

Index on Censorship submission to Online Harms White Paper consultation

19 June 2019

Index on Censorship is a registered charity and company that campaigns for freedom of expression worldwide. We publish work by censored writers and artists, promote debate, and monitor threats to free speech, and have more than 45 years' experience in this field.

Defending online freedom of expression is a core part of our work and Index has contributed to a number of consultations on this issue, including the provision of oral evidence to the current [Joint Human Rights Committee inquiry](#) into Democracy, free speech and freedom of association. Index is a council member of the global freedom of expression network [IFEX](#), a member of the multi-stakeholder [Global Network Initiative](#) and a member of [European Digital Rights](#).

While we recognise that the internet has provided new opportunities for some traditionally marginalised groups to speak out, we are increasingly concerned¹ about the way in which it is being used by governments and companies to silence legal speech. Index campaigns against restrictions of freedom of speech online, [including the growing use and abuse of national security legislation to silence government critics](#).

Index has played an active role in discussions related to the government's internet safety strategy. We note that the questions posed in this consultation do not address some of the key concerns voiced by freedom of expression and digital rights organisations regarding the Online Harms white paper, outlined in our recent briefing paper (attached).

We therefore focus below on the consultation questions that are most directly related to freedom of expression concerns and would urge the government to take into account broader concerns as it further develops this policy.

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?

As set out earlier in [a joint statement](#) with Open Rights Group and Global Partners Digital, a culture of transparency, trust and accountability is key to establishing good governance globally online.

Index places particular stress on the central importance of improved transparency from social media companies in which companies should have: clear terms of service, clear processes for assessing violations of terms of service and clear appeals procedures.

1

<https://www.indexoncensorship.org/2019/04/joy-hyvarinen-on-the-uk-governments-online-harms-white-paper-bbc-today-8-4-2019/>



There are a number of mechanisms by which these measures could be independently overseen, such as self-regulatory [Social Media Councils](#)² as proposed by ARTICLE 19 and whose efficacy is considered in a paper produced by the [Electronic Frontier Foundation](#)³, or an independent regulatory body mandated by parliament.

In all cases, however, the oversight body should be focused on regulating *process* rather than regulating *content*. As set out below, the current proposals risk effectively outlawing legal speech online. Rather than bringing laws that apply in the offline world into line with the online world, this would create a two-track system in which speech between individuals was subject to far stricter curbs online than off, thereby undermining the freedoms protected under Article 19 of the United Nations Declaration on Human Rights⁴ and Article 10 of the European Convention on Human Rights⁵.

The government already appears settled on a regulator. However, as set out in the attached briefing paper, Index is concerned the government has rushed to a solution that is ill-fitted to its ostensible purpose, and that - furthermore - the government has failed to establish clearly the parameters of the problem it is trying to solve.

We therefore urge the following:

1. Establish a clear evidence base

The wide range of different harms that the government is seeking to tackle in this policy process require different, tailored responses. Measures proposed must be underpinned by strong evidence, both of the likely scale of the “harm” and the measures’ likely effectiveness. This is particularly important as there remains considerable debate about the level of harm caused in general by internet use⁶. The evidence that formed the base of the Internet Safety Strategy Green Paper was highly variable in its quality. Any legislative or regulatory measures should be supported by clear and unambiguous evidence of their need and effectiveness.

Index urges the government to gather a more robust data on the scale of online harms.

2. Define “harm” precisely and in a manner consistent with Article 10 protections for speech

Index is concerned at the use of a duty of care regulatory approach. Although social media has often been compared the public square, the duty of care model is not an exact fit because this would introduce regulation – and restriction – of speech **between individuals** based on criteria that is **far broader than current law**.

² <https://www.article19.org/resources/self-regulation-hate-speech-social-media-platforms/>

³

<https://www.eff.org/deeplinks/2019/05/social-media-councils-better-way-firward-lipstick-pig-or-global-speech-police>

⁴ UN General Assembly, Universal Declaration of Human Rights, 10 December 1948, 217 A (III).

⁵ Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, 4 November 1950, ETS 5.

⁶ For example, a report published in January 2019 by The Royal College of Paediatrics and Child Health, which oversees the training of specialists in child medicine, said there was no good evidence that time in front of a screen is “toxic” to health, as is sometimes claimed: <https://www.bbc.co.uk/news/health-46749232>

A failure to accurately define “harmful” content risks incorporating legal speech, including political expression, expressions of religious views, expressions of sexuality and gender, and expression advocating on behalf of minority groups that are fundamental to effective democratic functions. Without precise definitions, and in conjunction with proposed sanctions and liability regimes (discussed below) the likelihood is that companies will resort to using algorithms to identify and eliminate harmful content. As recent cases have shown it is in practice extremely difficult for artificial intelligence to differentiate between many types of content. For example, vital documentation of war crimes by activists may be removed as 'hate speech', while anti-racist campaigners have found their sites removed for having used racial slurs as evidence of racism. Recent changes in YouTube's hate speech policy, which resulted in thousands of academic, journalistic and activist sites being removed is a powerful [example](#) of the risks inherent in a broad-brush approach policy that targets a wide range of ill-defined content.

3. De-link liability/sanctions to platforms over third party content

While well-meaning, proposals that combine platform liability with sanctions for third party content (such as the mention in the Online Harms white paper of possibly increasing the current liability basis as set out in the e-Commerce Directive) contain serious risks, such as requiring or incentivising wide-sweeping removal of lawful and innocuous content. The imposition of time limits for removal, heavy sanctions for non-compliance or incentives to use automated content moderation processes only heighten this risk, as has been evidenced by the approach taken in Germany via its Network Enforcement Act (or NetzDG).

Index has reported on various examples of the German law being applied incorrectly, including the [removal of a tweet](#) of journalist Martin Eimermacher criticising the double standards of tabloid newspaper Bild Zeitung and the [blocking of the Twitter account](#) of German satirical magazine Titanic.

Index recommends that the duty of care should be limited and defined in a way that addresses the risk that it will create a strong incentive for companies and others to censor legal content, especially if combined with fines and personal liability for senior managers.

4. Include explicit protections for freedom of expression

In addition, Index recommends that the government considers the call by United Nations Special Rapporteur David Kaye in [his 2018 report](#) for a human rights-based approach.

5. Adopt a genuinely multi-stakeholder approach to development of policy

In recognition of the UK's commitment to the multistakeholder model of internet governance, we hope all relevant stakeholders, including civil society experts on digital rights and freedom of expression, will be fully engaged throughout the development of the Online Harms bill.

A multistakeholder approach should also include consulting closely with media industry and media freedom organisations and providing further clarification regarding the relationship between online harms regulation and press freedom.

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?

As noted above, it is extremely important that a) Terms of Service be clear as to what a platform determines to be 'harmful' under its own terms of service; b) that there be a clear and easy-to-use mechanism for reporting such content; c) that there be a clear appeals process for content removed unfairly. As outlined below, the regulator should not outlaw content beyond that which is already illegal. A situation where activities and materials that are legal offline effectively became illegal online would create an unacceptable double standard. Much greater transparency must be at the centre of attempts to regulate online content.

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 duty of care would cover companies of all sizes, social media companies, public discussion forums, retailers that allow users to review products online, non-profit organisations (for example, Index on Censorship), file sharing sites and cloud hosting providers. This is too wide and would be very challenging to implement in practice. Index recommends a much narrower scope for any proposals, which would focus on the largest online platforms. We recommend an assessment of user numbers across major platforms as a basis for establishing a threshold above which online platforms would be within scope.

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

Index believes that private communications should not be in scope. Private channels are essential means for freedom of expression, including enabling campaigners and activists to make their voices heard.

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?

As outlined above, Index is extremely concerned about increasing liability in combination with possible personal liability for senior managers, possibly even including criminal liability, and the incentive this would create to remove and restrict user-generated content. The powers to disrupt business activities and undertake ISP blocking are far-reaching measures, which Index does not believe should be introduced because of their potential impacts on freedom of expression.

Question 16: What, if any, are the most significant areas in which organisations need practical guidance to build products that are safe by design?

Organisations need guidance on freedom of expression and how to implement freedom of expression considerations into product design from the earliest stage. Any future regulator should ensure that detailed guidance is produced to help ensure that freedom of expression impacts are integrated from the first step of design and planning.

Question 17: Should the government be doing more to help people manage their own and their children's online safety and, if so, what?

Better guidance regarding freedom of expression, including freedom of expression for children and young people, is needed. Index recommends that this is made a higher priority in the education system and training programmes for young people.

APPENDIX 1: The UK government's online harms white paper: implications for freedom of expression

Briefing paper 3 June 2019

Recommendations

- Parliament must be fully involved in shaping the government's proposals for online regulation as the proposals have the potential to cause large-scale impacts on freedom of expression and other rights.
- The proposed duty of care needs to be limited and defined in a way that addresses the risk that it will create a strong incentive for companies and others to censor legal content, especially if combined with fines and personal liability for senior managers.
- It is important to widen the focus from harms and what individual users do online to the structural and systemic issues in the architecture of the online world. For example, much greater transparency is needed about how algorithms influence what a user sees.
- The government is aiming to work with other countries to build international consensus behind the proposals in the white paper. This makes it particularly important that the UK's plans for online regulation meet international human rights standards. Parliament should ensure that the proposals are scrutinised for compatibility with the UK's international obligations.
- More scrutiny is needed regarding the implications of the proposals for media freedom, as "harmful" news stories risk being caught.

Introduction

The proposals in the government's online harms white paper risk damaging freedom of expression in the UK, and abroad if other countries follow the UK's example.

- A proposed new statutory duty of care to tackle online "harms" combined with substantial fines and possibly even personal criminal liability for senior managers would create a strong incentive for companies to remove content.
- The "harms" are not clearly defined but include activities and materials that are legal.
- Even the smallest companies and non-profit organisations are covered, as are public discussion forums and file sharing sites.

The proposals come less than two months after the widely criticised Counter-Terrorism and Border Security Act 2019. The act contains severe limitations on freedom of expression and access to information online (see Index report for more information).

The duty of care: a strong incentive to censor online content

The proposed new statutory duty of care to tackle online harms, combined with the possibility of substantial fines and possibly even personal criminal liability for senior managers, risks creating a strong incentive to restrict and remove online content.

Will Perrin and Lorna Woods, who have developed the online duty of care concept, envisage that the duty will be implemented by applying the “precautionary principle” which would allow a future regulator to “act on emerging evidence”.

Guidance by the UK Interdepartmental Liaison Group on Risk Assessment (UK-ILGRA) states:

The purpose of the Precautionary Principle is to create an impetus to take a decision notwithstanding scientific uncertainty about the nature and extent of the risk, i.e. to avoid 'paralysis by analysis' by removing excuses for inaction on the grounds of scientific uncertainty.

The guidance makes sense when addressing issues such as environmental pollution, but applying it in a context where freedom of expression is at stake risks legitimising censorship - a very dangerous step to take.

Not just large companies

The duty of care would cover companies of all sizes, social media companies, public discussion forums, retailers that allow users to review products online, non-profit organisations (for example, Index on Censorship), file sharing sites and cloud hosting providers. A blog and comments would be included, as would shared Google documents.

The proposed new regulator is supposed to take a “proportionate” approach, which would take into account companies’ size and capacity, but it is unclear what this would mean in practice.

Censoring legal “harms”

The white paper lists a wide range of harms, for example, terrorist content, extremist content, child sexual exploitation, organised immigration crime, modern slavery, content illegally uploaded from prisons, cyberbullying, disinformation, coercive behaviour, intimidation, under 18s using dating apps and excessive screen time.

The harms are divided into three groups: harms with a clear definition; harms with a less clear definition; and underage exposure to legal content. Activities and materials that are not illegal are explicitly included. This would create a double standard, where activities and materials that are legal offline would effectively become illegal online.

The focus on the catch-all term of “harms” tends to oversimplify the issues. For example, the recent study by Ofcom and the Information Commissioner’s Office Online Nation found that 61% of adults had a potentially harmful experience online in the last 12 months. However, this included “mildly annoying” experiences. Not all harms need a legislative response.

A new regulator

The white paper proposes the establishment of an independent regulator for online safety, which could be a new or existing body. It mentions the possibility of an existing regulator, possibly Ofcom, taking on the role for an interim period to allow time to establish a new regulatory body.

The future regulator would have a daunting task. It would include defining what companies (and presumably also others covered by the proposed duty of care) would need to do to fulfil the duty of care, establishing a “transparency, trust and accountability framework” to assess compliance and taking enforcement action as needed.

The regulator would be expected to develop codes of practice setting out in detail what companies need to do to fulfil the duty of care. If a company chose not to follow a particular code it would need to justify how its own approach meets the same standard as the code. The government would have



the power to direct the regulator in relation to codes of practice on terrorist content and child sexual exploitation and abuse.

Enforcement

The new enforcement powers outlined in the white paper will include substantial fines. The government is inviting consultation responses on a list of possible further enforcement measures. These include disruption of business activities (for example, forcing third-party companies to withdraw services), ISP blocking (making a platform inaccessible from the UK) and creating a new liability for individual senior managers, which could involve personal liability for civil fines or could even extend to criminal liability.

Undermining media freedom

The proposals in the white paper pose a serious risk to media freedom. Culture Secretary Jeremy Wright has written to the Society of Editors in response to concerns, but many remain unconvinced.

As noted the proposed duty of care would cover a very broad range of “harms”, including disinformation and violent content. In combination with fines and potentially even personal criminal liability, this would create a strong incentive for platforms to remove content proactively, including news that might be considered “harmful”.

Index has filed an official alert about the threat to media freedom with the Council of Europe’s Platform to promote the protection of journalism and safety of journalists. Index and the Association of European Journalists (AEJ) have made a statement about the lack of detail in the UK’s reply to the alert. At the time of writing the UK has not provided a more detailed reply.

Censorship and monitoring

The European Union’s e-commerce directive is the basis for the current liability rules related to online content. The directive shields online platforms from liability for illegal content that users upload unless the platform is aware of the content. The directive also prohibits general monitoring of what people upload or transmit.

The white paper states that the government’s aim is to increase this responsibility and that the government will introduce specific monitoring requirements for some categories of illegal content. This gets close to dangerous censorship territory and it is doubtful if it could be compatible with the e-commerce directive.

Restrictions on freedom of expression and access to information are extremely serious measures and should be backed by strong evidence that they are necessary and will serve an important purpose. Under international law freedom of expression can only be restricted in certain limited circumstances for specific reasons. It is far from clear that the proposals set out in the white paper would meet international standards.

Freedom of expression - not a high priority

The white paper gives far too little attention to freedom of expression. The proposed regulator would have a specific legal obligation to pay due regard to innovation. When it comes to freedom of expression the paper only refers to an obligation to protect users’ rights “particularly rights to privacy and freedom of expression”.

It is surprising and disappointing that the white paper, which sets out measures with far-reaching potential to interfere with freedom of expression, does not contain a strong and unambiguous commitment to safeguarding this right.