Carnegie UK Trust—written evidence (FEO0044)

House of Lords Communications and Digital Committee inquiry into Freedom of Expression Online

1. We welcome the Committee's inquiry into Freedom of Expression online and the opportunity to submit evidence. Our response to the questions below is drawn from the thinking that informs our work on the development of a statutory duty of care for online harm reduction – work that the Committee previously endorsed in its 2019 report ("Regulating in a Digital World") – and we provide references to relevant material below.

2. We would be happy to provide further information on our work in writing or to discuss it with Committee members at a future evidence session.

About our work

3. The Carnegie UK Trust was set up in 1913 by Scottish-American philanthropist Andrew Carnegie to improve the wellbeing of the people of the United Kingdom and Ireland. Our founding deed gave the Trust a mandate to reinterpret our broad mission over the passage of time, to respond accordingly to the most pressing issues of the day and we have worked on digital policy issues for a number of years.

4. In early 2018, Professor Lorna Woods (Professor of Internet Law at the University of Essex and member of the Human Rights Centre there) and former civil servant William Perrin started work to develop a model to reduce online harms through a statutory duty of care, enforced by a regulator. The proposals were published in a series of blogs and publications for Carnegie and developed further in evidence to Parliamentary Committees[1]. The Lords Communications Committee[2] and the Commons Science and Technology Committee[3] both endorsed the Carnegie model, as have a number of civil society organisations[4]. In April 2019, the government's Online Harms White Paper[5], produced under the then Secretary of State for Digital, Culture, Media and Sport, Jeremy Wright, proposed a statutory duty of care enforced by a regulator in a variant of the Carnegie model. France[6], and the European Commission's pre-Christmas publications (the Digital Services Act and the European Democracy Action Plan) both draw on a systemic, risk-based approach to harm reduction[7].

---

1 Our work, including blogs, papers and submissions to Parliamentary Committees and consultations, can be found here: https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/

2 https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/29902.htm

3 https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/82202.htm

4 For example, NSPCC: https://www.nspcc.org.uk/globalassets/ documents/news/taming-the-wild-west-web-regulate-social-networks.pdf; Children's Commissioner: https://www.childrenscommissioner.gov.uk/2019/02/06/childrens-commissioner-publishes-astatutory-duty-of-care-for-online-service-providers/; Royal Society for Public Health: https://www.rsph.org.uk/our-work/policy/wellbeing/new-filters.html

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

6 French-Framework-for-Social-Media-Platforms.pdf (thecre.com)

5. In December 2019, while waiting for the Government to bring forward its own legislative plans, we published a draft bill[8] to implement a statutory duty of care regime, based upon our full policy document of the previous April[9]. We are also supporting Lord McNally on his Private Bill (The Online Harm Reduction Regulator (Report) Bill)[10], introduced into the House of Lords on 14 January 2020, which would provide an opportunity for full Parliamentary debate on the nature of the regulatory regime and, if passed, empower OFCOM to prepare for its introduction.

6. We welcome the fact that the Government has now published its full response to the Online Harms White Paper[11] and we are now considering its proposals in detail.

Overview of Freedom of Expression Guarantees

7. Freedom of expression is protected by both the European Convention on Human Rights (ECHR) and by the International Covenant on Civil and Political Rights (ICCPR). Although the text of these two instruments differ slightly, there some commonalities. Freedom of expression is understood broadly, and political speech is highly protected. Nonetheless, the right is not unlimited. In both instances, an individual's right to free speech may be limited provided the grounds of derogation – Article 10(2) ECHR and Article 19(3) ICCPR – are satisfied. In both cases, this involves a three-stage test: the restriction must be provided by law; pursue a legitimate aim as identified in the provision; and be proportionate.

8. While the protection offered by these provisions is broad, covering in both instances speech that is offensive, protection is not awarded to speech that constitutes an abuse of rights.[12] In *Lopez Burgos v Uruguay,* Tomuschat opined that:

> "individuals are legally barred from availing themselves of the same rights and freedoms with a view to overthrowing the regime of the rule of law which constitutes the basic philosophy of the Covenant."[13]

9. To similar effect, the European Court of Human Rights has held that:

---

7 Please see Prof Lorna Woods' blog posts on both the European Democracy Action Plan (http://eulawanalysis.blogspot.com/2020/12/european-democracy-action-plan-overview.html) and the Digital Services Act (http://eulawanalysis.blogspot.com/2020/12/overview-of-digital-services-act.html)

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

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

10 https://services.parliament.uk/bills/2019-21/onlineharmsreductionregulatorreportbill.html

11 Our initial response is here: https://www.carnegieuktrust.org.uk/news/online-harms-initial-response/

12 Article 17 ECHR; Article 5(1) ICCPR

13 Individual Opinion of Mr Christian Tomuschat in *Burgos v Uruguay*, Communications no. R12/52 Supp. No. 40 (A/36/40), at 76, 29 July 1981

"[t]he general purpose of Article 17 is to prevent individuals or groups with totalitarian aims from exploiting in their own interests the principles enunciated by the Convention."[14]

10. These provisions might be said to be linked to the idea of militant democracy. Article 17 ECHR has been used in relation to the promotion and justification of terrorism and war crimes; incitement to violence; threat to territorial integrity and constitutional order of a state; promotion of totalitarian ideologies; incitement to hatred; and Holocaust denial. These provisions come into play only rarely and there is a high threshold[15] for the application of either provision. Should a claimant not fall foul of Art 5(1) ICCPR or Article 17 ECHR, an interference with that applicant's speech could still be justified under Article 19(3) ICCPR or Article 10(2) ECHR, as can be seen below.

Responses to the Committee's questions

Question 1: Is Freedom of Expression under threat online? Are there differences between exercising freedom of expression online versus offline?

11. The digital environment has changed the communication environment in a number of ways. Users communicate with one another at a distance, a fact which seems to have led to behaviours online occurring which would not be acceptable offline (disinhibition[16]). Disinhibition seems to be connected to the structure and nature of the online environment; it is an open question whether a failure to regulate content (or for platforms not to apply (equally) their own terms and conditions) has led to this problem getting worse, constituting a shift in public discourse. Such a shift would not be neutral as a more aggressive environment is more difficult for vulnerable and minoritised groups to tolerate.

12. The vast majority of cases from the latter half of the twentieth century and the first decade of the twenty-first deal with what might be termed 'public speech'. That is, the expression is formed and disseminated for public consumption, meaning that, on the whole, the content has been thought through and delivered bearing that in mind. Indeed, for some speakers (the traditional media) there are professional rules which provide a framework for such speech (journalistic ethics and content regulation). While these rules continue to apply to these speakers, the speech environment has changed. Search engines can find material that have been delivered to a small audience, protected by obscurity; discovery/targeting tools in general may expose users to content that is from a different or more extreme viewpoint, whether for good or bad. With social media in particular much speech that could be termed day-to-day chit chat is made public and recorded, arguably blurring the boundaries between public and private speech.

---

[14] *Kasymakhunov and Saybatalov v Russia (*26261/05 and 26377/06), judgment 14 March 2013, para 103

[15] *Perinçek v Switzerland* (27510/08), judgment 15 October 2015 [GC].

[16] Suler J. 'The online disinhibition effect' (2004) 7 *CyberPsychology and Behavior,* 321-326, doi:10.1089/1094931041291295.

13. While freedom of expression has traditionally been protected by Article 10 ECHR (and analogous rights in other instruments including Article 19 ICCPR), private speech – that which builds relationships – was probably protected by Article 8 ECHR (and analogous rights in other instruments). In the context of defamation and privacy claims made usually against media organisations, where freedom of expression and the right to private life are to a large extent in opposition, the European Court of Human Rights has held that the outcome of any such conflict should not depend on whether the claim is brought under Article 8 (by the subject of speech) or under Article 10 (by the speaker). The State is also under an obligation of the State to take action to protect those rights.[17]

14. Freedom of expression is a compound right: it is the right of the speaker and of the audience; the right to speak as well as to be silent and includes the right to hold an opinion. It does not include the right to be listened to; audiences have some choice in this regard. Insofar as freedom of expression is aimed at enabling an individual developing themselves through exposure to ideas and information, the right deals with various aspects of that process – though note the rights in the different human rights instruments recognise these differently.

15. Article 10 ECHR and Article 19 ICCPR both recognise the right to hold opinions, and in relation to Article 19 ICCPR it is clear that this element of the right is not subject to any limitation[18]. Some have questioned the impact of online tracking for the purpose of profiling individuals and determining their interests, views and emotional state on this right (this may have links also with the right to belief).[19] Put simply, where an individual receives their information from commercial algorithms online is their freedom to form opinions and to hold them (without expressing those beliefs) impeded; to what extent is an individual manipulated in the (commercial) interests of others? This concern is perhaps one that is distinctive in relation to the online context which is in the main currently based on what Zuboff termed the 'surveillance capitalism' model. Moreover, Article 19 ICCPR makes explicit what is implicit in Article 10 ECHR - that there is a right to seek out information.

16. Both Article 19 ICCPR and Article 10 ECHR recognise that there is a right to receive information as well as to impart it, though the jurisprudence here is less clear. Much of the case law has focussed on the right to specific pieces of information – freedom of expression seen as freedom of information. It is less clear whether there is a right to access information environments more generally, for example the right to access programming generally available on television. The European Court of Human Rights has recognised that the audience might have some interests worthy of protection, especially as regards the quality of information accessible to

---

17 Woods, L 'Social Media: it is not just about Article 10' in Mangan and Gillies (eds) *The Legal Challenges of Social Media* (Edward Elgar, 2017).

18 See Human Rights Committee General Comment No 24: Article 19- Freedom of Opinion and Expression, 12 September 2011, available: https://www2.ohchr.org/english/bodies/hrc/docs/GC34.pdf, paras 9-10.

19 Alegre, S. 'Rethinking freedom of thought for the 21st century' (2017) 3 *EHRLR* 221

them. While the position of the rights of the audience is far from fully understood or complete[20], this has mainly arisen in the European Court's jurisprudence on positive obligations and its view of the State as the ultimate guarantor of pluralism[21], though the meaning of pluralism itself is fluid[22] and the extent of the State's obligation uncertain.[23] The last sentence of Article 10(1) ECHR recognises that the State may regulate mass media (including the cinema), though this must be justified in the public interest; this seems to extent to controls over types of content to be made available, and in the context of public service broadcasting, that news is balanced and accurate[24]. General Comment 34 on the interpretation of Article 19 ICCPR reiterates that public service media operate in an independent manner and that the State should promote plurality of the media. The main concern identified at the point in time that General Comment 34 was written was that of media concentration (a point of concern to the European Court also), which may also be a concern in the social media environment in particular.

17. Another threat to freedom of expression, though not unique to the online environment is particularly noticeable there, is the impact of the speech of some individuals on others, especially those in vulnerable and minoritised groups. This may have an impact on the willingness of those affect to contribute to public debate (understood broadly to include academic debate) and to stand for public office. This latter issue could be seen to be, a violation of other international obligations, for example the Convention on the Elimination of Discrimination Against Women (CEDAW).[25] The UN Convention on the Elimination of All Forms of Racial Discrimination (CERD) recognises the link between hate speech and transmission of racists ideas. The distinctive element here could be said to be the continuous nature of the threat; it does not stop at the front door raising questions as to where people might feel safe. It might be said that States are under a positive obligation to remove threats from private individuals to others' speech; to limit the chilling effect of online intimidation.

Does the state have an obligation to protect people from harms online?

18. This point links to Article 8 and positive obligations the State would be under in that regard. Article 8 ECHR is broad and extends beyond privacy. It protects also the physical and moral or psychological integrity of the person[26] as well as 'multiple aspects of the person's physical and social

---

20 Flauss, J-F. 'The European Court of Human Rights and the Freedom of Expression (2009) 84 *Indiana Law Journal* 809, pp 812-3

21 *Informationsverein Lentia and Others v. Austria*, (A/276) 24 November 1993, para 38

22 Note McQuail's typology of pluralism: (1) reflecting proportionately existing differences in society; (2) giving equal access to any different points of view; and (3) by offering a wide range of choice for individuals – McQuail, D., *Media Performance Mass Communication and the Public Interest* (London: Sage, 1992)

23 Komorek, E. 'Is media pluralism? The European Court of Human Rights, the Council of Europe and the issue of media pluralism' (2009) 3 *EHRLR* 395, p. 404

24 *Manole v Moldova* (App no 13936/02), 17 September 2009, para 100, see also para 107

25 See CEDAW Recommendation 35, referring to violence against women in "technology-mediated environments, such as contemporary forms of violence occurring in the Internet and digital spaces".

26 *X and Y v Netherlands* (A/91), judgment 26 March 1985, para 22

identity'[27] and the right to personal development, whether in terms of personality or of personal autonomy. In *Beizaras and Levickas v Lithuania*[28] the applicants (two young men), posted a photograph of themselves kissing on a public Facebook page triggering hundreds of virulently homophobic comments. There was no prosecution. This failure, the Court of Human Rights determined, revealed a discriminatory frame of mind on the part of the relevant authorities and that there was a violation of Article 14 in relation to their Article 8 rights (seen as psychological well-being). States have been held to have a duty to protect a person from cyberbullying by that person's partner.[29] The Court has of late become increasingly aware of the problem of domestic violence. In the same way that a State has a narrower freedom of assessment in relation to political speech under Article 10, that margin will be narrow under Article 8 when a particularly important facet of an individual's existence or identity is at stake, as can be seen in *Goodwin v UK* who challenged the UK government's lack of legal recognition of the change of gender for a post-operative transexual[30]. Article 8 includes individuals' identity as part of a group (e.g. the dignity of the victims and the dignity and identity of modern-day Armenians in the context of the mass deportations and massacres suffered by the Armenians in the Ottoman Empire in 1915 and beyond in *Perinçek*.)[31]

19. When balancing rights, the Court will take into account the importance of the interest at stake and a State's obligations are particularly strong when involving "fundamental values" or "essential aspects" of private life (note that the case law[32] on balancing freedom of expression and the right to reputation in the context of the media would seem to be relevant in the context of the imposition or failure to impose criminal sanctions in a given case). Where these fundamental issues are at stake, Member States may be under an obligation to enact criminal provisions to effectively punish the behaviour undermining the victim's article 8 rights. Examples include the punishment of rape and the protection of children and other vulnerable individuals, as well as serious domestic violence.[33] In *KU v Finland*, the Court held that there was an obligation to protect a minor against a malicious misrepresentation where his details were posted on an Internet site, constituting a threat to his physical and mental welfare.[34] In less serious cases, the State has freedom to choose what in its view is the most appropriate framework (and whether to introduce a civil or criminal regime to ensure protection), but it is still under an obligation to ensure respect for these rights.

---

27 *S and Marper v UK* (App nos 30562/04 and 30566/04), judgment 4 December 2012 [GC], para 66

28 *Beizaras and Levickas v. Lithuania* (app no 41288/15), judgment 14 January 2020

29 *Buturugă v Romania* (app no 56867/15) judgment 11 February 2020, para 74, 78-79; on obligations with regard to vulnerable persons and groups more generally see Woods, L. 'Social Media: it is not just about Article 10' in Mangan and Gillies (eds) *The Legal Challenges of Social Media* (Edward Elgar, 2017).

30 *Goodwin v UK* (app no 28957/95), [GC] judgment 11 July 2002, para 90

31 *Perinçek v Switzerland* (27510/08), judgment 15 October 2015 [GC].

32 *Axel Springer AG v Germany* [GC], judgment 7 February 2012, paras 89-95

33 *M.C. v. Bulgaria*, (app no. 39272/98) ECHR 2003-XII, para 150; *BV v Croatia* (dec) 15th December 2015, para 151

34 *KU v Finland* (app no 2872/02), judgment 2nd December 2008

Question 3: Is online user-generated content covered adequately by existing law and, if so, is the law adequately enforced? Should 'lawful but harmful' online content also be regulated?

20. There are many content rules which in principle apply to online content, ranging from defamation and misuse of private information, through advertising regulation to the criminal law, which itself covers content of different levels of severity (e.g. Protection from Harassment Act, rules relating to revenge porn and the rules relating to child sexual abuse and exploitation). The difficulty seems to be in enforcement. As regards civil law, enforcement pre-supposes that the victim has the finance and the fortitude to take action, both of which constitute not inconsiderable barriers. By contrast to the position as regards, for example, data breaches, such claims are unlikely to be able to be dealt with as a class or representative action. If we saw victims taking action, this would also put a considerable burden on the courts. In principle, rules such as advertising rules, apply; the question is one regulatory priorities (especially in the face of a strong narrative that says online content is in some way beyond regulation) and the practicalities of enforcement (even insofar as identifying the person responsible), which may require the cooperation of the platforms; whether this is readily forthcoming is another question. Within the criminal law, enforcement may vary depending on the type of crime. While no-one doubts the severity of child sexual abuse and exploitation, questions might be raised as to whether the police are adequately trained and equipped to deal with harassment and online sexual offences (as can be seen from the discussion in the Law Commission Report on Communications Offences). The result is patchy and uneven enforcement.

21. In our view, if a systemic approach to regulation is adopted – that is, one that looks at the chain of communication and the various points at which the platform may affect content – the distinction between 'illegal' or 'criminal' on the one hand, and 'harmful but legal' on the other does not make sense. It we take an example, that of age verification, it is more likely to apply at the point that a user accesses the platform and therefore bites across all types of content. Design features may have cross-harm effects: the recommender algorithm may pull up all sorts of problematic content (misinformation about Covid and 5G, through to terrorist content and CSAEM). Moreover, if we envisage what the Victims Commissioner has terms a "pro-active" duty of care, this involves a company looking at its service and operations to see how they might result in harm (ie forward-looking) rather than waiting for the harm to occur so it can be classified as harmful or not, at which point the opportunity for many of the softer interventions has passed and we are likely left with the limited tool box of solutions of takedown or banning of the user. Boundaries between harmful and criminal may not be clear-cut in any event, in some instances being a matter of intensity, not type. Further, while the liability of platforms for the content of third parties under the current immunity rules is affected by the distinction between civil and criminal law, that distinction has less relevance here. The criminal law in particular deals with culpability of the speaker and rules designed with punishment are not well adapted for focussing on the harm suffered by victims; indeed, the requirement of *mens rea* may

exclude some actors from a criminal offence while including others although the impact on the victim in both sets of cases is the same.

22. We do not deal with broadcasting through criminal law but through a civil regulation regime. There is no reason why online user generated content should not be regulated through a civil regulation system. Such systems prove themselves to be more responsive and effective than simply 'enforcing the law' and are commonly used in other sectors of the economy.

Question 4: Should online platforms be under a legal duty to protect freedom of expression?

23. Human rights frameworks tend to bite on state actors and public bodies and the platforms are private actors. An obligation to comply with human rights norms would set them apart from other private actors, even large corporations. This seems to us to bring us into the terrain of regulation analogous perhaps to the universal service or must carry obligation imposed on some communications providers. In the absence of such regulation provision, we are of the opinion that companies should adopt a corporate social responsibility approach (Ruggie Principle) and bear in mind *all* human rights and not just freedom of expression. Freedom of expression is no more and no less important than any other right. To single it out might give an unfortunate signal diminshing the rights of victims and those silenced by the speech of others.

Question 5: What model of legal liability for content is most appropriate for online platforms?

24. Assuming that platforms are under a duty of care with regard to the design of their platforms and business operations, we are of the opinion that the current framework (rather than that adopted in the US model) is appropriate. It may be, however, that some consideration should be given to the introduction of a "good Samaritan" clause when a platform (or other intermediary) has taken reasonable steps for good cause.

Question 6: To what extent should users be allowed anonymity online?

25. The impact of anonymity on the types of speech is not well-understood; we are not aware that there is conclusive evidence that anonymity would silence trolls for example.[35] Moreover, there are groups (e.g whistle-blowers) for whom anonymity is important. On the other side, anonymity and pseudonymity can be abused (e.g catfishing). Given this ambivalent position we take the view that anonymity should not be automatically prohibited but that it would be a risk-factor that a platform offering services should consider.

Question 8: How do the design and norms of platforms influence the freedom of expression? How can platforms create environments that reduce the propensity for online harms?

---

[35] For example, https://www.pinknews.co.uk/2020/12/11/kate-scottow-appeal-conviction-trans-woman-abuse-twitter-troll/

26. Our proposal focussed on the underpinning social media service in terms of its design and business operations[36]. This differs from an approach which seeks to specify particular items of content as problematic and sees solutions only through the lens of take-down. This latter approach is the traditional publisher model which operates on the basis of a distinction between publisher, who has a role in the creation/selection of content, and the intermediary, who knows nothing of the content. Such a binary approach, we have argued, is inappropriate for the Internet context, specifically as regards social media. The Carnegie proposal aims to tackle the abuse-enabling environment that some platforms seem to have become by focussing on the question of whether service providers have adequately taken into account the risks of the way their systems have been designed and run – whether this be through design features influencing content or nudging users to certain sorts of behaviour at the point of creation or engagement (for example, through reward mechanisms); or in the context of how content is discovered (for example, through the use of algorithms to promote and recommend content – and problems have been recognised by the Council of Europe[37]). Consideration should be given to the tools the platforms give to third parties to push messages to targeted groups; and consider whether their controls about access to these tools is sufficient and indeed whether the characteristics by which the audience is segmented is appropriate (racist and discriminatory terms at the very least should be excluded). In this context, we emphasise that, while there is potential for overlap between the two claims, freedom of reach is very different from freedom of speech[38]. Takedown would be a mechanism of last resort – though in some instances may well be appropriate (consider, for example, the need to tackle child sexual abuse and exploitation material). In envisaging that platform operators undertake a risk assessment, the proposal aligns with the Council of Europe recommendation on the roles and responsibilities of internet intermediaries,[39] as well as by the UN Special Rapporteur on Freedom of Expression[40]. More recently still, in his 2019 report on hate speech, the Special Rapporteur noted ideas similar to those put forward under the duty of care, that is "restricting its virality, labelling its origin, suspending the relevant user, suspending the organization sponsoring the content, developing ratings to highlight a person's use of prohibited content, temporarily restricting content while a team is reviewing, demonetizing, minimizing its amplification, interfering

---

36 See all our work here: https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/ In particular, our full reference paper of 2019 which sets out the detail on our proposal for a statutory duty of care for online harm reduction, enforced by an independent regulator: https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2019/04/08091652/Online-harm-reduction-a-statutory-duty-of-care-and-regulator.pdf

37 Declaration on the manipulative capabilities of algorithmic processes (Decl (13/02/2019)1), para 9, available: https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b

38 See the 2019 comprehensive paper by Professor Lorna Woods on the duty of care and fundamental freedoms: https://www.carnegieuktrust.org.uk/publications/doc-fundamental-freedoms/

39 Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries, 7 March 2018, para 2.1.4, available: https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680790e14

40 Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (A/HRC/38/35), para 55

with bots and coordinated online mob behavior, adopting geolocated restrictions, and even promote counter-messaging" as a preferable alternative to take down regimes."[41]

27. Platforms can create better environments by performing user research into what system designs enable a healthier dialogue and implementing them. Facebook reportedly experimented with such a service but found that it was less commercially successful for them and so did not pursue it.[42]

Question 9: How could the transparency of algorithms used to censor or promote content, and the training and accountability of their creators, be improved? Should regulators play a role?

28. We recognise that algorithms play an important role in many if not all platforms' business model and that total openness is not possible. It is our opinion, however, that the platforms should be able to explain at a general level the values/metrics incorporated, and be able to justify this in the light of foreseeable harms on their platforms. Design of algorithms should therefore also allow for some form of auditing function to support any such explanation.

29. In this regard we would like to flag the importance of many platforms in terms of being a gatekeeper and affecting the news agenda. Further, recommender algorithms have come under particular criticism for prioritising extremist and unreliable content. Seemingly this is because it keeps users engaged. Done thoroughly, a review of the recommender algorithm would consider whether orientating the results towards this outcome (without further consideration) is safe. A recommender algorithm could prioritise material from reliable sources, though this might raise questions as to how. One suggestion could be that regulated services, or services which comply with an external code of ethics, are positively weighted within the recommender machine. In this the regulator would play a role in determining whether the platform had met its obligations.

30. If OFCOM is to be the regulator then the government should ensure that it is able to use the sweeping powers to request information (Section 135 of the Communications Act 2003[43]) that it uses with great success in regulating other sectors. An open-ended power to request information and for the platform to generate data from its systems that may not yet exist (ie run reports on databases in response to a regulator request) should enable adequate transparency.

31. We also support the proposals put forward by AWO/Ada Lovelace Institute on the importance of including in the Online Harms Bill powers for the regulator to carry out algorithmic auditing.[44]

---

41 Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (A/74/489)

42 "Facebook Struggles to Balance Civility and Growth" (New York Times, 24 November 2020) https://www.nytimes.com/2020/11/24/technology/facebook-election-misinformation.html

43 https://www.legislation.gov.uk/ukpga/2003/21/part/2/chapter/1/crossheading/information-provisions

44 https://www.adalovelaceinstitute.org/blog/algorithms-social-media-realistic-routes-to-

Question 10 How can content moderation systems be improved? Are users of online platforms sufficiently able to appeal moderation decisions with which they disagree? What role should regulators play?

32. In our view, adequate resources should be devoted to customer service (complaints and content moderation) and that the claim that a platform has got too big to do this effectively is unacceptable. While machine learning and AI tools may help, they carry their own risk (mainly bias, but also difficulty understanding context). In our proposal we suggested that, learning from the telecommunications regime, there should be an obligation on platforms to provide 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.

33. An appeals system is important. This should not be skewed in favour of just reviewing decisions to take content down, but should also include appeals against decisions to leave content up, so as to provide equal protection to Article 8 ECHR rights as Article 10 ECHR (or their equivalent rights in the ICCPR). Such systems must be visible and easy to use – and designed with the age or mental capabilities of the relevant user group in mind. Importantly, the platforms should not seek to exclude the ability of a user to bring an action in the courts if the user so desires. There should be transparency around any appeals process too, including the understanding of the community standards.

34. While most complaints will be dealt with it seems as a matter of community standards, it is important that users should be able to complain about a violation of the law too.

January 2021

---

regulatory-inspection/