/uk/crime/paedophiles-internet-sex-crimes-offences-nspcc-sajid-javid-apps-online-gaming-a8972901.html

The user-generated content services targeted by the White Paper have allowed an explosion of consumer harms including the sale of illicit medicines[31] and dangerous toys[32] as well as fraud[33] and phishing[34] by failing to take systemic action. A statutory duty of care would for the first time correct that, leading to the online space as a whole being a safe place in which businesses can operate and where consumers can have confidence that they will not be duped, scammed or misled. [35]

39. We would therefore recommend that the government frames the scope of the statutory duty of care at a higher-level: all online harms to individuals or society should be within scope, including those that cause consumer detriment. This will make the statutory duty of care simple, broad and future-proofed. It will then be for the regulator(s) to consider all harms that occur online and then to consult on scope, priorities and then act accordingly.

40. Moreover, many of the illegal online activities (and outwith the proposed scope) bring with them both individual harms and wider societal harms, such as gambling, addiction, criminality etc. Many of the individuals or gangs involved in the "clearly defined" harms (extremism, child sex abuse) identified in the White Paper will also be involved in other types of illegal activity online, such as fraud, romance scams, money laundering, sale of fake products, etc. More than two-fifths (43%) of older people – almost 5 million people aged 65 and over – believe they have been targeted by scammers.[36]

41. The impact of fraud goes well beyond the financial with it destroying the health and independence of victims. On top of the personal harm caused, this increases demand on under-pressure public services like the NHS and social care. People defrauded in their own homes are 2.5 times more likely either to die or go into residential care within a year. Vulnerable groups are at risk of being effectively "groomed"[37] by criminals who seek to cultivate a relationship in order to defraud them. Those who have been the victim of online economic harm may be less confident and willing to shop online (exposing them to paying higher prices for essential services) and in their engagement with other online activities defined within the consultation's scope.

42. The White Paper also avoids any mention of misogyny as a hate crime, despite the government having committed to Stella Creasy MP and others that it would begin a process of making misogyny a hate crime.[38]

43. In summary, we feel that the government's distinction between clearly defined and less clearly defined harms is not helpful. Even assuming the well-defined harms to be so, there will always be boundary cases needing to be assessed. In our view, many of the difficulties around harm can be

31 https://www.gov.uk/government/news/uk-seizes-more-than-2-million-of-fake-medicines-as-part-of-international-crackdown

32 https://www.btha.co.uk/wp-content/uploads/2019/06/Toy-Safety-Campaign_FV-19.06.2019.pdf

33 https://www.telegraph.co.uk/money/consumer-affairs/instagram-victims-lose-9000-get-rich-quick-investment-scams/

34 https://www.infosecurity-magazine.com/news/dramatic-increase-abuse-file-1/

35 Which? "Control, Alt or Delete?: the future of consumer data", July 2018 https:// www.which.co.uk/policy/digitisation/2659/control-alt-or-delete-the-future-of-consumer-datamain-report

36 https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/reports-and-briefings/safe-at-home/rb_mar18_applying_the_brakes.pdf

37 https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/reports-and-briefings/safe-at-home/rb_april15_only_the_tip_of_the_iceberg.pdf

38 Review announced following campaign by Stella Creasy MP. See: https://www.bbc.co.uk/news/ukpolitics-45423789

ameliorated by focusing on the means by which types of harm are likely to arise. Taking a real-world analogy, if an employer sees a floorboard sticking up that person would not think "will someone break their leg or just twist their ankle?" but would ensure the floorboard is fixed. The process is about the responsibility for companies to demonstrate that they are taking steps to identify the things that are likely to cause a risk of relevant harm, in which the precise nature of the harm does not need to be precisely identified or quantified.

Balancing the need for action and the need to get this right

44. This is a complex landscape. In many meetings during the consultation period, we have heard the challenge for the government around the competing demands of "the need for action" and the "need to get this right", with a particular challenge on the latter being the wider landscape in which this regulation will be drawn up and enacted. We do not think that this is overly problematic.

45. We note that government, in chapter one of the White Paper, places its proposals for a statutory duty of care to address its identified online harms in the context of work across government and elsewhere on "other online harms", including: privacy, hacking, data protection, unethical deployment of AI, persuasive design, competition in digital markets and online advertising. We also note that the policy responses to some of these are in the process of being consulted upon and implemented (such as the ICO's Age Appropriate Design Code[39], the code of practice for IoT security by design[40], and the implementation of the Audio Visual Media Services Directive in the UK), some are the subject of reviews (like the proposed review of digital advertising by the Competition and Markets Authority) and others – like the Competition Green Paper, the Consumer Markets White Paper, the response to the Furman Review and the National Data Strategy – will be published later in the year, around the time of the government response to the Online Harms White Paper.

46. We think that a duty of care approach is a simple and effective way to deal with the vast majority of online harms and that – referring again to our proposal – taking a systemic approach that requires companies to take responsibility for foreseeable harm caused by the design and operation of their services would provide consistency and clarity for services that, otherwise, risk being regulated in multiple, narrow ways by different regulatory bodies. The regulators who may be required to have a role in enforcing a broader duty of care have a strong track record in working together; a consistent, systemic approach will further consolidate this.[41]

47. As we set out in our own work, action to reduce harm on social media is urgently needed. We think that there is a relatively quick route to implementation in law. A short Bill before Parliament would create a duty of care, appoint, fund and give instructions to a regulator. We have reviewed the very short Acts that set up far more profound duties of care than regulating social media services – The Defective Premises Act 1972 is only seven sections and 28 clauses (very unusually this was a private members bill written by the Law Commission); the Occupiers Liability Act 1957 is slightly shorter. The central clauses of the Health and Safety at Work Act 1974 creating a duty of care and a duty

39 See our response to the ICO's consultation: https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2019/06/17125949/ICO-Consultation-Age-Appropriate-Design-Code.pdf

40 See our response to the DCMS consultation: https://www.carnegieuktrust.org.uk/publications/dcms-consultation-iot-security/

41 The relationship between the ICO and OFCOM will be critical here. For example, the ICO's recent report on adtech and real-time bidding exposed how the adtech business model is the driver of many poor design choices and exacerbates harms: https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf

to provide safe machines are brief. For social media services, a duty of care and key harms are simple to express in law, requiring less than ten clauses or less if the key harms are set out as sub clauses. A duty for safe design would require a couple of clauses. Some further clauses to amend the Communications Act 2003 would appoint OFCOM as the regulator and fund them for this new work.

48. We speculate that an overall length of six sections totalling thirty clauses might do it. This would be very small compared to the Communications Act 2003 of 411 Sections, thousands of clauses in the main body of the Act and 19 Schedules of further clauses. This makes for a short and simple bill in Parliament that could slot into the legislative timetable. There is broad political consensus in the House on this matter. So, if government did not bring legislation forward a Private Peers/ Members Bill could be considered.

Part two: responses to the consultation questions

Question 1: This government has committed to annual transparency reporting. Beyond the measures set out in this White Paper, should the government do more to build a culture of transparency, trust and accountability across industry and, if so, what?

49. We believe that the regulator should be in control of this and have set out in our work how transparency reporting would fit into the ongoing harm reduction cycle. In summary, the regulator, in conjunction with the companies in scope of the statutory duty of care and with input from academics, civil society and other experts, would be responsible for determining the format, type and level of detail of information as well as the frequency of reporting. A net output of the harm reduction cycle would be a transparency report – and we welcome the proposal in the White Paper that this should be produced on an annual basis – by each company to ensure that an accurate picture of harm reduction is available to the regulator and civic society organisations, as well as to government policymakers and Parliamentary decision-makers. This will enable all those with an interest in online harms to determine the effectiveness of the regulatory framework as well as to identify patterns and trends that amount to collective harm that may require a different focus from the companies or oversight from the regulator.

50. We also believe that the regulator should have powers for thematic investigations and also a power to obtain information as required from companies, particularly where they have been alerted by the media, academic researchers or civil society groups that there is enough evidence to indicate that either a collective or societal harm is likely to be taking place.

51. We see an important intersection here with user complaints and redress. If a regulator can determine from transparency reports that a company is receiving high levels of user complaints in a particular area, then it can push for action to address the source of these complaints via the harm reduction cycle; the complaints therefore don't just lead to individual resolution for the complainants but result in systemic change. Data transparency and analysis will be critical here to enable the regulator to identify where patterns of individual complaints are emerging and where these point to underlying systemic issues.

52. Coupled with this, there should be powers of inspection (such as those that are used by the Information Commissioner's Office to obtain access to companies for the purpose of investigations

and fulfilment of their enforcement powers). If companies fail to comply with any of these requirements, including transparency reporting, then they would be subject to penalties.

53. There are two specific areas where a statutory duty of care could deliver improved transparency and oversight on service design features that can contribute to the promotion of harmful content and the prevalence of harms, in particular to democracy and consumers: algorithmic auditing and ad transparency. We are grateful to Eric Kind, with whom we are working closely, for providing the thinking below.

Algorithmic auditing

54. There is a need to establish new systems for oversight or auditing of algorithmic decision-making and the types of problems they may contribute to such as the promotion of hate speech or polarising an already divisive political culture.

55. Existing efforts in this area are already underway, such as those between academia and the private sector to allow external researchers to analyse information amassed by companies to address societal issues. These efforts have been challenging to set up, have yet to prove themselves, and are limited in scope. But it is right that the online harm regulator will adopt a role to "encourage and oversee the fulfilment of companies' commitments to improve the ability of independent researchers to access their data".

56. The Online Harms White Paper also sets out that "[w]here necessary, to establish that companies are adequately fulfilling the duty of care, the regulator will have the power to request explanations about the way algorithms operate." (para 3.22) In order to give effect to the stated aims of the White Paper and to ensure true algorithmic accountability, we feel the regulator should have powers to undertake algorithmic audits themselves.

57. To do this, the regulator should be able to examine the purpose, constitution, and policies of the systems and to interview people who build and interact with different parts of that system and observe how people use the system. The regulator should be able to identify and assess what data was used to train the algorithm, how it was collected, and whether it is enriched with other data sources, and whether that data changed over time. It should be able to examine the model itself including considering the processing flow and the type of supervisory or monitoring mechanism used. The regulator should be able to undertake a code review or "white-box testing" to analyse the source code, or the statistical models in use, including how different inputs are weighted.

58. The regulator should also be empowered to run controlled experiments over time to determine if the algorithms subject to their review are producing unintended consequences that harm the public interest. Such experiments would be novel in this area, but such independent testing and experimentation are of course commonplace in other areas such as pharmaceuticals or food safety.

59. It is only by undertaking this sort of audit that the regulator, acting in the public interest, will be able to assess whether companies truly are acting responsibly, protecting the safety of their users and tackling harms on their platforms.

60. In order to achieve this, the staffing of the online harms regulator would need to include a number of technologists who are able to advise on policies and procedures on how such an audit should take place as well as undertaking the audit. The regulator could also bring in external experts as consultants or fellows to assist with particularly complex or novel issues, as other regulators have in the past. This work should draw on efforts already underway by the Information Commissioner's Office to create an AI auditing framework for data protection.[42]

61. This work should be done openly and transparently, with continuing consultation with civil society, industry and academia to jointly pool expertise to establish best practice in this new area. It is envisaged that processes could be co-constructed with external stakeholders and that this approach would be continuously developing, and responsive to the changing technological landscape.

62. There are a couple of different models that could be drawn from in terms of powers of audit. The ICO can undertake consensual audits to carry out an assessment of data controllers or processors are complying with good practice in the processing of personal data. Should the company not agree to a consensual audit, the ICO can (should they decide that there are reasonable grounds for suspecting a data controller or processor is failing to comply with the Data Protection Act) seek a warrant to enter, search, inspect, examine and operate any equipment in order to determine whether a company is complying with the act.

63. A similar power could be provided to the online harms regulator, to empower them with the consent of the company, to carry out an algorithmic audit, or if the company refuses consent, and there are reasonable grounds to suspect they are failing to comply with the duty of care, to seek a warrant to determine whether they are so failing.

64. Alternatively, a model could be drawn on that is used by the Investigatory Powers Commissioners Office (IPCO) who are responsible for keeping under review the use of investigatory powers by a number of public authorities including the security and intelligence agencies and law enforcement bodies. IPCO has powers to conduct investigations, inspections and audits as the Commissioner considers appropriate for the purpose of the Commissioner's functions including access to apparatus, systems or other facilities or services. In practice, this means IPCO are able to inspect on site, the entire system used by the body they are auditing, including the underlying data, any technologies processing the data, and the output provided.

65. It is, of course, important that commercial confidentiality is respected, as well as the data protection of the data being used by the algorithm. To avoid unnecessary risk, and to protect commercially confidential information and to ensure the protection of personal data, it is envisaged that the presumption would be that such audits would be undertaken entirely on the companies' systems within their premises or alternative secure location, and not that data, or commercially confidential information, would be removed offsite for analysis.

Ad transparency

66. Transparency around who is targeting citizens with political messaging is essential to a healthy democracy: it applies with broadcast and print advertising, and should apply online, too. Before the recent European Parliamentary elections, Facebook, Google, and other large online companies

42 https://ai-auditingframework.blogspot.com/2019/03/an-overview-of-auditing-framework-for_26.html

committed, among other things, to transparency around political advertising. However, as found by a group of organisations including Avaaz and Institute for Strategic Dialogue[43], these companies were unable to reliably differentiate between political and non-political online advertising. For instance, non-political adverts for companies like Ikea were tagged as political, and political adverts in support of the German far-right AfD were not tagged as political. Rather than leaving the companies to distinguish political from non-political ads, transparency should apply to all advertising.

67. How to operationalise this transparency is set out by Mozilla and a scores of experts in March 2019, in a practicable and detailed description of what an effective and transparent ad archive API should look like.[44] The description gives details of how such an API should:

- Show the content of the advertisement and information about targeting criteria;
- Have the functionality to empower, not limit, research and analysis;
- Allow up-to-date and historical data access;
- Be publicly accessible.

Question 2: Should designated bodies be able to bring 'super complaints' to the regulator in specific and clearly evidenced circumstances?

Question 2a: If your answer to question 2 is 'yes', in what circumstances should this happen?

68. We welcome the moves by the government to design a better system of redress for users. It is important, however, to differentiate between complaints against the regulator for the discharge of its responsibilities (which are often in the form of a "super complaint") and the handling of user complaints - either against the platform provider or as against other users - and related redress mechanisms.

Super complaints

69. The White Paper could have been clearer as to what a super complaint is, how it fits into the overall complaints and redress framework envisaged and what the distinction is between taking action to resolve or provide redress for individual harms and designing a mechanism to address collective or societal harms.

70. From our perspective, we interpret a super complaint as that by designated bodies in relation to systemic failures. The Enterprise Act 2002 (the Act) specifies that a super-complaint, as defined by section 11(1) of the Act, is a complaint submitted by a designated consumer body, such as Which? or Citizen Advice, that "any feature, or combination of features, of a market in the United Kingdom for goods or services is or appears to be significantly harming the interests of consumers". Super complaints have been made with regard to rail fares compensation and loyalty penalties, and can also be made regarding policing: s29A Police Reform Act specifies that complaints can be made by designated bodies in relation to "a feature, or combination of features, of policing in England and Wales by one or more than one police force is, or appears to be, significantly harming the interests of the public."

43 https://www.isdglobal.org/wp-content/uploads/2019/06/Joint_Submission-070619-letter.pdf

44 https://blog.mozilla.org/blog/2019/03/27/facebook-and-google-this-is-what-an-effective-ad-archive-api-looks-like/

71. There are similarities between the systems and it could be clearly applicable to the social media set-up as a backup for regulatory capture or a problem that the regulator is just not seeing. However, it is difficult to see which existing consumer or civil society body would be best suited to taking on this "super complaint" designation; we agree with doteveryone's view that civil society organisations in the digital space are largely nascent, small and financially precarious. We also note that the size of organisation is not necessarily a required criterion, as evidenced by the bodies designated to undertake super complaints in relation to the police. Indeed, some of these organisations might be appropriate for consideration as designated bodies under the Online Harms legislation[45].

72. Multiple organisations representing various perspectives or topics will also be required; for example, the NSPCC might lead as the designated body in relation to harms to children but other groups of people would also need representation so as not to be unserved. As well as more precisely defining the super complaint system, the government therefore needs to ensure that there are designated bodies that are resourced, capable and broadly based enough to be able to take on the role.

User complaints and redress

73. However, a "super complaint" process does not give compensation to users nor are the designated bodies the destination for the escalation and resolution of individual public complaints; on one reading, the White Paper suggests that this is what the super complaint might be for. Currently, actions by numbers of users (often in relation to sums that are individually relatively small) are dealt with through group litigation orders or the representative action procedure – but these depend on there being a right of action in the first place. The regulatory scheme, as we envisaged it in our earlier Carnegie UK Trust work, does not do that (users can rely on existing causes of action – and could presumably where relevant use group litigation order or representative action procedures).

Question 3: What, if any, other measures should the government consider for users who wish to raise concerns about specific pieces of harmful content or activity, and/or breaches of the duty of care?

74. We note the many complaints from individuals that social media services companies do not deal well with complaints – in this the companies may not be unusual. There are many issues with redress mechanisms inside regulatory processes in the UK more generally, but this does not mean that matters should be left there. Handling of complaints is an important part of harm reduction – a company's complaints process should function as an early warning of systemic problems (see also para 51). Some aspects of designing for a safe service will require companies to place more emphasis on their redress and complaint mechanisms. Where companies have made a business decision to encrypt and decentralise services, reducing their own ability to detect harm through sampling or software, then they will require a more robust and effective complaint service to investigate and address harms reported to them.[46]

45 See clause 3: https://www.legislation.gov.uk/uksi/2018/748/made

46 See Josh Constine, 'Whatsapp has an encrypted child porn problem', Tech Crunch, 20 December 2018, available: https://techcrunch.com/2018/12/20/whatsapp-pornography/

75. As noted, we also envisage that reporting on the operation of the complaints and redress mechanisms would form part of a provider's transparency and reporting obligations. This external review is important not just in terms of effectiveness of harm reduction but in ensuring that the companies are operating fairly, that the meaning of terms of service are clear and consistently applied, that the rights of some groups are not prioritised over those of others and that when take down is an issue, the decision to take content down or not can be objectively verified.[47] We have not had the capacity fully to consider ADR/ombudsman regimes within our work but it is an area that deserves detailed consideration and policy development.

Question 4: What role should Parliament play in scrutinising the work of the regulator, including the development of codes of practice?

76. We strongly argue throughout our work that the regulator must be independent, not just from government but also from industry so that it can make decisions based on objective evidence and not under pressure from other interests. We would extend that to Parliamentary pressure, which is often influenced by lobbyists from the businesses that are under scrutiny. As is common with other regulators, a regulator should be accountable to Parliament for the exercise of its duties and be required to include detail on the these in its annual report to Parliament. But the regulator must be independent and Parliamentary scrutiny must be limited to questions of its performance, and its delivery against its statutory objectives. Involvement in determining the substance of, e.g. the codes of practice, will be too close a relationship. As we set out in our opening remarks, we argue strongly that the independent regulator must remain in the lead on developing these, in consultation with other bodies but not accountable to them.

Question 5: Are proposals for the online platforms and services in scope of the regulatory framework a suitable basis for an effective and proportionate approach?

77. We have covered this question in the main body of our response and will reiterate here that the White Paper, with such a broad definition of services in scope, has raised a number of areas of concern and confusion, compounded by the overall focus on content removal rather than the explanation and application of a systemic duty of care.

78. We have seen White Paper responses during the consultation period warning that a duty of care would penalise start-ups and SMEs, lead to greater domination of the market by the large tech firms and stifle innovation in the UK, making it uncompetitive and undesirable for further investment[48]. We disagree. A level playing field will only be delivered by a baseline of regulation that requires all companies hosting user-generated content – no matter how big or small – to be responsible for the safety of users on their platforms. When we initially developed our duty of care proposals, we had considered a threshold based on the number of users a service had; we were persuaded by many civil society organisations, particularly those that represent child safety, that some of the most harmful behaviour can be carried out on the smallest, below-the-radar platforms and that, in the context of child sexual abuse and exploitation, children are often encouraged to leave the bigger platforms so that they can be targeted and groomed out of sight, elsewhere.

47 Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (A/HRC/38/35), 6 April 2018, paras 26-28

48 For example, from Coadec: https://twitter.com/coadec/status/1115160196832215040?lang=en

79. We therefore came to a view that there should be no de minimis user/customer threshold for the duty of care. 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. Beyond child protection, basic design and resourcing errors in a growth stage have caused substantial problems for larger services and much of the debate on AI ethics attempts to bake in ethical behaviour at the outset. The GDPR emphasis on privacy by design also sets basic design conditions for all services, regardless of size. We are struck that in other areas even the smallest businesses have to take steps to ensure basic safety levels – the smallest sandwich shops have to follow food hygiene rules almost all businesses have to follow health and safety measures for their workforce. In both these cases, risks are assessed in advance by the companies concerned within a framework with a regulator.

80. We do agree with the government that there should be a proportionate approach to the implementation of the regulation and that this is the best protection against the claims that applying a duty of care to all providers might discourage innovation and reinforce the dominance of existing market players. Good regulators take account of company size and regulation is applied proportionate to business size or capability. We would expect this to be a factor in determining what measures a company could reasonably have been expected to have taken in mitigating a harm. Clearly, what is reasonable for a large established company would be different for an SME. The 2014 statutory "Regulators Code" even requires some regulators to take a proportionate, risk managed approach to their work, the code says that: "Regulators should choose proportionate approaches to those they regulate, based on relevant factors including, for example, business size and capacity."[49]

81. The proportionality assessment proposed does not just take into account size, but also the nature and severity of the harm, as well as the likelihood of it arising (as discussed below in relation to risk-based regulation). For small start-ups, it would be reasonable for them to focus on obvious high risks, whereas more established companies with greater resources might be expected not only to do more in relation to those risks but to tackle a greater range of harms.

82. The regulator should determine, with industry and civil society, what is a reasonable way for an SME service provider to manage risk. Their deliberations might include the balance between managing foreseeable risk and fostering innovation (where we believe the former need not stymie the latter) and ensuring that new trends or emerging harms identified on one platform are taken account of by other companies in a timely fashion. The regulatory emphasis would be on what is a reasonable response to risk, taken at a general level. In this, formal risk assessments constitute part of the harm reduction cycle; the appropriateness of responses should be measured by the regulator against this.

83. We would envisage that a regulator would not be likely to apply severe sanctions in the case of a start-up, at least initially. A small company that refused to engage with the regulatory process or demonstrated cavalier behaviour leading to harms would become subject to more severe sanctions.

84. Regulatory action should be based upon evidence. This is a particular challenge in an area that moves as fast as online services. The precautionary principle provides a framework for potentially hazardous commercial activity to proceed relatively safely and acts as a bulwark against short term political attempts to ban things in the face of moral panic.

49 HMG, Regulators Code, April 2014, available: https://www.gov.uk/government/publications/ regulators-code

85. Rapidly propagating social media and messaging services, subject to waves of fashion amongst young people in particular, are an especial challenge for legislators and regulators. The harms are multiple, and may be context- or platform- specific, while the speed of their proliferation makes it difficult for policymakers to amass the usual standard of long-term objective evidence to support the case for regulatory interventions. The software that drives social media and messaging services is updated frequently, often more than once a day. Facebook for instance runs a "quasi-continuous [software] release cycle"[50] to its webservers. The vast majority of changes are invisible to most users. Tweaks to the software that companies use to decide which content to present to users may not be discernible. Features visible to users change regularly. External researchers cannot access sufficient information about the user experience on a service to perform long term research on service use and harm. Evidencing harm in this unstable and opaque environment is challenging, traditional long-term randomised control trials to observe the effect of aspects of the service on users or others are nearly impossible without deep co-operation from a service provider. In a continuous software release environment, there can be no control group of people against which to measure the impact of changes.

86. Regulation and economic activity must proceed in the face of indicative harm, but where scientific certainty cannot be achieved in the time frame available for decision making.

87. This is not the first time the government has been called to act robustly on possible threats to public health before scientific certainty has been reached. After the many public health and science controversies of the 1990s, the UK government's Interdepartmental Liaison Group on Risk Assessment (ILGRA) published a fully worked-up version of the precautionary principle for UK decision makers.[51]

    The precautionary principle should be applied when, on the basis of the best scientific advice available in the time-frame for decision-making: there is good reason to believe that harmful effects may occur to human, animal or plant health, or to the environment; and the level of scientific uncertainty about the consequences or likelihoods is such that risk cannot be assessed with sufficient confidence to inform decision-making.

88. The ILGRA document advises regulators on how to act when early evidence of harm to the public is apparent, but before unequivocal scientific advice has had time to emerge, with a particular focus on novel harms. ILGRA's work focuses on allowing economic activity that might be harmful to proceed "at risk", rather than a more simplistic, but often short-term politically attractive approach of prohibition. The ILGRA's work is still current and hosted by the Health and Safety Executive (HSE), underpinning risk-based regulation of the sort we propose.

89. A cynical actor in a regulatory setting might try to play an "evidence game" and the precautionary principle provides a tool with which to counter that. We should add that the precautionary principle is a process to bring indicative evidence into play and to act before scientific certainty is achieved. Once it has been invoked a normal risk management process continues based on the indicative risks. In our view, the precautionary principle does not imply that a more conservative approach is taken in risk management itself.

50 Continuous push software deployment – 'Rapid release at massive scale August 2017' https://code.fb.com/web/rapid-release-at-massive- scale/

51 United Kingdom Interdepartmental Liaison Group on Risk Assessment (UK-ILGRA): "The Precautionary Principle: Policy and Application": http://www.hse.gov.uk/aboutus/ meetings/committees/ilgra/pppa.htm

Question 6: In developing a definition for private communications, what criteria should be considered?

90. We strongly believe that private communications, such as one-to-one communications, should not be included in the duty of care. While we welcome the explicit inclusion of messaging, given that some private messaging groups may run into the hundreds if not thousands of users, and we acknowledge the government's willingness to engage on the difficult issue of where the boundary with private messaging lies, it is important to remember however that communications have traditionally formed part of constitutional and international law-based privacy guarantees and any state intrusion into that space must be limited, clearly justified and subject to safeguards (see for example: Copland[52]; Big Brother Watch[53] and Tele2/Watson)[54]. In our view, however, design choices about the way platforms operate are less likely to trigger privacy concerns than ex post monitoring of content.

Question 7: Which channels or forums that can be considered private should be in scope of the regulatory framework?

91. We think this is the wrong way to frame the question. Naming particular channels or forums in order to bring them into the duty of care ambit is not helpful, and is short-sighted as an exhaustive list is impossible, while new channels and forums will continually emerge as technology develops and as users migrate from channels that may be changing their terms of service in response to more stringent regulation. We think there should be a functional approach – and it should be how the channel/platform could be used, not how it is used in a particular context. This would cut down on scope of purely private, though we reiterate the need to be conscious of limitations imposed by the right to privacy.

92. For example, messaging services are not necessarily private and also give rise to risks to individuals. We encountered disturbing reports of harms arising in messaging services, for instance:

One teen chat app has featured in more than 1,100 child sexual abuse cases in the last five years, the BBC has found. Of 29 police forces that supplied information to the BBC, all but one had child exploitation cases involving Kik. [55]

Although some messaging service providers do carry out pro-active moderation, at least of unencrypted parts of their services, it is questionable if this is enough. Insofar as reasonably foreseeable harms arise, they should be risk managed by service providers.

93. As noted, we continue to take the view that private communication, for which the model in Article 8 ECHR is essentially one-to-one communication, lies outside our proposed regime. In the last year, it has become clearer that messaging services have gone beyond small groups supporting existing relationships – familial, friendship or work. We now observe a trend towards large groups and groups becoming findable to non-members who can join if there is room in the group. The size of these

52 Copland v. the United Kingdom, no. 62617/00, ECHR 2007-I.

53 Big Brother Watch v. the United Kingdom, nos. 58170/13 62322/14 24960/15, judgment 13 September 2018.

54 Joined Cases C-203/15 and C-698/15 Tele2/Watson, judgment 21 December 2016 ,ECLI:EU:C:2016:970.

55 Angus Crawford, 'Kik chat app 'involved in 1,100 child abuse cases', BBC News, 21 September 2018, available at https://www.bbc.co.uk/news/uk-45568276

groups suggests that the communication mediated via the service is neither private nor confidential. Other characteristics also indicate the non-private nature of the communication, notably the growing practice of public groups, sharing of group links and browsers and search apps for groups. Services that enable the creation of public groups and/or large groups would, in our view, become qualifying services under our proposal and fall under the statutory duty of care regime.

94. Reasonably foreseeable harms in a messaging service might be quite different to those in a public-facing social media service and may therefore require different responses. For instance, where the bulk of a service is not visible to the operator due to a business decision about encryption there should be a far more responsive and effective notice and remedy process for people in a group who have experienced harm. As we explain below, we propose a risk-managed harm reduction process which would lead to different measures to those for traditional social media.

Question 8: What further steps could be taken to ensure the regulator will act in a targeted and proportionate manner?

95. We set out in our April 2019 paper a full commentary on what can be learned from other regulatory models.[56] In adopting any such models, the government will need, at least until Brexit, to be aware of the constraints of the e-Commerce Directive and the AVMSD, as well as to be aware of the limitations on governmental action arising from human rights considerations, specifically (though not limited to) freedom of expression. Limitations on rights must meet certain requirements, notably that they be necessary and proportionate.

96. As we describe in our paper, British and European countries have adopted successful regulatory approaches across large swathes of economic and social activity, for example broadcasting, telecommunications, data, health and safety, medicine and employment. We judge that a regulatory regime for reducing harm on social media can draw from tried and tested techniques in the regulation of these fields, including effective regulatory tools and approaches such as risk assessments, enforcement notices etc. We also include consideration of the regulatory frameworks frequently cited in relation to social media services, namely the electronic communications and broadcasting sectors which represent the transmission/intermediary and content contexts respectively. We also consider data protection, as data processing is at the heart of social media services, as well as regimes which relate to the safeguarding of public or semi-public spaces. Harm emanating from a company's activities has, from a micro-economic external costs' perspective, similarity to pollution so environmental protection also provides some useful comparisons.

97. On the specifics of a statutory duty of care and the role of an independent regulator, we have set out our thinking in our April 2019 paper (chapter 9) where we also describe a detailed harm reduction cycle, and we include that material here for ease of reference.

98. Central to the duty of care is the idea of risk[57]. We argue above that all qualifying services should come under a statutory duty of care, regardless of size, but we are not proposing that a uniform

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

57 The HSE guidance on risk assessment demonstrates many approaches for companies. See: http:// www.hse.gov.uk/risk/ Also recent HMG guidance on risk and public bodies demonstrates an approach for board members in large organisations: https://www.gov.uk/government/publications/management-of-risk-in-government-framework

set of rules apply across very different services and user bases. Instead, the risks and appropriate responses to those risks should be assessed in the light of these differences. Harmful behaviours and risk have to be seen in the context of the platform. In assessing whether a statutory duty of care had been met, the regulator would examine whether a social media service operator has had particular regard to its audience (and in this there are similarities to the assessments made by OFCOM in relation to content regulation).

99. For example, a mass membership, general purpose service open to children and adults should manage risk by setting a very low tolerance for harmful behaviour, in the same way that some public spaces, such as say a family theme park, take into account that they should be a reasonably safe space for all. The risk profile would be different for a specialist site targeted at a more limited audience. Specialist audiences/user-bases of social media services may have online behavioural norms that on a family-friendly service could cause harm but, in the community in which they originate, are not harmful. Examples might include sports-team fan services or sexuality-based communities. This can be seen particularly well with Reddit: its user base with diverse interests self-organises into separate subreddits, each with its own behavioural culture and approach to moderation. Mastodon also has distinct communities each of which sets its own community rules (as opposed to Terms of Service imposed by the provider) within the overarching Mastodon frame.

100. Our proposed ongoing evidence-based process of harm reduction is another mechanism to ensure that the regulator's work is targeted and proportionate. The regulator would work with the industry and civil society to create a cycle that is transparent, proportionate, measurable and risk-based. As noted above, the regulator would prioritise high-risk services, and would only have minimal engagement with low-risk services. Our cycle relies on consultation and feedback loops with the regulator and civil society.

101. A harm reduction cycle begins with measurement of harms. Here we emphasise that as the companies' performance is to be managed at system level, we do not intend that the effect of social media and messaging use on each individual should be measured. Rather what is measured is the incidence of artefacts that – according to the code drawn up by the regulator – are deemed as likely to be harmful (to a particular audience) or if novel could reasonably have been be foreseen to cause harm. We use "artefact" as a catch all term for types of content, aspects of the system (e.g. the way the recommender algorithm works) and any other factors. The regulator would draw up a template for measuring harms, covering scope, quantity and impact. The regulator would then consult publicly on this template, specifically including the qualifying social media services. The qualifying social media services would then run a measurement of harm based on that template, making reasonable adjustments to adapt it to the circumstances of each service. The regulator would have powers in law to require the companies providing the qualifying services to comply. The companies would be required to publish the survey results in a timely manner. This would establish a first baseline of harm.

102. The companies would then be required to act to reduce these harms by setting out and implementing a harm reduction action plan. We expect those planned actions to be in two groups – things companies just do or stop doing, immediately; and actions that would take more time (for instance new code or terms and conditions changes). Companies should inform the regulator and publish their actions. Companies should seek views on their action plan from users, the victims of

harms, the NGOs that speak for users and victims etc. The companies' responses to public comment (though they need not adopt every such suggestion made) would form part of any assessment by the regulator of whether an operator was taking reasonable steps and satisfying its duty of care. Companies would be required to publish their action plans, in a format set out by the regulator; and our work suggests what those plans might cover. The regulator would take views on the plan from the public, industry, consumers/users and civil society and makes comments on the plan to the company, including comments as to whether the plan was sufficient and/or appropriate. The companies would then continue or begin their harm reduction work. Harms would be measured again after a sufficient time has passed for harm reduction measures to have taken effect, repeating the initial process. This establishes the first progress baseline.

103. If harms surveyed in the baseline have risen or stayed the same the companies concerned will be required to act and plan again, taking due account of the views of victims, NGOS and the regulator. In these instances, the regulator may take the view that the duty of care is not being satisfied and, ultimately, may take enforcement action (see below). If incidence of harms has fallen then companies will reinforce this positive downward trajectory in a new plan. Companies would prepare second harm reduction reports/ plans as in the previous round but including learning from the first wave of actions, successful and unsuccessful. Companies would then implement the plans. The regulator would set an interval before the next wave of evaluation and reporting. Well-run social media services would quickly settle down to a much lower level of harm and shift to less risky service designs. This cycle of harm measurement and reduction would continue to be repeated, as in any risk management process participants would have to maintain constant vigilance of their systems.

104. We anticipate that well-run services and responsible companies will want to comply with a harm reduction process. Where companies do not comply, or where the regulator has grounds to believe that they have not we propose that the regulator has information gathering powers as is normal in modern regulation (see for example the powers granted to the Information Commissioner). A net output of the harm reduction cycle would be a transparency report produced by each company to ensure an accurate picture of harm reduction was available to the regulator and civic society organisations.

Question 9: What, if any, advice or support could the regulator provide to businesses, particularly start-ups and SMEs, comply with the regulatory framework?

105. We would envisage that the independent regulator would, as in other sectors (notably HSWA, data protection and guidance on the Content Codes for broadcasting), give guidance on what is required by the regulatory regime and ways to achieve that standard. This saves businesses the cost of working out how to comply. In addition to guidance as to what risks are likely and immediate steps to mitigate those risks (provided in easier to understand language, perhaps even decision trees), another way to support companies would be encouraging the industry to develop of libraries of "good code" that provide appropriate solutions to some of the most common risks. Some large companies already share code on terror material. As in other sectors, regulation will create or bolster a market for training and professional development in aspects of compliance. We would expect the regulator to emphasise the need for training for start-ups and SME`s on responsibility for a company's actions, respect for others, risk management etc. The work on ethics in technology could usefully influence this type of training.

Question 10: Should an online harms regulator be: (i) a new public body, or (ii) an existing public body?

106. We strongly recommend that this is ii) an existing public body. A completely new regulator created by statute would take some years before it was operational. OFCOM, for instance, was first proposed in the Communications White Paper in December 2000, was created in a paving act of Parliament in 2002 but did not vest and become operational until December 29th 2003 at a cost of £120m (2018 prices). In our view harm reduction requires more urgent (and less expensive) action and for this reason we reject the idea, seen in some proposals, that a new sector specific regulator is required. Notwithstanding the long lead-in time to its establishment, a new body, with no track record, would take years to earn sufficient reputation to be taken seriously. A "shadow" process led by a nominated regulator that informs legislation and engages the parties in solving problems outside of government is vital to the regime working. That process would start to tackle problems now.

Question 10a: If your answer to question 10 is (ii), which body or bodies should it be?

107. As we set out in our overarching response, we strongly believe that the government should name OFCOM as the regulator immediately and the White Paper provides few reasons to justify consulting on whether it should be any other body. Our April 2019 paper (chapter 10) considers the merits of other potential regulators and puts forward the case for OFCOM to take on the role: it has over 15 years' experience of digital issues, including regulating harm and protecting young people in broadcasting, a strong research capability, proven independence, a consumer panel, and also resilience in dealing with multinational companies. OFCOM is of a size (£110-£120 annual income and 790 staff) where, with the correct funding it could support an additional organisational unit to take on this work without unbalancing the organisation. The Commons Science and Technology Select Committee supported this approach and recommended that OFCOM be given responsibility by October 2019.

108. In discussions with other stakeholders with an interest in economic harms, we have touched on the idea that responsibility for regulating a duty of care could be given to more than one regulator (also see para 46), in acknowledgement of the potential breadth of scope; for example, the Competition and Markets Authority might be more appropriate for some of the harms that are commercial or financial in nature, while the ICO might be given responsibility in relation to harms that are more closely aligned with its responsibility for data. Of course, any regulator responsible for this field – whether with sole or shared responsibility - would need to work with regulators with responsibilities in contiguous or overlapping areas; OFCOM has some existing experience in so doing, so our preferred approach would still be for it to be appointed to lead the initial process of establishing scope in consultation with industry, civic society and others.

Question 11: A new or existing regulator is intended to be cost neutral: on what basis should any funding contributions from industry be determined?

109. We note that the government intends the new regulator to be cost neutral but that set-up costs and ongoing running costs should be recouped via fees, charges or a levy on companies that are in scope. We would recommend a funding regime that is comparable with other similar regulators;

for example, both the ICO and Ofcom recoup costs in the form of fees from the companies they regulate coupled in some cases with grant-in-aid from DCMS to cover the running costs associated with fulfilling their statutory duties. The regulator should be given freedom to set their own fees, not have to negotiate the fee level with government. Fees for regulatory operation should be recouped from the largest players in the market, small companies should not have to pay. We note that the government has elsewhere proposed a digital services tax but we do not recommend that monies raised from this tax should be used to fund the costs associated with implementing a new regulatory regime, as this would be at odds with the framework under which other regulated businesses operated.

Question 12: Should the regulator be empowered to i) disrupt business activities, or ii) undertake ISP blocking, or iii) implement a regime for senior management liability? What, if any, further powers should be available to the regulator?

110. We have set out proposals for regulatory processes that would lead to imposition of sanctions in our detailed paper (chapter 9), all of which should be transparent and subject to a civil standard of proof. The regulator like any public body would be subject to judicial review. Sanctions would include:

- Administrative fines in line with the parameters established through the Data Protection Act/ GDPR regime of up to €20 million, or 4% annual global turnover – whichever is higher
- Enforcement notices – (as used in data protection, health and safety) – in extreme circumstances a notice to a company to stop it doing something. Breach of an enforcement service could lead to substantial fines.
- Enforceable undertakings where the companies agree to do something to reduce harm.
- Adverse publicity orders – the company is required to display a message on its screen most visible to all users detailing its offence. A study on the impact of reputational damage for financial services companies that commit offences in the UK found it to be nine times the impact of the fine.
- Forms of restorative justice – where victims sit down with company directors and tell their stories face to face.

111. We also consider criminal sanctions (either at corporate) to be appropriate for large companies with levels of resources where reputational or financial sanctions would be little deterrent. These sanctions would become applicable if it was found that a company had not carried out sufficient risk assessments nor put in place adequate processes to exercise its statutory duty of care: they would therefore enforce system-level compliance. Of course the decision to seek such sanctions would be that of the regulator and would be used where other measures had failed. In fact, this is normal regulatory practice. In extreme cases, there may be a case for considering director criminal responsibility. For instance, a network where that has played a role in repeated cases of child abuse and has not taken adequate steps to address the issue.

112. Our work has also explored whether, given the scale at which some of the qualifying social media services operate, there is the potential for exceptional harm. We used a couple of hypothetical examples to explore these circumstances: a) a social media service was exploited to provoke a riot in which people were severely injured or died and widespread economic damage was caused. The regulator had warned about harmful design features in the service, those flaws had gone

uncorrected, the instigators or the spreaders of insurrection exploited deliberately or accidentally those features; or, b) sexual harm occurs to hundreds of young people due to the repeated failure of a social media company to provide parental controls or age verification in a teen video service. In these cases, are fines enough or are more severe sanctions required, as seen elsewhere in regulation? In extreme cases should there be a power to send a social media services company director to prison or to turn off the service? We have noted that the regulation of health and safety in the UK allows the regulator in extreme circumstances which often involve a death or repeated, sustained breaches to seek a custodial sentence for a director. In the USA the new FOSTA-SESTA package apparently provides for criminal penalties (including we think arrest) for internet companies that facilitate sex trafficking. The introduction of FOSTA-SESTA led swiftly to closure of some dating services and a sex worker forum having its DNS service withdrawn in its entirety. We note, however, the potential interference with freedom of expression that the existence of such a possibility could have and therefore emphasise that if such sanctions were to be envisaged, their use must be limited to extreme cases and as a measure of last resort.

113. Further, short of sanctions against directors, could a service be turned off? The Digital Economy Act contains power (Section 23) for the age verification regulator to issue a notice to internet service providers to block a website in the UK. While this may appear to be a model for enforcement here, some caution is advisable. It is likely that sites falling within the DEA pornography provisions will mainly carry a similar sort of content; sites such as Twitter which may carry pornography lie outside the regime. Ancillary effects arising from a S23 blocking of a pornography site would therefore be limited compared with the situation in which a large and general platform were blocked – this could give rise to concerns about collateral censorship and therefore require really compelling grounds to justify it.

Question 13: Should the regulator have the power to require a company based outside the UK and EEA to appoint a nominated representative in the UK or EEA in certain circumstances?

114. Article 80 of the GDPR requires data controllers and processors to establish a representative. There is therefore precedent for this approach. It ensures that there is a point of contact, though that person need not be an employee of the data controller or processor.

115. We also note that the Indian Government is consulting on guidelines for intermediary liability for technology companies in the Information Technology Act (IT Act), 2000. In the draft guidelines the government is consulting on the requirement for basic establishment in India for companies with more than 5 million subscribers in India:

(7) The intermediary who has more than fifty lakh users in India or is in the list of intermediaries specifically notified by the government of India shall:

(i) be a company incorporated under the Companies Act, 1956 or the Companies Act, 2013;

(ii) have a permanent registered office in India with physical address; and

(iii) Appoint in India, a nodal person of contact and alternate senior designated functionary, for 24x7 coordination with law enforcement agencies and officers to ensure compliance to their orders/requisitions made in accordance with provisions of law or rules.

This seems a more stringent requirement than the GDPR: a 'permanent registered office' implies some assets and employers of the company.

116. If the requirement to nominate a representative for all non-UK based companies was felt to be disproportionate, another option might be to impose the requirement on some platform providers. The Indian example, for example looks at the company's size seen from the Indian market. A different analysis does not consider how many UK users there are (provided there are some) and instead considers the risk posed by the service. This could arise from being a widely used service or where the nature of the service is high risk (e.g. live streaming; aimed at children). There are wider questions for the United Kingdom in a post Brexit future as to whether it wishes to require representatives for a range of services.

Question 14: In addition to judicial review should there be a statutory mechanism for companies to appeal against a decision of the regulator, as exists in relation to Ofcom under sections 192-196 of the Communications Act 2003?

117. The issue of appeals will be a major focus in consultation for well-resourced companies with a track record of litigation. The UK economic regulatory system has substantial experience of designing, running and amending appeals systems that balance justice, timely regulatory action and gaming of the system by big companies. We refer back here to work some years ago to reform appeal regimes where the balance was not well made as this could be useful to DCMS. When OFCOM was created in 2003, its decisions could be appealed on their merits to the Competition Appeals Tribunal. In part, the government's intention had been to provide a check or balance to a powerful new regulator. The broad basis on which appeal was allowed led to OFCOM defending 10 appeals per year and frustrated OFCOM's ability as a regulator to take timely decisions.[58] Appeals against OFCOM made up '85-90%' of the workload of the Competition Appeals Tribunal, whose work was intended to cover appeals from all competition issues in the wider economy.

118. The government consultation document "Streamlining Regulatory and Competition Appeals: Consultation on Options for Reform"'[59] is a good overview of appeals processes in regulators, seeking the balance between justice, gaming the system and the need for prompt decision making. There is also a good discussion of the issue by former regulator Martin Stanley on his UK regulation site[60]. The government, satisfied that OFCOM was operating well, in the Digital Economy Act 2017 constrained the ability to appeal against OFCOM's decision on the merits, reducing the scope of appeal to process and judicial review. In the explanatory notes[61] to the Act the government said:

58 Ed Richards, OFCOM CEO Exit Interview – Lords Communications Committee Uncorrected Evidence, 18 November 2014 "..what we are trying to do here is balance efficiency and justice. It is important that justice is seen to be done so that you can have an appeal, and if we make a bad decision it is overturned….What you should not do is have a system that emphasises justice to the degree that you can no longer make timely decisions. This is really important. Parliament has asked us to act proactively to promote competition and to protect the consumer. When you can be tied up in the courts for years on end—and I mean years, not months—it is hard to do that in a fast-moving— [environment]… he solution we have laid out on many occasions, which is that you adopt a slightly different threshold so that appeal is still possible, but you do something that we refer to as enhanced JR, which permits all the procedural defences of judicial review, but the enhanced element also means that if we make a material error in the decision we can be overturned as well." https://www.parliament.uk/documents/lords-committees/communications/Ed_Richards_181114/ucCommsEv1181114OfcomRichards.pdf

59 Streamlining Regulatory and Competition Appeals Consultation on Options for Reform BIS June 2013 Ref: BIS/13/876 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/229758/bis-13-876-regulatory-and-competition-appeals-revised.pdf

60 Regulatory Appeals - this note examines the background to a 2013 government review of appeals against regulators' decisions. See https://www.regulation.org.uk/competition-appeals.html

61 Explanatory Notes: Policy Background Ofcom Appeals: paragraphs 45-49 http://www.legislation.gov.uk/ukpga/2017/30/notes/division/3/index.htm

The government believes that an "on the merits" standard of review is overly burdensome. A 2013 analysis by the Department for Business, Innovation & Skills found that the average length of an appeal to the CAT reviewed "on the merits" was 11 months. The high costs of continuing litigation and subsequent delays in the regulatory regime can hinder effective regulation that must be able to keep pace with technological advances in the sector. The CAT itself has stressed "an appeal before the Tribunal is not a de novo hearing" (BT v Ofcom [2015] CAT 6).

We would not suggest the ability to review decisions on the merits, given that OFCOM's decision making ability is well established and experience of operating an on the merits regime leads to an imbalance between quantity of appeals and the ability to deliver regulatory judgements fast enough to keep up with market development.

119. We have not carried out detailed consideration in our work of complaints and judicial review processes. Under this proposal there is a question of who would be responsible, because CAT is unlikely to have the right expertise; so this could be a new specialist tribunal or it could possibly come under the First Tier Tribunal (General Regulatory Chamber). Using this Tribunal would have the benefit of alignment with the mechanism for appeals under a parallel regulatory regime, as it currently deals with information rights appeals against the ICO. Clarity would be needed as to what sorts of complaints would go there; for example, user complaints, competition complaints, or complaints from companies about competitors and their own compliance.

Question 15: What are the greatest opportunities and barriers for (i) innovation and (ii) adoption of safety technologies by UK organisations, and what role should government play in addressing these?

120. Requiring a continuous process of risk assessment and mitigation for the roll out of software in online services should be simple, straightforward and well understood. This would be a substantial leap forward.

121. The government can play a role in championing the work of UK-based start-ups and other innovative companies that are designing and scaling services and products that have inbuilt safety-by-design features and that champion a "duty of care" to their users above, making trust and safety a business offering and providing a viable alternative to other more established services. Working with industry, the government can also use techniques that have proved successful in growing UK capability and innovation in other parts of the tech sector, such as providing UKRI investment programmes and accelerators to boost development of new products and services in high-priority or high-skilled "safety by design" areas; intensive start-up development support, network/information sharing and ecosystem building via Tech Nation; or establishing sandboxes to trial innovative new approaches to address emergent harms.

Question 16: What, if any, are the most significant areas in which organisations need practical guidance to build products that are safe by design?

122. Companies need greater education and awareness of the nature of harms online, how design choices and business models can exacerbate these and how safe-by-design principles translate into viable and profitable products and services. The ICO's Age Appropriate Design Code is an

example of how practical guidance can be combined with regulatory compliance advice to support companies in these design choices.

Question 17: Should the government be doing more to help people manage their own and their children's online safety and, if so, what?

123. We believe that education is an important aspect of online safety; however, there is no shortage of civic society organisations, consumer groups or initiatives funded by tech companies that provide information, tools and guidance to users of online services. Many commentators call for more media literacy training for children, and it is clear that parents and teachers need greater support to keep up with the risks that arise for children online. However, the speed and pace of change is overwhelming and for too long, the onus has been placed on users to keep themselves safe or (if they are unaware of, or do not follow, the guidelines on how to do so) to accept that any harm that arises is their responsibility. We do not think that media literacy, therefore, is a substitute for implementing a comprehensive statutory duty of care that bites at a systemic level and puts responsibility on companies to reduce reasonably foreseeable harms to their users regardless of their own actions or behaviours or levels of resilience.

124. We note that the government will develop a new online media literacy strategy and will start a "comprehensive mapping exercise" to determine its objectives. We are worried that this is referred to as being "ahead of the new regulator", indicating that the government does not envisage the implementation of any new regulatory regime for a number of years. Public information campaigns also have limited effectiveness and we do not wish to see the development of a "media literacy strategy" take precedence and divert attention from the design of an effective regulatory system. Many commentators call for more media literacy training for children[62] and we note the good work carried out by multi-stakeholder groups, such as the UK Council on Internet Safety, to identify the points of intervention across the education system to have an impact and improve digital literacy and resilience[63]. But this is not enough.

125. We think the need for training goes much further. Education is an important tool, not just in developing resilience in users, but also in introducing would-be software developers and service operators to some of the ethical and legal issues, not just add on tools to help users help themselves. Education could be a mechanism to disseminate that guidance and risk-tested code libraries should be developed and available.[64]

Question 18: What, if any, role should the regulator have in relation to education and awareness activity?

126. We believe the regulator can play an important role in identifying emerging harms and raising public awareness about them, and that it should have powers to commission research to improve the knowledge base and the effective actions both to reduce the risk of reasonably foreseeable harm occurring and to identify the measures that will help users understand risk and keep themselves safer online. Ofcom already has substantial capacity and expertise on media literacy research and

62 See, for example, LSE Media Policy Project blog: https://blogs.lse.ac.uk/mediapolicyproject/tag/ media-literacy/

63 https://www.gov.uk/government/publications/education-for-a-connected-world

64 The House of Lords Select Committee on Communications in its report Growing up with the Internet 2nd Report of Session 2016–17 (HL Paper 130), 21 March 2017 also emphasised the importance of guidance on design.

education[65] and, if it were to be given powers in relation to online harm reduction, could expand its capabilities to support this aim. We agree with the proposal that the regulator should also be given power to require companies to report on their education and awareness activity.

127. The regulator can also play a role in signposting users to education and support materials and resources and ensuring that users who feel they may have been harmed can understand their rights under the regulatory framework and the process they need to follow to raise a complaint and, if necessary, seek redress.

Carnegie UK Trust
June 2019

Contact: maeve@carnegieUK.org

65 https://www.ofcom.org.uk/research-and-data/media-literacy-research