CarnegieUK TRUST

CHANGING MINDS • CHANGING LIVES

Response to the DCMS Select Committee Interim Report

November 2018

Proposal for a Duty of Care for Harm Reduction on Social Media

Summary

1. This paper is submitted to the Digital, Culture, Media and Sport Select Committee in response to the recommendations in "Disinformation and 'fake news' – an interim report", published on 29 July 2018, and the invitation for "submissions to the Committee from readers of this interim Report, based on these recommendations, and on specific areas where the recommendations can incorporate work already undertaken by others" (para 5).

2. We congratulate the Committee on their thorough inquiry and the focus and determination with which they have addressed the myriad complex issues which have arisen throughout the course of their evidence sessions. In this submission, we do not set out to respond to the individual recommendations but to present to the Committee a proposal which, we believe, supports its observation that these "*complex, global issues [..] cannot easily be tackled by blunt, reactive and outmoded legislative instruments" and which has led the Committee instead to develop "principle-based recommendations which are sufficiently adaptive to deal with fast-moving technological developments*" (para 4).

3. This paper sets out a detailed regulatory proposal that would address many of the points the Committee raised in its interim report. The proposal is for a short bill to create a statutory duty of care for social media companies in respect of their users, enforced by an existing regulator such as OFCOM.

4. Earlier in the year we shared our work with Jack Walker in Damian Collins's office after your original call for evidence closed.

5. Our work for Carnegie UK Trust to design a regulatory system for "Harm Reduction in Social Media" starts from a similar position as the Committee's interim report: rapidly-propagating social media services, subject to waves of fashion amongst young people, are a particular 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. That is why we have looked to other regulatory regimes for a workable, principle-based approach to reduce the risk of harms to individuals: our proposal applies the "duty of care" principle found in health and safety regulation to social media platforms.

6. We set out the main elements of our proposal below, with links to further detail in the annex. While not limited to the challenges posed by digital advertising, our "duty of care" approach has most relevance to the Committee's further consideration of regulation in this area, where it believes the process: "*should establish clear legal liability for the tech companies to act against harmful and illegal content on their platforms. This should include both content that has been referred to them for takedown by their users, and other content that should have been easy for the tech companies to identify for themselves. In these cases, failure to act on behalf of the tech companies could leave them open to legal proceedings launched either by a public regulator, and/or by individuals or organisations who have suffered as a result of this content being freely disseminated on a social media platform*" (para 60).

7. We would welcome the opportunity to provide further evidence on our proposal to the Committee, either in written or oral form, before the publication of its final report.

Background

8. Lorna Woods (Professor of Internet Law, Essex University) and William Perrin (a former government advisor on regulation and Trustee of Carnegie UK Trust) have been working with Carnegie UK Trust (CUKT) to design a regulatory system to reduce harm on social media. The proposals have been published via a series of blogs[1] and in detailed evidence[2] submitted to the Lords Communications Committee Inquiry ("The Internet: to regulate or not to regulate?").

9. This paper sets out, in response to the DCMS Select Committee's call for further submissions to its inquiry, a high-level description of the Duty of Care proposal, including its aims and objectives, how it would work and next steps for development.

Social media harms: The problem

10. Ofcom's recent research, carried out jointly with the Information Commissioner's Office, has set out clearly the scale of the problem of harms on social media platforms, with 79% of adult internet users having concerns about going online. Specifically, among UK adult internet users:

- 69% report concerns about harmful content or conduct online, with around a quarter (26%) saying they have personally experienced some form of harm
- 58% are concerned about control over personal data and privacy, with over a quarter (28%) reporting experiencing harm; and
- 54% are concerned about hacking and security, with a quarter (25%) reporting having been harmed.[3]

11. The Government has made clear that it will legislate to protect safety online and its Internet Safety Strategy White Paper is due in early 2019. In recent months, the pressure on social media companies – in Europe and in the US – to act to address some of the most damaging harms on their platforms has intensified and, in the UK, the Opposition and many civic society organisations are now calling for a new Internet regulator, or indeed a "super-regulator", to bring together existing functions spread across multiple bodies.

1 https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/

2 http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/communications-committee/the-internet-to-regulate-or-not-to-regulate/written/82684.html

3 Ofcom/ICO Internet users' experience of harm online: summary of survey research, 18 September 2018

12. We do not debate the pros and cons of regulation in this paper but take it as a given – as does the Committee – that some form of regulation is urgently required. Instead, we focus on what we see as the most effective way to design and enact such regulation to reduce harms on social media: a risk-based approach that will deliver a simpler, quicker and future-proofed route to harm reduction online; and how that duty would work in practice.

The regulatory landscape and social media service providers

13. Social media service providers are not un-regulatable, and recent national and international policy has broadly taken a view that internet issues should be tackled wherever possible using 'physical world techniques'.

14. To inform this work, we have surveyed existing regulatory regimes for communications, the digital economy, health and safety and the environment.[4] There are many similarities in these regimes which, we believe, should be replicated in any regime for social media service providers: many ensure that changes in policy take place in a transparent manner and after consultation with a range of stakeholders; all have some form of oversight and enforcement – including criminal penalties; and the regulators are independent from both Parliament and industry. Breach of statutory duty may also lead to civil action. These matters of standards and of redress are not left purely to the industry.

15. While the telecommunications model may on the face of it seem an appropriate model for social media, there are some key considerations that argue against this:

- the regulator's power to stop the operator from providing the service itself feels inappropriate in the context of social media service providers, in the light of freedom of speech concerns;

- the regulator's specification of the conditions for compliance may be too 'top-down' for a fast-moving sector; allowing operators to make their own assessment of how to tackle risks means that solutions may more easily keep up with change, as well as be appropriate to the service; and

- the telecommunications regime is specific to the telecommunications context; the data and workplace regimes are designed to cover the risk entailed from broader swathes of general activity.

16. Conversely, we have drawn out some important principles from the telecoms as well as other regimes, which have informed our Duty of Care proposal.

- The data protection and health and safety regime highlight that there may be differing risks with two consequences:

    - that measures should be proportionate to those risks; and

    - that in areas of greater risk there may be greater oversight.

- The telecoms regime emphasises the importance of transparent complaints mechanisms – this is against the operator (and not just other users).

- The environmental regime introduces the ideas of prevention and prior mitigation, as well as the possibility for those under a duty to be liable for

4 https://www.carnegieuktrust.org.uk/blog/harm-reduction-social-media-can-learn-models-regulation/

the activities of others (eg in the case of fly-tipping by a contractor).

- The Digital Economy Act has mechanisms in relation to effective sanctions when the operator may lie outside the UK's jurisdiction.

The precautionary principle and duties of care

17. While social media companies are not unregulatable, they are unlike many other businesses in comparable sectors, such as telecoms or broadcasting. As Lawrence Lessig argued back in 1999, computer code sets the conditions on which the Internet (and all computers) is used. Code is the architecture of cyberspace and affects what people do online: code permits, facilitates and sometime prohibits. It is becoming increasingly apparent that it also nudges us towards certain behaviour – as the Committee itself acknowledges "*People's behaviour is being modified and changed as a result of social media companies*" (para 55) - with some of the most damaging consequences of this being considered in the Inquiry's focus on the proliferation of "disinformation" and "fake news". In this context, it is impossible for regulatory mechanisms to anticipate and limit the harmful consequences before they take place. That is why we think the answer to reducing online harms on social media needs to be built on two bases for action that have been developed in other areas of regulation: the precautionary principle and a duty of care.

18. Traditional, evidence-based policymaking requires that policy decisions should be informed by rigorously established objective evidence. Typically, action on an issue is only taken after consultation and the collection of such evidence. But, in innovative areas, there is often no long-term scientific research; or such evidence arrives too late to provide an effective measure against harms.

19. Rapidly-propagating social media services, subject to waves of fashion amongst young people, are a particular challenge for long-term objective evidence. In the face of such scientific uncertainty, the precautionary principle provides one basis for risk-based harm prevention. 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[5].

'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.'

20. 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. The ILGRA's work is still current and hosted by the Health and Safety Executive (HSE) and is one basis for our focus on the Duty of Care as an approach to harm reduction.

21. The other basis is the parallel between online spaces and the physical world: Parliament has long imposed statutory

5 http://www.hse.gov.uk/aboutus/meetings/committees/ilgra/index.htm

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, so as to make good shortcomings in the previous common law position[6]. While the company has freedom to adopt its own approach, the issue of what is 'reasonable' is subject to the oversight of a regulator, with recourse to the courts in case of dispute. If harm does happen, the victim may have rights of redress in addition to any enforcement action that a regulator may take action against the company.

22. We see social media platforms as no different to other public places, like an office, bar or theme park. Millions of people go to social networks owned by companies to do a vast range of different things. In our view, they should be protected from harm when they do so.

23. Duty of care is simple, broadly based and largely future-proof. It focusses on the objective and outcome (ie the prevention of harm) and leaves the detail of the means and processes to those best placed to come up with context-appropriate solutions, enforced in a risk-based manner by a regulator. Such an approach is also flexible, allowing operators to respond to changing technology and services where relevant. Duties of care set out in law 40 years ago or more still work well; for instance, the duty of care from employers to employees in the Health and Safety at Work Act 1974 still performs well[7], despite today's workplaces being profoundly different from 1974's.

24. In our view, the generality and simplicity of a Duty of Care works well for the breadth, complexity and rapid development of social media services, where writing detailed rules in law is as impossible as collecting evidence quickly enough to provide a basis to act. Making owners and operators of the largest social media services responsible for the costs and actions of harm reduction – therefore internalising these costs to the provider ("the polluter pays" principle) rather than generating external costs to society – will also make markets work better.

25. The preventive element of duty of care will also reduce the suffering of victims. It may also prevent behaviours reaching a criminal threshold. We envisage that platforms may take different approaches, and that a market could arise in which platforms develop aimed at particular groups, with different levels and types of safeguards. Content or speech/behaviour patterns that are not acceptable on one platform may find a home elsewhere and users could, of course, participate on more than one service if they so choose.

Designing a Duty of Care regime for social media

Statutory specification of harms

26. A Duty of Care would start with the statutory specification of harm. The categories would be specified by Parliament in statute at a high level of generality, as is the case in the 1974 Health and Safety at Work Act. Those under a Duty of Care would be expected to identify the level of specified harms occurring through set-

6 'In the course of its development by numerous decisions of the courts, the Common Law has hardened into rigid categories, representing the different classes of visitors to whom varying duties of care are owed. These classifications no longer represent the needs of the present day… I hope your Lordships will agree that this Bill does so in a way which does no unnecessary violence to the spirit of the common law, but frees it from the shackles imposed on it by past decisions, and enables it to go forward with renewed strength.' (Viscount Kilmuir: Lord Chancellor Occupiers Liability Bill (Lords) Lords Second reading 21 June 1956) https://api.parliament.uk/historic-hansard/lords/1956/jun/21/occupiers-liability-bill-hl

7 'The 1974 Health and Safety at Work etc Act has provided an effective framework for businesses and individuals for almost 40 years.' Common Sense Health and Safety – Review by Lord Young October 2010

up, design and/or use of their respective platforms and take steps to reduce them.

27. We list here some areas that are already a criminal offence, or are recognised as harmful behaviours in other spheres, which should be priorities for the statutory specification of harm. The duty of care aims to prevent an offence happening and so requires social media service providers to take action before activity reaches the level at which it becomes an offence. We have grouped these as separate categories, though there may be overlap between them. Examples of harm are:

- Harmful threats – eg statement of an intention to cause pain, injury, damage or other hostile action; psychological harassment, threats of a sexual nature, threats to kill, racial or religious threats known as hate crime; hostility or prejudice based on a person's race, religion, sexual orientation, disability or transgender identity. We would extend the understanding of "hate" to include misogyny.
- Economic harm – financial misconduct, intellectual property abuse.
- Harms to national security – violent extremism, terrorism, state sponsored cyber warfare.
- Emotional harm – preventing significant emotional harm suffered by users such as cyber-bullying and trolling[8].
- Harm to young people – as well as the harmful threats listed above, exposure to harmful or disturbing content, grooming, child abuse.
- Harms to justice and democracy – prevent intimidation of people taking part in the political process beyond robust debate, protecting the criminal and trial process, and preventing deliberate circulation of disinformation.

It is important to note that the proposal is not focussed on proscribing particular types of content, but aimed at the system underneath that encourages, facilitates or amplifies harmful or risky behaviours.

Role of the regulator

28. We note that the DCMS Select Committee has been considering the role of existing regulators in relation to disinformation and "fake news", and has highlighted that a number of existing bodies have an interest in this landscape, including the Information Commissioner's Office, Ofcom, the Electoral Commission and the Advertising Standards Authority.

29. In relation to online safety and harm reduction on social media, we have set out in detail in our Lords Communication Committee evidence why be believe an independent regulator is necessary to oversee the duty of care; and also why an existing regulator is preferable to a new one, given the potential costs and timescales involved in establishment by statute. (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)).

30. In our view, harm reduction requires more urgent (and less expensive) action, which can be met by an existing regulatory body assuming responsibilities for the functions set out below. Having reviewed the roles and scope of existing regulators, our recommendation is to vest the powers to reduce harm in social media services in OFCOM, given its experience, proven independence and resilience in dealing with multinational companies. With the correct

8 https://www.gov.uk/government/collections/intimidation-in-public-life

funding (potentially just a small fraction of the revenue planned to be raised by HM Treasury from taxing the revenues of internet companies), it could support an additional organisational unit to take on this work without unbalancing the rest of the organisation.

Regulatory responsibilities under Duty of Care

31. The regulator would ensure that companies have measurable, transparent, effective processes in place to reduce harm, so as to help avoid the need for individuals to take action in the first place, with a level of differentiation between high- and low-risk services as is common in other regulatory regimes, such as GDPR or health and safety regulation. This approach corresponds with the Committee's recommendation that "social media companies should not be in a position of 'marking their own homework' [and] the Government need[s] to carry out proactive work to find practical solutions to issues surrounding transparency that will work for both users, the Government, and the tech companies." (para 65)

32. Under a duty of care, the regulator would have the following responsibilities:

- Producing (through a consultative process, with users, civil society bodies, whistleblowers, researchers, journalists, consumer groups, social media companies themselves, etc), the detailed criteria for qualifying social media services; and drawing up a list of the qualifying social providers with more than one million users/members etc in the UK. Being on or off that list is challengeable by judicial review. Services could be added to the list with due process at any time, but the regulator should review the entire list periodically, perhaps every two years.
- monitoring the harm reduction processes run by the companies and supervising them into a continuous harm reduction cycle as described below (though the regulator may choose to target from time to time particular services that at that time pose the greatest risks).
- providing advice as to scope of the harms, best practice on harm reduction;
- enforcing the duty of care using tools and sanctions such as enforcement notices, prohibition notices and fines.

33. Individuals may be able to bring court action but we emphasise that this should only be in respect of systemic failures and not as a substitute for a civil action in relation to specific items of content. The regulator would not get involved in individual items of speech or be a censor. We would expect companies addressing harm reduction to run a competent ombudsman or mediation service to address individual issues that arise in their complaints process.

The Duty of Care harm reduction cycle

34. We envisage an evidence-based harm reduction cycle in which the regulator would work with the industry to create an on-going process that is transparent, proportionate, measurable and risk-based. It might look something like this:

i) Measurement of harms: the regulator would draw up a template covering scope, quantity and impact, using as a minimum the harms set out in statute. The service provider works with the regulator, consulting civil society on the template, and then surveying the extent and occurrence of harms, as set out by Parliament, in respect of the services provided by that provider;

ii) Service provider action: each service provider then runs a measurement of harm based on that template and produces and implements a plan to reduce them. The regulator would have powers in law to require the qualifying companies to comply. The companie would be required to publish for consultation their survey results and plans of action in a timely manner, establishing a first baseline of harm, and including details on:

- what actions they have taken immediately;
- actions they plan to take;
- an estimated timescale for measurable effect; and
- basic forecasts for the impact on the harms revealed in the baseline survey and any others they have identified

iii) Re-measurement and assessment: periodically, the harms are re-measured, the effectiveness of the plan assessed and, if necessary, further changes to company practices and to tools available to users introduced. The re-assessment process would provide the first progress baseline and would show four likely outcomes; that harms:

- have risen;
- stayed the same;
- have fallen; or
- new harms have occurred.

35. If harms surveyed in the baseline have risen or stayed the same, the companies concerned will be required to act and plan again; the regulator may also take the view that the Duty of Care is not being satisfied and, ultimately, may take enforcement action (see below). If harms have fallen then companies will reinforce this positive downward trajectory in a new plan.

36. The cycle then repeats, with harms measured and new plans produced by the service providers, while the regulator monitors progress towards overall harm reduction, taking action where necessary.

37. It is important to emphasise that we do not envisage the harm reduction processes to necessarily involve take-down processes. Moreover, we do not envisage that a system that relied purely on user notification of problematic content or behaviour and after the event responses would be taking sufficient steps. Tools/techniques that could be developed and deployed include:

- the development of a statement of risks of harm, prominently displayed to all users when the regime is introduced and thereafter to new users;
- an internal review system for risk assessment of new services prior to their deployment (so that the risk is addressed prior to launch);
- the provision of a child protection and parental control approach, including age verification, (subject to the regulator's approval/ adherence with standards);
- the display of a rating of harm agreed with the regulator on the most prominent screen seen by users;
- development – in conjunction with the regulator and civil society – of model standards of care in high risk areas such as suicide, self-harm, anorexia, hate crime etc; and
- provision of adequate complaints handling systems with independently assessed customer satisfaction targets and also produce a twice-yearly report on the breakdown of complaints (subject, satisfaction, numbers, handled by humans, handled in automated method etc.) to a standard set by the regulator.

The regulator would also:

- publish model policies on user sanctions for harmful behaviour, sharing research from the companies and independent research;
- set standards for and monitoring of response time to queries (as the European Commission does on extremist content through mystery shopping);
- co-ordinate with the qualifying companies on training and awareness for the companies' staff on harms;
- contact social media service companies that do not qualify for this regime to see if regulated problems move elsewhere and to spread good practice;
- publish a forward-look at non-qualifying social media services brought to the regulator's attention that might qualify in future;
- support research into online harms – both funding its own research and co-ordinating work of others;
- establish a reference/advisory panel to provide external advice to the regulator – the panel might comprise civil society groups, people who have been victims of harm, free speech groups; and
- maintain an independent appeals panel (in relation to complaints about the companies' responses rather than being a first port of call for content disputes).

Sanctions and compliance

38. Some of the qualifying social media services will be amongst the world's biggest companies. In our view the companies will want to take part in an effective harm reduction regime and comply with the law. The companies' duty is to their shareholders – in many ways they require regulation to make serious adjustments to their business for the benefit of wider society. The scale at which these companies operate means that a proportionate sanctions regime is required.

39. Throughout discussion of sanctions, there is a tension with freedom of speech. The companies are substantial vectors for free speech, although by no means exclusive ones. The state and its actors must take great care not to be seen to be penalising free speech unless the action of that speech infringes the rights of others not to be harmed or to speak themselves. The sanctions regime should penalise bad processes or systems that lead to harm and all processes leading to the imposition of sanctions should be transparent and subject to a civil standard of proof.

40. Sanctions would include:

- Administrative fines in line with the Data Protection Act 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
- Enforceable undertakings.
- Adverse publicity orders – the company is required to display a message on its screen most visible to all users detailing its offence.
- Forms of restorative justice – where victims sit down with company directors and tell their stories face to face.

Sanctions for exceptional harm

41. The scale at which some of the qualifying social media services operate is such that there is the potential for exceptional harm, where activity on its platforms (potentially as a result of design flaws that the regulator may have flagged) has, for example, provoked a riot or resulted in sexual harm to hundreds of young people.

42. In extreme cases, should there be a power to send a social media services company director to prison or to turn off the service? Regulation of health and safety in the UK allows the regulator in circumstances which often involve a death or repeated, persistent breaches to seek a custodial sentence for a director. The Digital Economy Act contains power for the age verification regulator to issue a notice to internet service providers to block a website in the UK.

43. None of these powers sit well with the protection of free speech on what are generalist platforms – withdrawing the whole service due to harmful behaviour in one corner of it deprives innocent users of their speech on the platform. However, the scale of social media services mean that acute large-scale harm can arise that would be penalised with gaol elsewhere in society. Further debate on this aspect is needed.

Next steps: simple legislation to pass quickly

44. As the Select Committee have noted: "*Within social media, there is little or no regulation. Hugely important and influential subjects that affect us—political opinions, mental health, advertising, data privacy—are being raised, directly or indirectly, in these tech spaces. People's behaviour is being modified and changed as a result of social media companies. There is currently no sign of this stopping*" (para 55). That is why action to reduce harm on social media, and limit the manipulation of users' behaviour by the platforms, 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.

45. 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; 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 to provide safe machines are brief.

46. For social media services, a Duty of Care and key harms are simple to express in law, requiring ten clauses or fewer 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. The most clauses might be required for definitions and parameters for the list the regulator has to prepare. 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.

47. This makes for a short and simple Bill in Parliament that could slot into the legislative timetable, even though it is crowded by Brexit legislation. We are considering drafting such a Bill to inform debate and test our estimate.

48. In the meantime, we continue to talk to policymakers, Parliamentarians, regulators, academics, campaigners and civic society groups to build a consensus around the need for action and make the case for the adoption of this particular approach.

49. We would welcome the opportunity to talk to the Select Committee as they continue their deliberations and work up their final report.

Professor Lorna Woods, University of Essex
William Perrin, Trustee Carnegie UK Trust
11 October 2018

Annex: Further reading

Carnegie Trust: Online Harms in Social Media (including links to the series of blog posts by Lorna Woods and William Perrin on the Duty of Care proposals): https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/

William Perrin: "It's quite possible for social media firms to protect children. Here's how they could do it" (Daily Telegraph, 21 June 2018) https://www.telegraph.co.uk/news/2018/06/21/quite-possible-social-media-firms-protect-children-could-do/

Lorna Woods: Evidence to Lords Communication Committee Enquiry "The Internet: to regulate or not to regulate?": http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/communications-committee/the-internet-to-regulate-or-not-to-regulate/written/82684.html

Lorna Woods: "A proposal for harm reduction in social media": https://inforrm.org/2018/07/17/carnegie-uk-trust-a-proposal-for-harm-reduction-in-social-media-lorna-woods/