CENTER FOR DEMOCRACY & TECHNOLOGY

Department of Culture, Media, and Sport Consultation on Online Harms

Comments of the Center for Democracy & Technology

1 July 2019

The Center for Democracy & Technology (CDT) respectfully submits the following comments to the Department of Culture, Media, and Sport's consultation on Online Harms. CDT is a non-profit public interest advocacy organization dedicated to advancing human rights and civil liberties in Internet and technology law and policy. CDT has offices in Washington, DC and Brussels and regularly engages in policy advocacy in the US and Europe concerning freedom of expression online, intermediary liability laws, corporate transparency and accountability, and communications privacy. We have previously engaged in consultations in the United Kingdom concerning the Investigatory Powers bill and participated as intervenors in the case of Big Brother Watch vs. the United Kingdom before the European Court of Human Rights.

Below, we provide responses to the specific consultation questions. As a threshold matter, however, we emphasize that the "duty of care" concept advanced in the white paper, and in particular the contemplation of legal duties for intermediaries to police speech that is "legal but harmful", pose significant risks to individuals' fundamental rights to freedom of expression and access to information.

Article 10 of the European Convention on Human Rights requires that limitations on freedom of expression be prescribed by law,[1] which "means that the law must be accessible, clear and sufficiently precise to enable individuals to regulate their behaviour."[2] Limitations or restrictions on expression that are aimed at intermediaries nevertheless implicate the fundamental rights of individuals to speak and to access information.[3] The white paper proposes an extremely broad concept of "online harms"[4] and describes a regulatory approach in which a regulator, not Parliament, articulates standards for restricting information that intermediaries must follow.[5] Intermediaries would face a variety of proposed penalties (as determined by the regulator, not a judge) for failing to meet these standards, which would create significant incentives for intermediaries to restrict content more broadly on their services, out of an abundance of caution. The lawful expression and access to information of individuals around the world would be the collateral damage of such a regulatory approach.

We welcome the white paper's proposal that a regulator would "have an obligation to protect users' rights online, particularly rights to privacy and freedom of expression. It will ensure that the new

[1] European Convention on Human Rights, Article 10.
[2] Council of Europe, Freedom of Expression and Information, Explanatory Memorandum, https://www.coe.int/en/web/freedom-expression/freedom-of-expression-and-information-explanatory-memo.
[3] Cengiz and Others v. Turkey, para. 55 (Eur. Ct. H.R. 2015).
[4] *See* Graham Smith, "The Rule of Law and the Online Harms White Paper", 5 May 2019 (discussing the white paper's expansion in scope from "harms to individuals" to "harms to society") https://www.cyberleagle.com/2019/05/the-rule-of-law-and-online-harms-white.html.
[5] HM Government, Online Harms White Paper, p. 64 para. 7.4 (April 2019).

regulatory requirements do not lead to a disproportionately risk-averse response from companies that unduly limits freedom of expression, including by limiting participation in public debate."[6] We respectfully submit, however, that the government will need to narrow significantly the scope of such regulator's activity and to ensure that any limitations on freedom of expression meet the standards of legality, necessity, and proportionality, in order to meet these goals.[7]

* * *

Consultation Responses:

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?

The most critical component of transparency reporting and building a culture of transparency is ensuring that it is *meaningful*. Data simply for the sake of data is neither meaningful nor helpful. For example, numbers without context, examples, and anecdotes provide little insight into the policies and practices that impact internet users. Any transparency reporting that the government promotes should hew closely to guidelines for meaningful transparency.

Engagement with external stakeholders will be critical to achieving this goal. Members of civil society, academia, and the tech sector have spent extensive time studying the challenges of reporting as well as the impact and accessibility of reports.[8] Their expertise will be invaluable in ensuring that whatever efforts the government engages in take into account the needs of those who read the reports as well as the potential impacts and concerns with aggregation and publication of disparate data from different companies.

A natural model for reporting around content moderation practices will be existing transparency reports around government requests for data. Reporting on this data has become an industry standard for social media and tech platforms. While the practice was in place by one company (Google) prior to the Edward Snowden leaks of 2013, its popularity has skyrocketed since then. The intervening years have seen a number of iterations in reporting, with companies working in conjunction with outside stakeholders to understand the best way to report. Reporting around content moderation practices is in

[6] *Id.* at p. 56 para. 5.12.

[7] We also note that the process of implementing the Audio-Visual Media Services Directive, which is already underway, will provide an important opportunity for working through the legal and practical challenges of engaging in regulatory oversight of Internet intermediaries and their handling of third-party content.

[8] New America, the Open Technology Institute, and the Berkman Klein Center at Harvard University; The Transparency Reporting Toolkit - Guide and Template https://www.newamerica.org/oti/policy-papers/transparency-reporting-toolkit-reporting-guide-and-template/.

a nascent stage and necessarily should take a similar path, requiring consultation and consideration of what approaches, types of data, information, and context are the best.

Beyond the practice of transparency reporting about companies' specific content moderation policies, there are other facets of online content practices that would benefit from greater insight. Engagement between companies and government bodies on content moderation and removal is effectively a black hole of information. The UK's Counter-Terrorism Internet Referral Unit (CTIRU), for example, would offer one starting point for exploring what meaningful transparency around these engagements might look like. At the same time, the government can contribute to building a culture of transparency by ensuring that its agencies operate in a transparent manner. CTIRU flags online content related to terrorism that it considers in violation of terms of service, but not necessarily illegal. There have been repeated calls[9] for improving the transparency of CTIRU's operations in order to make sure that content notifications are justified and proportionate.

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

The White Paper does not explain for what purposes "super complaints" would be used. On this basis, CDT does not see a justification for this concept. It is unclear whether "super complaints" would be related to concerns over companies' handling of allegedly illegal content, enforcement of their own content policies, or potential obligations under the code of practice. If they are related to companies' enforcement of their own content policies, and not to companies' obligations under the code of practice, then a "super complaints" procedure could create significant confusion about the role and scope of the regulator and the obligations under the code of practice. Moreover, if the "super complaints" procedure is envisioned as an element of user redress in situations where the company has enforced its own content policy, it is highly unlikely that the regulator would be able to address users' appeals of specific enforcement decisions at a scale or frequency that would have a meaningful impact on user redress overall. This is a similar concern that has been raised about, for example, the Facebook Oversight Board or "Supreme Court" concept.

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?

Safeguards against abusive notices are a key component of any system that enables notification or flagging of content that may be in violation of a rule or law. As has been demonstrated countless times

[9] Jim Killock, Informal Internet Censorship: The Counter Terrorism Internet Referral Unit (CTIRU), https://www.openrightsgroup.org/blog/2019/informal-internet-censorship-the-counter-terrorism-internet-referral-unit.

across social media websites, online fora, and other interactive online spaces, any content flagging system will be vulnerable to abuse by bad actors who use the system to try to target for removal speech they disagree with.

Especially in the case of systems for flagging potential violations of a code of practice, which will raise for the company the prospect of significant fines, there is a risk of overbroad removal in response to flags or notifications. Companies will be less likely to reject suspect notifications if there is a legal risk to them to do so; this means that the provisions around submitting and receiving notification need to include clearly specified components of a notice and a clear description of the legal consequence of sending, receiving, and rejecting notices. If the regulator is the one receiving the notifications, it too will need to be aware of the risk of abusive notices and will need to develop procedures for identifying and rejecting improper notifications, particularly if it ends up receiving large quantities of notices.

One option would be to include a penalty or liability that could be invoked against the notifier for sending fraudulent or bad-faith notices. An example of this is found in the Digital Millennium Copyright Act in the United States, which requires notifiers to affirm under the penalty of perjury that they are authorized to make a claim by the rightsholder of the copyrighted work in question. This provides a legal basis for cases challenging fraudulent, unauthorized, and otherwise overbroad takedown notices.

Moreover, the government should not create legal obligations on companies based on notifications about lawful content that they host. Article 10 of the Convention on Human Rights establishes that all restrictions of speech must be prescribed by law, have a legitimate aim, and be necessary for a democratic society. The white paper proposes to enable an independent body to punish a company that fails to eliminate *lawful content* with potentially harmful effects (*see e.g.* page 7, para. 14; page 42, para. 3.5; page 17, box 6). Punishing companies for hosting legal content restricts freedom of expression with no legal basis and would require the deletion of large amounts of content with pervasive collateral effects and without specifying any precise system to prevent abuses. Restrictions to access legal content also infringe users' right to access information, also enshrined in Article 10.

The Government should ensure that any policy initiatives it takes are consistent with the ECHR and respect the rights of those whose speech is targeted by notifications.

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

Parliament must ensure that any regulatory activity does not interfere disproportionately with people's access to information or freedom of expression. Periodic review of the proposed regulator's activities would need to consider carefully whether the regulator has met its objective to promote a 'free, open and secure' internet, and 'protect freedom of expression online', in addition to the other objectives it would be charged with furthering.

If the regulator considers that new types of content and behaviour cause actual and demonstrable harm on a significant scale, and those types of content and behaviour are not proscribed by legislation, Parliament would need to legislate for it. That means that Parliament must describe in terms that are clear, understandable and meaningful, the type of behaviour or content it wishes to intervene against. It will also need to demonstrate how such interference with free expression rights is necessary and proportionate and meets international Human Rights standards. These decisions cannot be left to an administrative body.

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?

The White Paper defines the proposed scope of regulation to include any service that "allows users to share or discover user-generated content or interact with each other online." It argues that this broad scope is necessary in order to capture the many different types of services that people use to connect and share information. The definition also possibly covers services and companies that have no relationship with end-users. This includes security, infrastructure, and payment providers, for example. This is counter-productive. The definition should be amended to exclude any company that does not have a direct relationship with users who provide content that may be disputed. These companies are not able to decide on notifications of discrete pieces of content. The Government should avoid creating responsibilities and liabilities for companies providing the services used by companies that host third-party content.

The scope does not include any limits as concerns size. It is argued that a "comprehensive approach is important for the efficacy of the new regulatory framework". The scope will include a global social media network as well as a privately run blog with a comment section. The Government should recognise that such a broad approach will inevitably create problems for start-ups and entrepreneurs in the internet sector. A broad scope and vaguely defined responsibilities will create legal and compliance risks that will raise costs and deter venture capital early phases of operation. A recent study by the Copia Institute noted the strong correlation between protections from liability for Internet intermediaries and rates of investment in start-ups and innovative online services.[10]

It is important that the White Paper recognises the need for the regulator to "take a risk-based and proportionate approach" and "focus on those companies that pose the biggest and clearest risk of harm to users, either because of the scale of the platforms or because of known issues with serious harms". Prioritisation based on risk would enable the regulator to develop the necessary expertise and to focus its resources most effectively. But it will be crucial for all determinations of the scope of covered

[10] Michael Masnick, Don't Shoot the Message Board (June 2019) https://copia.is/wp-content/uploads/2019/06/DSTMB-Copia.pdf.

companies to be supported by clear and justifiable evidence of risk, including an examination of whether size/scale of user base alone presents a heightened risk or whether risk derives from other factors.

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

Given the importance of respecting the privacy of individuals' communications, the government should treat as private (and therefore out of scope of the regulation) any communication where the person has taken any steps to make their communication non-public. This includes one-to-one conversations as well as communications made available only to groups, friends, followers on social media networks. This also includes encrypted communications.

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

Any private channel should be exempted from the regulation.

Question 7a: What specific requirements might be appropriate to apply to private channels and forums in order to tackle online harms?

We agree with the White Paper that companies should not have a legal duty to police, monitor, or scan users' private communications. Long-standing principles of the privacy and confidentiality of private communications should be maintained. In the United States, companies are permitted to view users' private communications only in limited circumstances and may (but are not obliged to) report information to law enforcement if they inadvertently encounter evidence of the potential commission of a crime or of an emergency involving the risk of serious physical harm or death.[11] In case of suspicion of distribution of illegal content or evidence of a crime, law enforcement authorities can obtain the necessary authorisations to access communications data and devices to carry out investigations. The government should not pursue any mandates on companies to report potential illegal activity/content to law enforcement because such a mandate would circumvent privacy protections built into existing legal processes for compelling the disclosure of such information and because such a mandate would result in over-reporting. To promote the security and fidelity of private communications, the government should encourage companies to develop and deploy strong encryption.

[11] See 18 U.S.C. 2702, *available at* https://www.law.cornell.edu/uscode/text/18/2702.

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

The regulator should be required to publish comprehensive transparency reports accessible to both Parliament, the public and human rights organisations. As mentioned, the regulator should be required to demonstrate how it has met its obligations to protect free expression and promote an open, innovative internet ecosystem. It would be important not to measure the regulator's effectiveness on how much content it has succeeded in suppressing or how many fines it has issued.

As with transparency reporting around companies' content moderation practices, reporting by the regulator should not simply offer up data for the sake of transparency. Rather, the regulator should engage with outside stakeholders to understand what meaningful transparency around its practices would look like. Such engagement would allow for consideration of the type of information necessary to enable external accountability as well as the unintended consequences of reporting. These engagements can also help to better understand how reports are read and who reads them, as well as the type of information and context that will be necessary to make such reporting more than a rote exercise in transparency.

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

CDT does not take a position on this question. Regardless of how regulatory oversight is put in place, it would be essential to ensure that the regulator has the technical expertise to understand online content moderation systems, and the necessary legal knowledge of international human rights and free expression standards the regulator's activities will need to meet.

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?

These proposed powers would not only be punitive for the company but would also have significant consequences for the fundamental rights of third parties; such measures should only be possible further to a decision by an independent judicial authority. These proposed measures are likely to have a dramatic impact on people's ability to access and share lawful content.

In the case of ISP blocking, the European Court of Human Rights has ruled (Cengiz v. Turkey) that such measures render large quantities of information inaccessible and affect the rights of many users. UK law already provides for ISP blocking (for example, for copyright infringement), and evidence shows that nearly 40% of such blocking orders are erroneous.[12] The White Paper proposes administrative blocking, without court review. This is likely to lead to exacerbating these problems. The White Paper does not put forward evidence to suggest that there is a need for expanding the use of ISP blocking.

Other types of 'disruptions of business' would have equally problematic consequences, and should not be used unless ordered by an independent court. CDT does not consider that there is a justification for introducing new liabilities for company managers.

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?

Both the GDPR and draft EU legislation on access to electronic evidence include provisions requiring companies to appoint legal representatives if they meet certain thresholds according to which they are deemed to provide services in a country. The concerns discussed in the white paper about illegal and harmful content typically involve large social media platforms. These companies already have business representatives in the UK. It would not seem justified or necessary to introduce new requirements for the purposes discussed in the White Paper.

Furthermore, there is a risk that such a step would legitimise authoritarian regimes with poor records of respecting human rights and the rule of law to take similar initiatives. It is important to note that in such states, people benefit from information that state-controlled media cannot carry. In countries where the state runs strict political censorship, dissidents and others rely on internet platforms to access information that governments would otherwise stop them from. If those regimes were to mandate local representatives, legitimised by UK policy, it would have serious repercussions for access to information and free expression. For these reasons, CDT recommends that the UK government refrains from introducing requirements of the sort discussed in the White Paper.

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?

[12] Open Rights Group, Internet Regulation Part 1: Internet Censorship in the UK Today (December 2018), https://www.openrightsgroup.org/assets/files/pdfs/reports/Internet_Regulation_Part_I_Internet_Censorship_in_the_UK_today-web.pdf.

It is not clear that there are barriers to the deployment of safety technologies. The industry is increasingly using technical tools, such as filters, to flag potentially problematic content and prevent uploads where possible. Even the most advanced companies are far from developing solutions that can understand and interpret complex and nuanced speech. For example, technical solutions cannot determine when speech is intended to, e.g., incite hatred, or is meant to satirise or ridicule those who traffic in hateful content.[13] There are multiple examples of platforms taking down content completely legitimate speech, having mistaken it for violating the site's terms of service.

The government should be conscious that efforts to accelerate the use of filtering technology by platforms is likely to suppress broad categories of legitimate and lawful speech. It should also bear in mind that mandating use of filtering technologies is likely to have disproportionate cost impact on start-up companies. Instead, the government should focus on supporting the development of user-empowerment tools and safety technologies that give parents and other users the ability to set their own parameters and standards for what online content they access.

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

The Regulator, if it has the necessary expertise, could help create awareness and digital literacy, working with schools, companies and civil society. It could also take an active role in helping users develop sound practices in their online activities. If necessary, it can provide materials about the boundaries between lawful and illegal speech, and it can emphasise that users are accountable for their behaviour online in so far as it violates the law. A better appreciation of the limits of the law can also help people understand that being exposed to points of view and statements one finds offensive does not mean that norms, let alone laws have been broken. Offensive, outrageous and provocative statements are explicitly protected under human rights law, offline and online.

[13] Center for Democracy & Technology, Mixed Messages? The Limits of Automated Social Media Content Analysis (November 2017), https://cdt.org/insight/mixed-messages-the-limits-of-automated-social-media-content-analysis/.