

Written evidence from Carnegie UK Trust (COV0165)

Background

1. [Carnegie UK Trust](#) (CUKT) is a not-for-profit organisation focused on improving wellbeing through a range of research, advocacy and community programmes. Since early 2018, it has supported work on new proposals for internet harm reduction instigated by William Perrin (a former UK Civil Servant, who is now a Carnegie UK Trustee) and Professor Lorna Woods (Professor of Internet Law, University of Essex, and an EU national expert on free speech and communications regulation). Their work has focussed on the development of a statutory duty of care to reduce reasonably foreseeable harms on social media enforced by a regulator. Full details on the proposals can be found on the CUKT website.¹
2. We have previously submitted evidence to the Joint Committee's Inquiry on Democracy, Free Speech and Freedom of Association². This submission provides an update on our thinking on how a statutory duty of care for online harm reduction would fit with the right to freedom of expression, taking health misinformation and disinformation in the context of the pandemic as a case study.

Freedom of Expression and the Disinfodemic

3. Covid-19 has sparked an "infodemic"³ or as UNESCO termed it, a "disinfodemic"⁴ which sows confusion about the existing medical knowledge, disrupts public health information campaigns and spreads rumours and conspiracy theories that are demonstrably false, some of which may adversely impact particular minority groups. Access to reliable and good quality information is important at any time to allow people to come to valid conclusions on issues of public interest, but this is particularly so during a public health crisis when people's lives may depend on it. While misinformation and disinformation are not novel to coronavirus, "[i]t is more toxic and more deadly than disinformation about other subjects"⁵. As such, it raises questions not just about the speaker's right to freedom of expression, but the audience's right to information (Article 10 ECHR) and also their rights to bodily integrity (Article 8 ECHR) and even to life (Article 2 ECHR). As regards Article 10, the Court has stated that freedom of expression imposes:
"a duty on the State to ensure, first, that the public has access through television and radio to impartial and *accurate information* and a range of opinion and comment"⁶ [emphasis added].
While diversity and pluralism are fundamental to a democratic society, so is the accuracy of facts.
4. Covid-19 constitutes a case study on the role of social media platforms (and potentially other internet intermediaries such as search engines) in the creation and dissemination of misinformation and disinformation. With the exception of Pinterest (which has had a health

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

2 <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/human-rights-committee/democracy-free-speech-and-freedom-of-association/written/101682.html>

3 See WHO, available: <https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-and-cybercrime-covid-19> [accessed 14 July 2020].

4 Posetti and Boncheva *Disinfodemic: Deciphering Covid-19 Disinformation*, https://en.unesco.org/sites/default/files/disinfodemic_deciphering_covid19_disinformation.pdf [accessed 14 July 2020]

5 Posetti and Boncheva, *ibid*, p. 2

6 *Manole v Moldova* (App no. 13936/02), judgment 17 September 2009, para 100.

misinformation policy since 2017 and took a decision last year not to have anti-vax material on its platform⁷), the social media platforms did not have public health harm policies in place on their platforms. Their orientation towards a radical and unbalanced version of freedom of expression has blinded them to the risks of speech, save that arising from the most egregious criminal offences. In the Covid-19 context, this meant that they were unable and unready to take action once those harms – largely through the spread of mis/dis-information in the form of fake health claims, harmful health advice or 5G conspiracy theories – emerged online, despite the fact that these were not new issues. In any event, in the absence of a regulatory system we have no way to verify the claims of social media companies about the effectiveness of their actions.

5. The Covid-19 pandemic has emphasised the ever-more urgent need for online harms legislation to be brought forward at the earliest opportunity as well as the challenge that arises when the determination of action on “acceptable” content and behaviour is left to the platforms to define, rather than setting regulatory expectations that address the systemic design and information flows that facilitate the spread of many types of harms online. If we accept that the State’s positive obligations under Article 10 require it to take steps to ensure the reliability of the public information environment, even while maintaining diversity and plurality of facts and opinion, this suggests that content that causes harm should be limited, no matter whether that content is considered illegal or not. Indeed, there is a risk that excluding harmful but not criminal content would mean that there is a discrepancy between the approach to content on social media platforms and content disseminated across other media. An approach which focusses on the categorisation of speech as illegal or not overlooks the fact that it is not only the expression of the harmful content in itself that causes problems but the speed and scale of its spread and promotion – a spread encouraged and facilitated by the platforms’ own system design, for example their algorithms, recommender models, reliance on user profiling and micro-targeting⁸, or nudges to users to like or share content without time for reflection. A significant part of the problem, in our view, relates to these information flows, and this is an aspect that does not readily fit a framework designed round illegal/harmful.
6. This systemic approach demonstrates that actions related to disinformation are not limited to take down, but can focus on the realignment of incentives for creation of content, the extent to which the platform contributes to virality of content, the extent to which this can be weaponised, and the promotion of reliable content. Indeed, some of the measures that are somewhat belatedly being taken by the major platforms reflect this approach. For example, WhatsApp’s “velocity limiter” to reduce the number of times things can be forwarded which, it claims, has led to a 70% reduction in “highly forwarded” messages on its services.⁹ Most social media companies have also introduced changes to the design of their services to promote information from authoritative services while reducing the prominence given to unverified information by their discovery and search functions. In this, a systemic approach to platform regulation (as detailed in Carnegie’s proposals), which does not rely on takedown of content to address all problems on the Internet, is more proportionate in terms of the impact on the speaker’s freedom of expression, whilst also taking into account the rights of others and the public interest.

7 Pinterest’s community guidelines say that: “Medically unsupported health claims that risk public health and safety, including the promotion of false cures, anti-vaccination advice, or misinformation about public health or safety emergencies” It also won’t have conspiracy theories or content that originates from disinformation campaigns. (See <https://policy.pinterest.com/en-gb/community-guidelines>)

8 See the recent report by the Centre for Data Ethics and Innovation, which recommended that online targeting be subject to the duty of care: <https://www.gov.uk/government/publications/cdei-review-of-online-targeting>

9 <https://techcrunch.com/2020/04/27/whatsapps-new-limit-cuts-virality-of-highly-forwarded-messages-by-70/?guccounter=1>

7. While the social media companies, after some pressure, have taken some ameliorating steps, it remains the case, however, that the design and business model underlying the platforms encourages and facilitates the problem of mis/disinformation creating a crisis on their platforms that is damaging trust in both national governments' handling of the pandemic and trust in previously authoritative sources of information and news. The actions taken are also reactive rather than systemic: whether they are superficial changes to platform design to attempt to reduce the spread of material that is already out of control; the promotion and signposting of authoritative sources; or the flagging, correction or takedown of untrue or harmful content, once it has been established as such. These steps may be part of the solution but, as they stand, are insufficient; they do not recognise the role of the platforms in creating the environment where these problems thrive.
8. We set out in our detailed work on our website how our proposal for a systemic duty of care, enforced by a regulator, enables regulation to bite at a platform design level – tackling these information flow issues - and requires risk mitigation rather than regulating individual pieces of content. Such a systemic approach should cover not just disinformation, whether resulting in electoral harms or public health harms, but also consumer harms (including online scams, fraud and the sale of unsafe products). As we have argued, this approach is entirely consistent with the protection of people's fundamental rights, including the right to freedom of expression¹⁰.
9. In the course of its deliberations in this important area of inquiry, we would therefore urge the Committee to consider whether the Government's apparent reticence to address mis- and disinformation through the online harms regime amounts to a misinterpretation of the nature of a statutory duty of care and an unnecessarily generous interpretation of the right to freedom of speech online as being equivalent to a right to unrestricted freedom of reach. We would argue that – where harms to public health and/or, as in the case of the 5G conspiracy theories, to critical national infrastructure emerge – that a focus on systemic regulation, rather than content flagging and takedown, is far more protective of the former right, while addressing the harmful impact of the latter.

16/07/2020

10 https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2019/12/10111353/The-Carnegie-Statutory-Duty-of-Care-and-Fundamental-Freedoms.pdf