HOUSE OF COMMONS LIBRARY

BRIEFING PAPER

Number 8743, 26 February 2020

Social media regulation

By John Woodhouse

Contents:

Contents

Summary

There is increasing concern about harmful content and activity on social media. This includes cyberbullying, the intimidation of public figures, disinformation, material promoting violence and self-harm, and age inappropriate content.

Critics, including parliamentary committees, academics, and children's charities, have argued that self-regulation by social media companies is not enough to keep users safe and that statutory regulation should be introduced.

The Online Harms White Paper (April 2019) – a new regulatory framework?

An Online Harms White Paper was published in April 2019. This set out the then Government's approach for tackling "content or activity that harms individual users, particularly children, or threatens our way of life in the UK."

According to the White Paper, existing regulatory and voluntary initiatives had "not gone far or fast enough" to keep users safe. The Paper proposed a single regulatory framework to tackle a range of harms. At its core would be a statutory duty of care for internet companies, including social media platforms. An independent regulator would oversee and enforce compliance with the duty. A consultation on the proposals closed on 1 July 2019.

Reaction to the White Paper

The White Paper received a mixed reaction. Children's charities were positive. The NSPCC said that the Paper was a "hugely significant commitment" that could make the UK a "world pioneer in protecting children online". However, some commentators raised concerns that harms were insufficiently defined and that the Paper blurred the boundary between illegal and harmful content. The Open Rights Group and the Index on Censorship warned that the proposed framework could threaten freedom of expression.

Interim response to the White Paper consultation (February 2020)

On 12 February 2020, the Government published its initial response to the consultation on the White Paper. The response states, among other things, that the Government is minded to make Ofcom the regulator for online harms.

To protect freedom of expression, the new regulatory framework would not require companies in scope to remove specific pieces of legal content. Instead, they would be required to explicitly state what content and behaviour would be considered acceptable on their sites and to enforce this consistently and transparently. Illegal content would have to be removed "expeditiously". Robust action would be required in relation to terrorist content and child exploitation and abuse. The regulator would not investigate or adjudicate on individual complaints.

The proposed duty of care would only apply to companies that "provide services or use functionality on their websites which facilitate the sharing of user generated content or user interactions, for example through comments, forums or video sharing". According to the Government, fewer than 5% of UK businesses would be in scope.

The Government is still considering issues such as the enforcement powers of the regulator and senior management liability.

A final response is due in spring 2020. In a Commons debate on 13 February 2020, the DCMS said that legislation would be introduced in this parliamentary session.

1. Background

The criminal law applies to online activity in the same way as to offline activity. As the Government has noted, legislation passed before the digital age "has shown itself to be flexible and capable of catching and punishing offenders whether their crimes are committed by digital means or otherwise".[1] The Crown Prosecution Service has published guidance on offences on social media.

In addition, many regulators have a role in relation to certain types of online activity e.g. Ofcom, the Competition and Markets Authority, the Advertising Standards Authority, the Information Commissioner's Office, and the Financial Conduct Authority.[2]

The internet is therefore not quite an unregulated "Wild West" as some have claimed.[3]

However, there is no overall regulator and there is no specific content regulator.[4] Also, under the e-Commerce Directive, social media companies are exempt from liability for illegal content they host if they "play a neutral, merely technical and passive role" towards it. Once they become aware of illegal content, they must remove or disable access to it.

For content that is harmful or inappropriate, but not illegal, social media platforms self-regulate i.e. through "community standards" and "terms of use" that users agree to when joining a platform. This type of material can promote violence, self-harm, or cyberbullying. It can also include indecent, disturbing or misleading content.[5] The role of social media in enabling access to such material, and claims that self-regulation is an unsatisfactory response, has led to calls for statutory regulation and for the e-Commerce Directive to be revised or replaced.

1.1 Social media companies – increased liability?

The e-Commerce Directive makes provision about the liability of "information society services", a category that includes most internet service providers and online platforms.[6]

The Directive provides exemptions to liability for three types of online services - those that host (Article 14), cache (Article 13) or transmit (Article 12) content under certain circumstances.

---

1 HMG, *Internet Safety Strategy - Green paper*, October 2017, p37
2 House of Lords Select Committee on Communications, *Regulating in a digital world* HL Paper 299, March 2019, Appendix 4
3 House of Lords Select Committee on Communications, *Regulating in a digital world*, p9; Science and Technology Committee, *Impact of social media and screen-use on young people's health*, HC 822, January 2019, p52
4 House of Lords Select Committee on Communications, *Regulating in a digital world*, p3
5 Ibid, p48
6 Directive 2000/31/EC; In the UK, the *Electronic Commerce (EC Directive) Regulations 2002* implemented the Directive

Social media companies are generally regarded as information society services that host material, rather than create it. Under Article 14, they are not liable for content that they host, provided they do not have actual knowledge of illegal activity and act expeditiously to remove the information if they are informed that it is illegal.[7] Only services that "play a neutral, merely technical and passive role towards the hosted content are covered by the exemption".[8] This model of liability is commonly called the "notice and take-down" model.[9]

Under Article 15, Member States must not impose a general obligation on internet intermediaries to monitor or actively to seek facts or circumstances indicating illegal activity.

Criticism

The liability framework under the e-Commerce Directive has been criticised by different parliamentary committees. A March 2019 report by the Lords Select Committee on Communications noted that the Directive was nearly twenty years old and was developed before platforms began to curate content for users. It concluded, among other things, that notice and take-down was "not an adequate model for content regulation".[10]

In a July 2018 report, the Digital, Culture, Media and Sport Committee argued that social media companies could not "hide behind the claim of being merely a 'platform'" with no role in regulating the content of their sites":

> …they continually change what is and is not seen on their sites, based on algorithms and human intervention…[11]

A December 2017 report by the Committee on Standards in Public Life said that the Government should consider legislation to "rebalance" liability for online content after the UK had left the EU.[12]

The application of the Directive to online platforms has its defenders. An October 2017 paper, produced for the European Parliament's Internal Market and Consumer Protection Committee, argued that the current framework should be "maintained, since it is needed to ensure the

---

7 For discussion of the Directive see: House of Lords Select Committee on Communications, *Regulating in a digital world*, chapter 5; Lorna Woods, "When is Facebook liable for illegal content under the E-commerce Directive?", EU law analysis blog, 19 January 2017; Giovanni Sartor, *Providers Liability: From the eCommerce Directive to the future*, Report requested by the European Parliament's Internal Market and Consumer Protection Committee, October 2017

8 European Commission website, e-Commerce Directive [accessed 10 December 2019]

9 House of Lords Select Committee on Communications, *Regulating in a digital world*, chapter 5; Lorna Woods, "When is Facebook liable for illegal content under the E-commerce Directive?", EU law analysis blog, 19 January 2017

10 House of Lords Select Committee on Communications, *Regulating in a digital world*, p5 and para 193

11 Digital, Culture, Media and Sport Committee, *Disinformation and 'fake news': Interim Report*, HC 363 2017-19, July 2018, para 57

12 Committee on Standards in Public Life, *Intimidation in Public Life: A Review*, Cm 9543, December 2017, pp35-7

diverse provision of intermediation services and the freedoms of the users of such services".[13]

Graham Smith, an internet lawyer, has defended Article 15 of the Directive and its role in protecting freedom of expression:

> (…) [Article 15] prevents the state from turning internet gateways into checkpoints at which the flow of information could be filtered, controlled and blocked.
>
> The principle embodied in Article 15 is currently under pressure: from policymakers within and outside Brussels, from antagonistic business sectors, from the security establishment and potentially from all manner of speech prohibitionists. The common theme is that online intermediaries – ISPs, telecommunications operators, social media platforms - are gatekeepers who can and should be pressed into active service of the protagonists' various causes.
>
> Article 15 stands in the way of the blunt instrument of compulsory general monitoring and filtering. It does so not for the benefit of commercial platforms and ISPs, but to fulfil the policy aim of protecting the free flow of information and ultimately the freedom of speech of internet users…[14]

1.2 Social media companies - a duty of care?

The e-Commerce Directive does not prevent Member States from requiring service providers to apply duties of care "which can reasonably be expected from them and which are specified by national law, in order to detect and prevent certain types of illegal activities".[15]

Lorna Woods (Professor of Internet Law at the University of Essex) and William Perrin (Trustee of Carnegie UK Trust) have proposed a regulatory regime, centred on a new statutory duty of care, to reduce online harm. The regime, developed under the "aegis" of the Carnegie UK Trust, was put forward in a series of blog posts in 2018. A "refined" proposal was published in January 2019.[16] Woods and Perrin explain the duty as follows:

> Social media service providers should each be seen as responsible for a public space they have created, much as property owners or operators are in the physical world. Everything that happens on a social media service is a result of corporate decisions: about the terms of service, the software deployed and the resources put into enforcing the terms of service.
>
> In the physical world, Parliament has long imposed statutory duties of care upon property owners or occupiers in respect of people using their places, as well as on employers in respect of their employees. Variants of duties of care also exist in other sectors where harm can occur to users or the public. A statutory duty of care is simple, broadly based and largely future-proof. For instance, the duties of care in the 1974 Health and Safety at Work

---

13 Giovanni Sartor, *Providers Liability: From the eCommerce Directive to the future*, Report requested by the European Parliament's Internal Market and Consumer Protection Committee, October 2017, Abstract and Executive summary

14 Graham Smith, "Time to speak up for Article 15", Cyberlegal Blog, 21 May 2017

15 Recital 48 of Directive 2000/31/EC

16 Lorna Woods and William Perrin, *Internet harm reduction: an updated proposal*, Carnegie UK Trust, January 2019

> Act still work well today, enforced and with their application kept up to date by a competent regulator. A statutory duty of care focuses on the objective – harm reduction – and leaves the detail of the means to those best placed to come up with solutions in context: the companies who are subject to the duty of care. A statutory duty of care returns the cost of harms to those responsible for them, an application of the micro-economically efficient 'polluter pays' principle…
>
> Parliament should guide the regulator with a non-exclusive list of harms for it to focus upon. These should be: the stirring up offences including misogyny, harassment, economic harm, emotional harm, harms to national security, to the judicial process and to democracy…[17]

The regime would regulate services that:

- have a strong two-way or multiway communications component;
- display user-generated content publicly or to a large member/user audience or group.

It would cover "reasonably foreseeable harm that occurs to people who are users of a service and reasonably foreseeable harm to people who are not users of a service".[18]

Woods and Perrin argue that the regulator should be an existing one with experience of dealing with global companies (they suggest Ofcom). The regulator would use a harm reduction method, similar to that used for reducing pollution:

> [The regulator would] agree tests for harm, run the tests, the company responsible for harm invests to reduce the tested level, test again to see if investment has worked and repeat if necessary. If the level of harm does not fall or if a company does not co-operate then the regulator will have sanctions.
>
> In a model process, the regulator would work with civil society, users, victims and the companies to determine the tests and discuss both companies harm reduction plans and their outcomes. The regulator would have the power to request information from regulated companies as well as having its own research function. The nature of fast-moving online services is such that the regulator should deploy the UK government's formalised version of the precautionary principle, acting on emerging evidence rather than waiting years for full scientific certainty about services that have long since stopped…[19]

To make companies change their behaviour, Woods and Perrin suggest penalties of large fines, set as a proportion of turnover.[20]

Criticism

The duty of care proposed by Woods and Perrin has attracted criticism. Graham Smith, for example, has challenged the idea that social media

---

17 Lorna Woods and William Perrin, "Internet Harm Reduction: a Proposal", Carnegie UK Trust Blog, 30 January 2019
18 Ibid
19 Ibid
20 Penalties are issued on a similar basis under the General Data Protection Regulation and under the Competition Act

platforms should be viewed as having responsibilities for a public space, similar to property owners in the physical world:

> (…) The relationship between a social media platform and its users has some parallels with that between the occupier of a physical space and its visitors.
>
> A physical public place is not, however, a perfect analogy. Duties of care owed by physical occupiers relate to what is done, not said, on their premises. They concern personal injury and damage to property. Such safety-related duties of care are thus about those aspects of physical public spaces that are less like online platforms.
>
> That is not to say that there is no overlap. Some harms that result from online interaction can be fairly described as safety-related. Grooming is an obvious example. However that is not the case for all kinds of harm…[21]

He also raised concerns about the possible impact of the duty on freedom of expression:

> (…) We derive from the right of freedom of speech a set of principles that collide with the kind of actions that duties of care might require, such as monitoring and pre-emptive removal of content. The precautionary principle may have a place in preventing harm such as pollution, but when applied to speech it translates directly into prior restraint. The presumption against prior restraint refers not just to pre-publication censorship, but the principle that speech should stay available to the public until the merits of a complaint have been adjudicated by a legally competent independent tribunal. The fact that we are dealing with the internet does not negate the value of procedural protections for speech…[22]

Support

A February 2019 NSPCC report drew heavily on the work of Woods and Perrin and argued for a regulator to enforce a duty of care to protect children on social media.[23] According to the NSPCC, 90% of parents support making social networks legally responsible for protecting children.[24]

In a January 2109 report, the Science and Technology Committee noted the work of Woods and Perrin[25] and recommended, among others, that a duty of care should be introduced that would require social media companies "to act with reasonable care to avoid identified harms" to users aged under 18.[26] The duty would extend beyond that age for other vulnerable groups, as determined by the Government.

---

21 Graham Smith, "Take care with that social media duty of care", Cyberlegal Blog, 19 October 2018
22 Ibid
23 NSPCC, *Taming the Wild West Web: How to regulate social networks and keep children safe from abuse*, February 2019, p1 and chapter 3
24 "9 out of 10 parents back social network regulation", NSPCC News, 12 February 2019
25 House of Commons Science and Technology Committee, *Impact of social media and screen-use on young people's health*, paras 223-5
26 Ibid, para 228

Woods and Perrin submitted evidence to the Lords Select Committee on Communications during its inquiry into regulating the digital world.[27] The Committee's March 2019 report recommended that a duty of care should be imposed on online services hosting user-generated content. This would be enforced by Ofcom.[28]

In April 2019, after consulting on its internet safety strategy, the Government's Online Harms White Paper set out plans to introduce a statutory duty of care, to be overseen by an independent regulator.[29]

1.3 The Internet Safety Strategy

In October 2017, as part of its Digital Charter work, the Government published an Internet Safety Strategy green paper to help make "Britain the safest place in the world to be online."[30] Among other things, the paper looked at the role of social media companies in keeping users safe and sought views on a social media code of practice - section 103 of the *Digital Economy Act 2017* requires the Government to publish a voluntary code for social media platforms. The code would introduce minimum standards and ensure regular review and monitoring. The code has to address conduct that involves bullying, insulting, intimidating or humiliating behaviour. A consultation on the proposals closed in December 2017.

Government response

The Government's response was published in May 2018. This noted that a growing number of people, including parents and education and health professionals, were concerned about safety online. The consultation highlighted three main issues:

- Online behaviours too often fail to meet acceptable standards;
- Users can feel powerless to address these issues;
- Technology companies can operate without proper oversight, transparency or accountability, and commercial interests mean that they can fail to act in users' best interests.[31]

Various charities, including those representing children,[32] expressed concern about the existing self-regulatory approach and suggested making the social media code of practice legally binding, with an independent regulator and sanctions regime. However, the Internet Watch Foundation noted that self-regulation, in partnership with the internet industry, was "hugely effective" in removing online child abuse images. The industry companies that responded to the consultation also favoured continued self-regulation.[33]

The Government's view was that the "disconnect" between user and industry responses strongly suggested that companies needed to do

---

27 See House of Lords Select Committee on Communications, *Regulating in a digital world*, paras 198-202
28 Ibid, paras 205-6
29 HMG, *Online Harms White Paper*, April 2019
30 HMG, *Internet Safety Strategy - Green paper*, October 2017, p3
31 HMG, *Response to the Internet Safety Strategy Green Paper*, May 2018, p5
32 e.g. the NSPCC and the Children's Charities' Coalition on Internet Safety
33 *Response to the Internet Safety Strategy Green Paper*, May 2018, p9

more to manage content and behaviour on their platforms.[34] A white paper would be published, setting out areas for legislation, including making the social media code of practice legally binding.[35] The white paper would also look at increasing the liability of social media platforms for harmful and illegal content. The Government's response to the consultation claimed that the status quo was "increasingly unsustainable as it becomes clear many platforms are no longer just passive hosts."[36]

[34] Ibid, p13
[35] Ibid, p15
[36] Ibid, p14

2. Online Harms White Paper (April 2019)

The Online Harms White Paper was published in April 2019. It set out the then Government's approach to "tackle content or activity that harms individual users, particularly children, or threatens our way of life."[37]

According to the Government, the current "patchwork of regulation and voluntary initiatives" had not gone far or fast enough to keep UK users safe. The Paper therefore proposed a single regulatory framework to tackle a range of online harms.[38] The core of this would be a new statutory duty of care for internet companies, including social media platforms. An independent regulator would oversee and enforce compliance with the duty.

The Paper covered three categories of harms:

- harms with a clear definition;
- harms with a less clear definition;
- underage exposure to legal content.

Examples of harms in each category were set out in the following table:[39]

Harms with a clear definition	Harms with a less clear definition	Underage exposure to legal content
• Child sexual exploitation and abuse. • Terrorist content and activity. • Organised immigration crime. • Modern slavery. • Extreme pornography. • Revenge pornography. • Harassment and cyberstalking. • Hate crime. • Encouraging or assisting suicide. • Incitement of violence. • Sale of illegal goods/ services, such as drugs and weapons (on the open internet). • Content illegally uploaded from prisons. • Sexting of indecent images by under 18s (creating, possessing, copying or distributing indecent or sexual images of children and young people under the age of 18).	• Cyberbullying and trolling. • Extremist content and activity. • Coercive behaviour. • Intimidation. • Disinformation. • Violent content. • Advocacy of self-harm. • Promotion of Female Genital Mutilation (FGM).	• Children accessing pornography. • Children accessing inappropriate material (including under 13s using social media and under 18s using dating apps; excessive screen time).

Part 1 of the White Paper gives further detail on the above harms.

37 HM Government, *Online Harms White Paper*, April 2019, p6
38 p30
39 p31

The regulatory model

Parts 2 and 3 of the Paper set out in detail the Government's plans for a new regulatory framework. A brief overview of some of the key elements is set out below.

What would the duty require?

The statutory duty of care would require companies to take greater responsibility for the safety of their users and to tackle the harms caused by content or activity on their services.

The regulator would issue codes of practice setting out how to do this. For terrorist activity or child sexual exploitation and abuse (CSEA), the Home Secretary would sign off the codes.

As an indication of fulfilling the duty of care, companies would be expected to:

- ensure their relevant terms and conditions meet standards set by the regulator and reflect the codes of practice as appropriate.
- enforce their own relevant terms and conditions effectively and consistently.
- prevent known terrorist or CSEA content being made available to users.
- take prompt, transparent and effective action following user reporting.
- support law enforcement investigations to bring criminals who break the law online to justice.
- direct users who have suffered harm to support.
- regularly review their efforts in tackling harm and adapt their internal processes to drive continuous improvement.[40]

To help with the above outcomes, codes of practice would include:

- steps to ensure products and services are safe by design.
- guidance about how to ensure terms of use are adequate and are understood by users when they sign up to use the service.
- measures to ensure that reporting processes and processes for moderating content and activity are transparent and effective.
- steps to ensure harmful content or activity is dealt with rapidly.
- processes that allow users to appeal the removal of content or other responses, in order to protect users' rights online.
- steps to ensure that users who have experienced harm are directed to, and receive, adequate support.
- steps to monitor, evaluate and improve the effectiveness of their processes.[41]

Section 7 of the Paper sets out specific areas that codes of practice would be expected to cover in relation to the following types of harm: CSEA, terrorism, serious violence, hate crime, harassment,

---

40 p64
41 pp64-5

disinformation, encouraging self-harm and suicide, the abuse of public figures, cyberbullying, and children accessing inappropriate content.

Who would the duty apply to?

The White Paper noted that harmful content and behaviour originates from a wide range of online platforms or services and that these cannot easily be categorised by reference to a single business model or sector. It therefore focused on the *services* provided by companies. According to the Paper, there are two main types of online activity that can give rise to the online harms in scope:

- hosting, sharing and discovery of user-generated content (e.g. a post on a public forum or the sharing of a video);
- facilitation of public and private online interaction between service users (e.g. instant messaging or comments on posts).

As a wide variety of companies and organisations provide the above services, the regulatory framework would cover social media companies, public discussion forums, retailers that allow users to review products online, non-profit organisations, file sharing sites and cloud hosting providers.[42]

The Paper said that users should be protected from harmful behaviour and content in private as well as public online space. Given the importance of privacy, the framework would "ensure a differentiated approach for private communication". The Paper sought views on how 'private' and 'public' should be defined as well on what regulatory requirements should apply to private communication services.[43]

The regulator

An independent regulator would oversee and enforce the new framework. It would issue codes of practice, setting out what companies would need to do to comply with the duty of care. The regulator's other functions would include:

- establishing a framework to assess compliance with the duty of care;
- overseeing the implementation of user redress mechanisms;
- promoting education about online safety for users;
- taking enforcement action against non-compliant companies;
- promoting the development and adoption of safety technologies to tackle online harms;
- undertaking and commissioning research on online harms and their impact.[44]

42 p49
43 p50
44 p54

The regulator's initial focus would be on companies posing the "biggest and most obvious risk" to users, either because of the size of a service, or because of known harms.[45]

Companies would be required to do what is "reasonably practicable" to meet regulatory requirements. This would be enshrined in legislation.[46]

The Paper sought views on whether the regulator should be a new or an existing body with an extended remit.[47]

Enforcement

The core enforcement powers of the regulator would be:

- Issuing civil fines for proven failures in clearly defined circumstances. Civil fines can be tied into metrics such as annual turnover, volume of illegal material, volume of views of illegal material, and time taken to respond to the regulator.
- Serving a notice to a company that is alleged to have breached standards, and setting a timeframe to respond with an action plan to rectify the issue.
- Requiring additional information from the company regarding the alleged breach.
- Publishing public notices about the proven failure of the company to comply with standards.[48]

The powers would be used in a proportionate manner, "taking the impact on the economy into account".[49]

The paper sought views on other possible powers for the regulator:

- disruption of business activities in the event of extremely serious breaches – e.g. a company failing to take action to stop terrorist use of its services;
- internet service provider (ISP) blocking – an option of "last resort" where a company had "committed serious, repeated and egregious violations of the outcome requirements for illegal harms, failing to maintain basic standards after repeated warnings and notices of improvement";
- senior management liability – holding certain individuals personally accountable in the event of a major breach of the duty of care. This could involve personal liability for civil fines, or could even extend to criminal liability.[50]

Why did the Paper not seek changes to liability?

In the White Paper, the Government explained why it was not seeking changes to the liability regime under the e-Commerce Directive:

45 p54
46 p55
47 pp57-8
48 pp59-60
49 p59
50 p60

> The existing liability regime…does not provide a mechanism to ensure proactive action to identify and remove content. In addition, even if reforms to the liability regime successfully addressed the problem of illegal content, they would not address the full range of harmful activity or harmful behaviour in scope. More fundamentally, the focus on liability for the presence of illegal content does not incentivise the systemic improvements in governance and risk management processes that we think are necessary…

According to the Government, its proposed framework would take a more thorough approach:

> (…) It will increase the responsibility that services have in relation to online harms, in line with the existing law that enables platforms to operate. In particular, companies will be required to ensure that they have effective and proportionate processes and governance in place to reduce the risk of illegal and harmful activity on their platforms, as well as to take appropriate and proportionate action when issues arise. The new regulatory regime will also ensure effective oversight of the take-down of illegal content, and will introduce specific monitoring requirements for tightly defined categories of illegal content.[51]

Consultation

A consultation on the Paper's proposals closed on 1 July 2019. The Government published its interim response on 12 February 2020 (see section 3 of this Paper).

2.1 Comment

The White Paper received a mixed response.

Children's charities

The NSPCC was positive. Chief Executive, Peter Wanless, said that the Paper represented a "hugely significant commitment" that could make the UK a "world pioneer in protecting children online":

> …For too long social networks have failed to prioritise children's safety and left them exposed to grooming, abuse, and harmful content. So it's high time they were forced to act through this legally binding duty to protect children, backed up with hefty punishments if they fail to do so….[52]

The Children's Charities' Coalition on Internet Safety (CHIS) "applaud[ed]" the Paper:

> (…) It recognises self-regulation has failed as the core principle for addressing the challenges facing children as internet users. It sees a statutory Regulator as an essential feature of the future landscape. As a statement of intent, the White Paper's aspirations are crystal clear, and the Government is doubly to be commended for publishing it, written as it was in the face of nearly zero meaningful co-operation from a great many important high-tech companies.[53]

---

51 pp62-3

52 "Government listens to our Wild West Web campaign and launches White Paper", NSPCC News, 8 April 2019

53 CHIS, *Comments on the Online Harms White Paper*, July 2019, p1

Anne Longfield, the Children's Commissioner for England, said that the problem of harmful content on social media was getting worse, that self-regulation had to end, and that the Government's plans for a statutory duty of care "cannot come a moment too soon". She called for the new regulator to "have teeth with strong powers to represent children" and for the balance of power to "to decisively shift" away from the companies.[54]

Digital, Culture, Media and Sport Committee

In a June 2019 report, the Digital, Culture, Media and Sport Committee said it was "pleased" that its recommendations[55] for a duty of care, overseen by an independent regulator, had been included in the White Paper.[56] The Committee said that the regulator's success would depend on its enforcement powers:

> (…) We urge the Government to take an ambitious approach to equipping the regulator with sufficient means, including adequate sanctions. This must go beyond fines to include the ability to disrupt the activities of businesses that are not complying, and ultimately custodial sentences…[57]

Most of the Committee's report discussed the White Paper's "scant focus on electoral interference and online political advertising", both of which the Committee had said required "urgent action."[58]

Carnegie UK Trust

In a June 2019 summary response, Lorna Woods, William Perrin and Maeve Walsh said that the White Paper was a "significant step in attempts to improve the online environment". However, a number of concerns were raised.

The response noted that the White Paper envisaged online operators moderating and taking proactive action in relation to certain forms of content. Although the Paper acknowledged that the e-Commerce Directive prohibits general monitoring, Woods et al claimed that its explanation of how it resolved this conflict was "weak".

The Trust said that it could not support the Government drafting some of the codes of practice, even in relation to the most extreme and harmful speech. According to the Trust, the drafting should, at most, be the responsibility of Parliament or, "more likely", the independent regulator after consultation with bodies such as the police, the security services, the Crown Prosecution Service and possibly the Home Secretary.

The White Paper did not identify an existing body, such as Ofcom, to be the regulator. The Trust said this was a "significant weakness". Ofcom had a proven track record of effective engagement with some of the

---

54 "What does the Government's Online Harms White Paper mean for children?", Children's Commissioner News, 8 April 2019

55 The recommendations were set out in the Committee's February 2019 report, *Disinformation and 'fake news': Final Report* (HC 1791)

56 Digital, Culture, Media and Sport Committee, The Online Harms White Paper, HC 2431, July 2019, p3

57 Ibid, p6

58 Ibid, p6

largest media groups, whereas a new body would take "years to earn sufficient reputation to be taken seriously".

The Trust said that the Paper's distinction between clearly defined and less clearly defined harms was "not helpful". It was also a mistake to exclude economic harm from the framework's scope. In addition, there was no mention of making misogyny a hate crime, despite the Government committing to do so.

Without "urgent clarification" of the points it had raised, the Trust claimed that the Government had "opened itself up to (legitimate) criticism" that its proposed regime was "about moderation, censorship and takedown".

A full response, including the Trust's responses to the White Paper's questions, was published in June 2019.

Further criticism of the framework

Other commentators have drawn attention to the White Paper's "extremely broad" definition of "online harms", that it treats legal and illegal harms as "indistinguishable", risks conflating legal and social issues, and could restrict freedom of expression.[59]

The concept of "harm"

Paul Wragg, Editor in Chief of *Communications Law*, noted that the Paper "moves, awkwardly and confusingly, between criminality and immorality, between commerce and health and safety, between social cohesion and personal development". This is not a "description of *a* problem, or even *some* problems. It is a description of *all* our problems".[60] Wragg claimed there was a tension throughout the Paper between questions of law and ethics, between what is illegal and what is unacceptable. According to Wragg, freedom of expression becomes the "obvious casualty" when attempting to prevent harm to internet users.

Graham Smith criticised the Paper's "all-encompassing" approach:

> [The White Paper] sets out to forge a single sword of truth and righteousness with which to assail all manner of online content from terrorist propaganda to offensive material.
>
> However, flying a virtuous banner is no guarantee that the army is marching in the right direction. Nor does it preclude the possibility that specialised units would be more effective…
>
> (…) An aversion to fragmentation is like saying that instead of the framework of criminal offences and civil liability, focused on specific kinds of conduct, that make up our mosaic of offline laws we should have a single offence of Behaving Badly.

---

59 See, for example, Huw Edwards, "Uncharted territory – the UK sets sail towards regulation of 'Online Harms'", *Social Media Law Bulletin*, 18 April 2019; "The Guardian view on online harms: white paper, grey areas", *Guardian*, 8 April 2019; Emma Goodman, "The Online Harms White Paper: its approach to disinformation, and the challenges of regulation", Inforrm Blog, 13 April 2019; Ashley Hurst, "Tackling misinformation and disinformation online", Inforrm Blog, 16 May 2019

60 Paul Wragg, "Tackling online harms: what good is regulation?", *Communications Law,* Vol 24(2), 2019, pp49-51

> We could not contemplate such a universal offence with equanimity. A Law against Behaving Badly would be so open to subjective and arbitrary interpretation as to be the opposite of law: rule by ad hoc command. Assuredly it would fail to satisfy the rule of law requirement of reasonable certainty. By the same token we should treat with suspicion anything that smacks of a universal Law against Behaving Badly Online.
>
> In placing an undefined and unbounded notion of harm at the centre of its proposals for a universal duty of care, the government has set off down that path.

Smith goes on to discuss how the Paper's "impermissibly vague" concept of harm, including the "nebulous" notion of "harm to society", could cause problems for legislation, hand too much power to the regulator, and restrict freedom of expression.[61]

Freedom of expression

Organisations working to protect freedom of expression have raised concerns about the White Paper.

Open Rights Group

In a May 2019 paper, the Open Rights Group (ORG) noted that "the Internet in general and social media in particular play a central role in protecting free expression in society. They have particular importance for children and young people's expression and access to information".[62]

According to ORG, the Government's proposed framework was "unrealistically vast" and a "poor" conceptual approach.[63] In ORG's view, any regulatory scheme should be "explicitly rooted in the international human rights framework":

> • (…) This provides an established, universally-applicable standard capable of holding both companies and States to account.
>
> • Regulation should encourage internet companies to adopt and implement the UN Guiding Principles on Business and Human Rights . These establish principles of due diligence, transparency, accountability and remediation, and would commit companies to implementing human rights standards throughout their product and policy operations. **We would welcome incorporation of these principles into any regulatory framework so that they become directly enforceable.**
>
> • **It is critical to acknowledge and understand that regulation will ultimately bite on social media users and directly impact the fundamental rights of ordinary citizens**. Regulating social media is essentially different from regulating newspapers or broadcasters because internet media platforms driven by user generated content facilitate the day-to-day freedom of expression of their users.
>
> • **Protection of the right to free speech must infuse how legislative and regulatory schemes are developed,**

---

61 Graham Smith, "Users Behaving Badly: the Online Harms White Paper, Inforrm Blog, 30 April 2019; See also: Graham Smith, "The Rule of Law and the Online Harms White Paper", Informm Blog,12 May 2019

62 Open Rights Group, *Policy responses to Online Harms White Paper*, May 2019, p1

63 Ibid, pp2-3

> **implemented and enforced**. If a harms-based approach is used (which we would not recommend), harms to freedom of expression must themselves be recognised as a harm, to be weighed in any balancing exercise.
>
> • What is legal offline must remain legal online.
>
> • **Platforms must not be obligated to generally scan or monitor content**. Proactive monitoring is inconsistent with the right to privacy and will lead to increased censorship.
>
> • Regulation should promote non-discrimination in decision-making, both human and algorithmic.[64]

ORG also noted the importance of any policy intervention being underpinned "with a clear, objective evidence base which demonstrates that actions are necessary and proportionate":

> Regulation impacting on citizen's free speech needs to be based on evidence of harm traceable to specific pieces or types of content, activity or behaviour, rather than expectations or social judgements that these may be related to possible harms...

There were limitations of research in this area that had to be taken into account when assessing the weight to be given to evidence:

> Risk encounters cannot easily be measured except by asking children directly, which raises ethical (children might be unaware of harm until asked specifically) and measurement (risk of under- or over-reporting) questions. Research is also not able to predict which children will experience harm as a result of encountering risk. Risk refers to the probability of harm, and e.g. encountering hostile messages or pornographic images is not necessarily harmful. Some risks may also be rare but severe in their consequences, and this, too, is difficult to assess. Since children are no more homogeneous than the adult population, a host of factors affect the distribution of risk and harm, vulnerability and resilience.[65]

ORG's response to the White Paper was published in July 2019.

Index on Censorship

The Index on Censorship warned that the White Paper poses "serious risks to freedom of expression online":

> These risks could put the United Kingdom in breach of its obligations to respect and promote the right to freedom of expression and information as set out in Article 19 of the International Covenant on Civil and Political Rights and Article 10 of the European Convention on Human Rights, amongst other international treaties.
>
> Social media platforms are a key means for tens of millions of individuals in the United Kingdom to search for, receive, share and impart information, ideas and opinions. The scope of the right to freedom of expression includes speech which may be offensive, shocking or disturbing. The proposed responses for tackling online safety may lead to disproportionate amounts of legal speech being curtailed, undermining the right to freedom of expression.[66]

---

64 Ibid, p2, emphasis in original
65 Ibid, p4
66 Index on Censorship, *Online harms proposals pose serious risks to freedom of expression*, April 2019

In a June 2019 paper, Index made these 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.[67]

Proportionality

Other commentators have pointed out that attempts to prevent online harm must be proportionate to the risks. Emma Goodman, for example, has written: "While it is true that the most vulnerable need most protection, how to sufficiently protect them without over-protecting the less vulnerable is also a challenge".[68]

In a May 2019 blog, Ashley Hurst argued that the Government "should scale back its ambition to focus on what is illegal and defined, not legal and vague". According to Hurst, the way forward should be focussed on technology and education, approaches that are mentioned in the White Paper, but not in sufficient detail.[69]

A differentiated duty of care?

In a June 2019 paper, Damian Tambini, Associate Professor in the Department of Media and Communications at LSE, acknowledged some of the criticisms of the White Paper. He said that its proposed framework could be "significantly damaging for freedom of expression and pluralism". On the other hand, it could be "a proportionate and

---

67 Index on Censorship, The UK government's online harms white paper: implications for freedom of expression, June 2019; See also Index on Censorship, "Duty of care does not translate well from the offline to online context", Letter to the Secretary of State for Digital, Culture, Media and Sport, 1 July 2019

68 Emma Goodman, "The Online Harms White Paper: its approach to disinformation, and the challenges of regulation", Inforrm Blog, 13 April 2019

69 Ashley Hurst, "Tackling misinformation and disinformation online", Inforrm Blog, 16 May 2019

effective response" to internet harms. Much would depend on what decisions were taken at the next stage of policy development.[70]

Tambini agreed that social media companies have a duty of care to protect users from online harms. He also agreed that there should be a regulator (Ofweb). However, harms were "insufficiently defined" in the White Paper and there was "a blurring of the boundary between illegal and harmful content". In addition, there was a risk of "significant chilling of freedom of expression".[71]

According to Tambini, many of the problems with the Paper's approach could be addressed by "a clear distinction between the illegal/clearly defined and the legal/less clearly defined categories of content".[72] He argued:

> (…) There is a need for a single regulator in order to centralize information, transparency, research, and expertise in one organization. However, the legal and constitutional basis of regulation of illegal content and legal content are distinct and should remain distinct, otherwise there is a risk of chilling of free speech, regulatory uncertainty, and a failure of the regulatory model. The new regulator should work closely with industry to create a new code on illegal online harms. For the category of legal/'difficult to define' content, the regulator should provide research, transparency, good practice, and oversight of a plurality of self-regulatory codes, and work to promote good practice in self-regulation.[73]

The central challenge would be to clarify the process of writing codes of conduct and adjudicating on their application.

Tambini explained how a "differentiated" duty of care would work:

> For illegal content requiring urgent action to be removed or blocked, such as terrorism, child abuse, and hate speech that meets the legal threshold of incitement, an increased pressure to enforce should be ensured by a regulator through a code of conduct that clearly sets out the standards and expectations regarding procedures for removing or blocking content and other relevant responsibilities. These categories of harm are such that they may justify some form of prior restraint, and where the urgency of content removal by platform providers would justify effective deterrent sanctions for their failure effectively to do so. A code would ensure that this process is transparent and standards are supported by Parliament.

For legal but harmful content:

> Ofweb should promote best practice and provide guidance to improve self-regulation, and promote, through monitoring of self-regulation, improved consumer awareness, and competitive pressure for a culture of responsibility. This category of content includes political speech and other sensitive areas where prior restraint would traditionally be regarded as highly undesirable, and where too great a deterrent effect on speech would be

---

70 Damian Tambini, *Reducing Online Harms through a Differentiated Duty of Care: A Response to the Online Harms White Paper*, The Foundation for Law, Justice and Society, June 2019

71 Ibid, p1

72 Ibid, p1

73 Ibid, p8

> socially undesirable. State control of this type of content would too easily tip the balance toward censorship and the restriction of free speech.

Both categories of content would be part of the regulator's remit, but the regulator would adopt a 'one country two systems' approach and not confuse the distinction between illegal and harmful content.[74]

[74] Ibid, p9

3. Interim response to the White Paper consultation (February 2020)

On 12 February 2020, the Government published its initial response to the Online Harms White Paper consultation. The response said that the Government was minded to make Ofcom the regulator for online harms. This was because of its "organisational experience, robustness, and experience of delivering challenging, high-profile remits across a range of sectors".[75] Ofcom welcomed its possible new role.[76]

A brief overview of some of the key themes that emerged from the consultation is set out below. A final Government response is due in spring 2020.

Freedom of expression

Some respondents raised concerns that the White Paper's proposals could impact freedom of expression online.[77] In its response, the Government said that it recognised the "critical importance of freedom of expression, both as a fundamental right in itself and as an essential enabler of the full range of other human rights protected by UK and international law".

To protect freedom of expression, the new regulatory framework would not require companies in scope to remove specific pieces of legal content. Instead, they would be required to explicitly state what content and behaviour would be considered acceptable on their sites and to enforce this consistently and transparently.

Illegal content would have to be removed "expeditiously" and services would also have to minimise the risk of it appearing. Robust action would be required in relation to terrorist content and child exploitation and abuse.

The regulator would not investigate or adjudicate on individual complaints:

> …Companies will be able to decide what type of legal content or behaviour is acceptable on their services, but must take reasonable steps to protect children from harm. They will need to set this out in clear and accessible terms and conditions and enforce these effectively, consistently and transparently. The proposed approach will improve transparency for users about which content is and is not acceptable on different platforms, and will enhance users' ability to challenge removal of content where this occurs.

---

75 DCMS/Home Office, *Online Harms White Paper - Initial consultation response*, February 2020, "Our response" para 11; A Written Ministerial Statement on the Government's response was made on 12 February 2020

76 *Ofcom response to Government announcement on online harms regulation*, 12 February 2020

77 DCMS/Home Office, *Online Harms White Paper - Initial consultation response*, February 2020, Ministerial foreword and "Our response" para 1

> Companies will be required to have effective and proportionate user redress mechanisms which will enable users to report harmful content and to challenge content takedown where necessary. This will give users clearer, more effective and more accessible avenues to question content takedown, which is an important safeguard for the right to freedom of expression. These processes will need to be transparent, in line with terms and conditions, and consistently applied.[78]

Comment

Graham Smith has commented that the different approaches to legal but harmful content, as opposed to illegal content, is "perhaps the most significant policy development compared with the white paper".[79] However he raises two points. Firstly, in relation to legal content that might be seen by children:

> (…) the intermediaries' freedom is restricted to content seen by adults. But at the same time "All companies in scope will need to ensure a higher level of protection for children, and take reasonable steps to protect them from inappropriate or harmful content." If there is any possibility that the audience includes children (which, under UK law, includes anyone under the age of 18), does that trigger stricter duty of care requirements? If so, does that render illusory the apparent freedom to decide on content? The Response says:
>
> "Under our proposals we expect companies to use a proportionate range of tools including age assurance, and age verification technologies to prevent children from accessing age-inappropriate content and to protect them from other harms."
>
> Would non-compliance with age verification standards laid down by the regulator automatically disapply the intermediary's freedom to determine what is permissible on its platform?

Secondly, in one place the response refers to services in scope having to enforce terms and conditions "consistently and transparently". In another it states that this must be done "effectively, consistently and transparently". According to Smith, this raises issues about the regulator's supervisory remit:

> (…) Would (as at any rate the first version appears to suggest) the regulator's remit be strictly limited to assessment by reference to the content standards set in the intermediary's own terms and conditions? Or could the regulator go beyond that and assess effectiveness in reducing harm?
>
> If the latter, then the regulator could enter the realm of (for instance) requiring algorithms to be designed so as to make particular kinds of lawful content less readily accessible to users, recommendation algorithms to be tweaked, dissemination tools to be modified and so on – all while still adhering to the letter of the government's commitment that the platform has freedom to decide what material is permitted on its platform.[80]

---

78 DCMS/Home Office, *Online Harms White Paper - Initial consultation response*, "Our response" paras 1-4

79 Graham Smith, "Online Harms Deconstructed: the Initial Consultation Response", Inforrm Blog, 20 February 2020

80 Ibid

Smith points out that the interim response does not mention the e-Commerce Directive. David Barker also observes that the response is "silent" on this:

> [the white paper] itself said that the "new regulatory framework will increase the responsibility of online services in a way that is compatible with the EU's e-Commerce Directive, which limits their liability for illegal content until they have knowledge of its existence, and have failed to remove it from their services in good time". The response paper is notably silent on this point. Perhaps the UK's departure from the EU may have meant it was off message to refer to the E-Commerce Directive safe harbours, and the Regulations bringing them into force in UK law. Ensuring that legislation takes proper account of the law on safe harbours will be critical. This issue cannot simply be ducked…[81]

The Open Rights Group has claimed that the proposals in the response "could cause vast restrictions on free speech". ORG also thinks that the proposed 'duty of care' "remains an open-ended and vague concept that needs vastly clearer definition".[82]

The duty of care - businesses in scope

Respondents to the White Paper emphasised the need for clarity for businesses. The initial response states that the proposed duty of care would only apply to companies that: "provide services or use functionality on their websites which facilitate the sharing of user generated content or user interactions, for example through comments, forums or video sharing".

According to the Government, fewer than 5% of UK businesses would be in scope:

> (…) Just because a business has a social media page that does not bring it in scope of regulation. Equally, a business would not be brought in scope purely by providing referral or discount codes on its website to be shared with other potential customers on social media. It would be the social media platform hosting the content that is in scope, not the business using its services to advertise or promote their company. To be in scope, a business would have to operate its own website with the functionality to enable sharing of user-generated content, or user interactions…[83]

The regulator would issue guidance to help businesses understand whether they would be within scope of the duty of care.

Comment

According to Cathryn Hopkins and Dan Tench, the White Paper and the initial response do not properly address jurisdictional issues. The response refers to "UK businesses" but does not clarify whether only businesses registered in the UK would be within scope of the framework. If businesses registered elsewhere,

---

81 David Barker, "Online harms: the good, the bad and the unclear", Inforrm Blog, 22 February 2020

82 "Online Harms Regulation Threatens Free Speech", ORG press release, 12 February 2020

83 DCMS/Home Office, *Online Harms White Paper - Initial consultation response*, "Our response" paras 7-9

offering services to people in the UK, would be within scope, then complex jurisdictional issues would arise.[84]

David Barker has suggested that the regulator not investigating individual complaints could lead to confusion about what a duty of care means in terms of claiming redress:

> (…) we currently face the prospect of a regulatory framework which will have the badge of a duty of care but which will leave individuals distinctly confused about what that duty of care means for them in terms of claiming redress. They will not be able to pursue a complaint to the regulator about their individual circumstances. They may be able to bring a claim through ordinary legal proceedings concerning what has happened to them, but that claim will need to be based on the existing framework of their legal rights. They will not be able to contend that they are entitled to remedies based on the new duty of care because the new regime will not establish a new private right of action in tort. How are judges supposed to navigate all of this?[85]

Protecting children

Many respondents to the White Paper stressed the importance of higher levels of protection for children.[86] In its response, the Government said that companies would be expected to use a range of tools, such as age verification technologies, to prevent children from accessing age-inappropriate content and to protect them from other harms. This would, among other things, protect children from online pornography and fulfil the aims of the *Digital Economy Act 2017*.[87] In October 2019, the DCMS announced that it would not be introducing age verification for online pornography through provisions in the 2017 Act but would instead deliver its objectives through the regulation of online harms. For background, see the Library Paper *Online pornography: age verification* (CBP 8551, 18 October 2019).

Comment

David Barker has warned that the protection of children online is perhaps the most difficult issue of all and that legislators will have "to think about the way in which children live now, and the way in which technology is an absolutely integral part of their lives".[88]

According to Graham Smith, the duty of care approach "seems to presuppose a system of age verification" if services are "to avoid the more onerous duties that would apply for lawful content that could be seen by children". He comments: "How that might operate, and how it

---

84 Cathyrn Hopkins and Dan Tench, "Online Harms White Paper: The Government's Initial Consultation Response", Inforrm Blog, 14 February 2020
85 David Barker, "Online harms: the good, the bad and the unclear", Inforrm Blog, 22 February 2020
86 DCMS/Home Office, *Online Harms White Paper - Initial consultation response*, Ministerial foreword
87 Ibid, "Our response" para 20
88 David Barker, "Online harms: the good, the bad and the unclear", Inforrm Blog, 22 February 2020

would mesh with the Information Commissioner's Age Appropriate Design Code, is anyone's guess at present".[89]

Enforcement

One of the issues that the Government is still considering is the enforcement powers of the regulator and senior management liability. The initial response states:

> 19. We recognise the importance of the regulator having a range of enforcement powers that it uses in a fair, proportionate and transparent way. It is equally essential that company executives are sufficiently incentivised to take online safety seriously and that the regulator can take action when they fail to do so. We are considering the responses to the consultation on senior management liability and business disruption measures and will set out our final policy position in the Spring.

Next steps

The Government's interim response stresses that it is an "iterative step" in developing its policy on online harms. While legislation is prepared, wider measures will be taken to ensure that progress is made on online safety. These include:

- Interim codes of practice - providing guidance to companies on how to tackle online terrorist and child sexual exploitation and abuse. The codes will be voluntary but are intended to bridge the gap until the regulator becomes operational;
- A Government transparency report – to be publised in the next few months;
- A media literacy strategy - to be published in summer 2020;
- A forthcoming publication on the the safety technology ecosystem.[90]

In response to an Urgent Question on 13 February 2020, the DCMS said that legislation would be introduced in this parliamentary session.[91]

---

89 Graham Smith, "Online Harms Deconstructed: the Initial Consultation Response", Inforrm Blog, 20 February 2020; The Information Commissioner's Office (ICO) published its *Age appropriate design: a code of practice for online services* in January 2020

90 DCMS/Home Office, *Online Harms White Paper - Initial consultation response*, Executive summary

91 Urgent Question on Online harms legislation, HC Deb 13 February 2020 cc971-81; The December 2019 Queen's Speech said that the Government would "develop legislation to improve internet safety for all"

BRIEFING PAPER

Number 8743

26 February 2020

About the Library

The House of Commons Library research service provides MPs and their staff with the impartial briefing and evidence base they need to do their work in scrutinising Government, proposing legislation, and supporting constituents.

As well as providing MPs with a confidential service we publish open briefing papers, which are available on the Parliament website.

Every effort is made to ensure that the information contained in these publicly available research briefings is correct at the time of publication. Readers should be aware however that briefings are not necessarily updated or otherwise amended to reflect subsequent changes.

If you have any comments on our briefings please email papers@parliament.uk. Authors are available to discuss the content of this briefing only with Members and their staff.

If you have any general questions about the work of the House of Commons you can email hcenquiries@parliament.uk.

Disclaimer