



Submission to the Australian government consultation on the Online Safety Bill

1. This submission outlines for the Australian Government our work for Carnegie UK Trust in the UK on a statutory duty of care for online harm reduction¹, an approach that is systemic: biting at the platform design level rather than the level of individual pieces of content. It is underpinned by a requirement for the platforms to undertake continuous risk assessments of the core design and operational features of their services to reduce the risk of reasonably foreseeable harms to users and, more broadly, to society. Our work has greatly influenced the UK government's own proposals on Online Harms and the anticipated Online Safety Bill, due for introduction into the UK Parliament later this year. The European Commission's Digital Services Act also reflects an approach that brings in a requirement for the biggest platforms to undertake risk assessments of their systems and processes, such as recommender algorithms.
2. We respectfully offer this submission as a point of comparison to the approach set out in the Australian Online Safety Bill, which addresses a series of the serious harms (including online abuse of adults, which is a significant addition) through an approach that is more overtly focused on the rapid notification and takedown of the most harmful categories of content to individual users, albeit backed by significant reporting and enforcement measures and enhanced powers for the e-safety Commissioner. In addition, we have attached our own draft Bill which we produced last year to demonstrate how a systemic statutory duty of care for online harm reduction, enforced by an independent regulator, could be introduced through a short piece of legislation².

Background to the Project

3. Carnegie UK Trust was set up in 1913 by Scottish-American philanthropist Andrew Carnegie to improve the well-being of the people of the United Kingdom and Ireland, a mission it continues to this day. Carnegie particularly charged the trustees to stay up to date and the trust has worked on digital policy issues for some years.

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

² <https://www.carnegieuktrust.org.uk/news/online-harms-paving-bill-introduced/>

4. In 2016 Woods and Perrin carried out work with an MP (on the private members bill ‘Malicious Communications (Social Media) Bill’) to try to ensure that social media platforms gave adequate tools to users to help them defend themselves from online abuse. This focus on design features and tools formed the basis for a larger project that Woods and Perrin commenced in early 2018 after the UK Government’s Internet Safety Strategy Green Paper in Autumn 2017 detailed extensive harms but few solutions. Initially published as a series of blogs, the work developed into a public policy proposal to improve the safety of users of internet services through a statutory duty of care, enforced by a regulator. A full reference paper³ drawing together their work on a statutory duty of care was published in April 2019, just prior to the publication of the UK Online Harms White Paper⁴. The UK government has since published both its interim⁵ and full⁶ responses to the White Paper, with significant shifts in each iteration towards a more systemic approach to regulation of harm that is closer to our model than the initial White Paper version, which was framed around a series of content-based codes of practice.
5. This work has influenced the recommendations of a number of bodies in the UK including select committees in the UK Parliament, charities and the UK Chief Medical Officers.⁷ More broadly, Woods gave evidence to the International Grand Committee on Fake News; while it did not make specific reference to this work, a report to the French Ministry of Digital Affairs referenced a “duty of care” as the proposed basis for social media regulation.⁸ This suggests that the underlying model is not specific to the UK but could be adapted by other jurisdictions.

Systems-based Regulation

6. Carnegie UK proposed a shift from the regulation of specific items of content to a focus on the design of platforms (including business models and resourcing of complaints systems). This is based on the assumption that design choices can have an impact on the content posted and the way information flows across communications platforms – including but not limited to recommender algorithms. Rather than specify individual rules, which might quickly become

³ https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2019/04/08091652/Online-harm-reduction-a-statutory-duty-of-care-and-regulator.pdf

⁴ <https://www.gov.uk/government/consultations/online-harms-white-paper>

⁵ <https://www.gov.uk/government/consultations/online-harms-white-paper/public-feedback/online-harms-white-paper-initial-consultation-response>

⁶ <https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response>

⁷ <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-a-statutory-duty-of-care-for-online-service-providers/>; <https://www.gov.uk/government/publications/uk-cmo-commentary-on-screen-time-and-social-media-map-of-reviews/>;

<https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/82202.htm>;

<https://www.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/news/immersive-technology-report-17-19/>; <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-to-regulate/>;

<https://www.rsph.org.uk/our-work/policy/wellbeing/new-filters.html>

⁸ <http://www.iicom.org/images/iic/themes/news/Reports/French-social-media-framework---May-2019.pdf>

outdated both as regards to technologies and services available and the problems faced, we proposed an overarching duty on operators to ensure, so far as possible, that their services were 'safe by design'. Borrowing from the tort of negligence the concept of a duty of care (which as a private law tool has an analogue in many countries), the Carnegie proposal suggested a statutory duty of care that would set down this general obligation to take *reasonable* steps to address *foreseeable* harm. Note, it is not expected that the duty of care will lead to a perfect environment – it cannot solve all problems on the Internet. It may improve the general environment so as to allow more targeted, content focused measures if needed; it can therefore be seen as working in tandem with rules aimed at improving notice and action requirements in relation to specific categories of speech. In this context, we should note also the importance of data protection rules.

7. In general, the systems-based approach is neutral as to the topics of content. Moreover, most interventions allow speech to continue, but affect its visibility (e.g changes to recommender algorithm/ autoplay switched off), velocity of spread (number of people to whom one message may be forwarded) and – perhaps – manner of expression (reminders as to rules relating to harassment and hate speech). Such interventions are less intrusive as regards freedom of speech.
8. The obligation has, in essence, four aspects:
 - the overarching obligation to exercise care in relation to user harm;
 - risk assessment process
 - establishment of mitigating measures; and
 - ongoing assessment of the effectiveness of the measures.
9. The regime envisages an independent regulator with a double-role:
 - informing and facilitating good practice (e.g. through the drafting of codes or guidance); and
 - verification of compliance of the operators with the duty and, where necessary, enforcement.
10. Enforcement action should be context specific and proportionate, especially given the fundamental rights in play (including but not limited to freedom of expression).
11. While the proposal envisaged that the underpinning statute should set out the types of harm, this does not take away from the fact that this is a general duty. The generality is important for two reasons. First, it allows the regime to develop as technology does, as services and the market change and as understanding of risk and harm increases. It is an element of future-proofing the regime – harm is consistent while the technical state of the art advances. Secondly, the general duty allows operators to take into account their respective services and the risk that those services pose to the sorts of user the services have. It also allows the platform operators to bring their technical and service knowledge into the regime. Finally, the fact that there is a general obligation does not mean that statute cannot specify specific obligations within the general duty – for example, the need to have an effective complaints

mechanism, obligations of transparency for particular issues, the need to take particular steps with regard to specific types of content (e.g. child sexual abuse and exploitation material). The general obligation acts as a form of basket for any such specific obligation, giving coherence and structure to the regime.

12. We note that the Australian proposals set out a set of “Basic Online Safety Expectations” which service providers will need to take into account when designing and updating their services and processes, covering children’s access to services and reporting and complaint mechanisms, and gives the e-Safety Commissioner greater powers in relation to transparency reporting directions. These and many other features are necessary components of the systemic approach that we have described above. Effective moderation and requirements for takedown of content that is harmful are also necessary features of an effective regulatory regime; however, we would see the most effective action to limit the broadest range of harms happening further upstream, before that content is widely viewed and disseminated and before the harm has been caused.
13. We look forward to following the progress of the Bill as it is introduced into the Australian Parliament for debate and we would be happy to provide further information or discuss this work with officials, if helpful.

February 2021



Attachment: Carnegie UK Trust Draft Online Harm Reduction Bill