Free Speech Union briefing

You're On Mute:

The Online Safety Bill and what the Government should do instead

Dr. Radomir Tylecote, Victoria Hewson, Matthew Lesh, Dr. Bryn Harris

April 2021

Contents

Executive Summary

The Government plans to put an Online Safety Bill, previously called the Online Harms Bill, before Parliament in 2021.

There have been some changes to the Bill, but most of the concerns the FSU expressed before still apply. The Bill remains a serious threat to freedom of speech.

The Government wants to make the UK "the safest place in the world to go online" but its plans will restrict online free speech to a degree almost unprecedented in any democracy. If the Government keeps to the plans laid out in its latest consultation response this will result in the censorship of legal speech.

Many of the so-called "harms" the Government describes are dangerously vague, including "offensive material", as if offense is a harm the public need protecting from by the state.

The Government wants to impose a "duty of care" on internet companies so that they "take responsibility for the safety of their users" and prevent harms resulting from other people's conduct, in effect making companies responsible for how members of the public treat each other, an unprecedented proposal that is both unenforceable and diminishes individual responsibility.

This duty will mean internet companies are obliged to remove "harmful content", including that which risks "significant adverse… psychological impact on individuals". "Psychological impact" could mean almost anything. Social media platforms and websites that breach this duty could be fined up to £18m or 10 per cent of their global annual turnover. This will lead to companies being over-cautious and removing any material that might fall foul of the new rules.

Companies will also be obliged to remove "disinformation and misinformation" which "could cause significant harm to an individual". What does the Government mean by this? When discussing how to tackle "disinformation",

the Government links to YouTube's page outlining its policy of censoring content that "contradicts the World Health Organization". YouTube's censorship of talkRADIO earlier this year was dictated by this exact policy. The Government seems to regard this as best practice, rather than an unacceptable act of censorship.

The proposed new internet regulator is Ofcom, which is worrying. Its head, Dame Melanie Dawes, has said that on gender and transgender issues broadcasters should "steer their way through these debates without causing offence and without bringing inappropriate voices to the table", which shows that Ofcom is no friend of free speech. This year, Ofcom enlarged the number of "protected characteristics" in its broadcasting code from four to 18, with broadcasters told that "hate speech" in their programmes now means: "all forms of expression which spread, incite, promote or justify hatred based on intolerance on the grounds of disability, ethnicity, social origin, sex, gender, gender reassignment, nationality, race, religion or belief, sexual orientation, colour, genetic features, language, political or any other opinion, membership of a national minority, property, birth or age".

The Government will set out priority categories of "legal but harmful" material in secondary legislation, including so-called hate content. Alarmingly, this is pending Ofcom's advice.

Even though the Government claims the Bill will protect press freedom, their recent changes are limited to a newspaper's own website or newsfeed and does not include journalists' Facebook pages – so if a journalist posts a story they've written to Facebook, it could be taken down.

The Government has stated that its proposals are partly inspired by Germany's 2017 "NetzDG" internet law. President Lukashenko of Belarus, Vladimir Putin's United Russia Party and Venezuela have also cited NetzDG as the model for their online laws.

The authors of this report oppose the Government's planned approach and propose an alternative that involves removing all references to "legal but harmful" material and providing more resources to law enforcement to better address genuine harms like child abuse and terrorist activity. However, if the Government goes ahead with the legislation the protections for freedom of expression should:

- Remove any obligation imposed on internet companies not to publish “offensive” views;
- Insist that Ofcom pay due regard to freedom of expression and preventing it from punishing an internet company that refused to remove content that is protected under English common law or the European Convention on Human Rights;
- Make sure the codes of practice that internet companies are bound by refer to the protection of freedom of expression as one of their objectives;
- Remove any reference to disinformation and misinformation from the Bill;
- Include a mechanism whereby individuals can appeal a decision to remove content they have posted to Ofcom.

Authors

Dr Radomir Tylecote is Research Director of the Free Speech Union. He is frequently published in the UK press including The Telegraph and the Mail. Radomir has a PhD from Imperial College Business School and an MPhil from Cambridge University. He is also a Fellow of the Institute of Economic Affairs and Director of Defence and Security for Democracy (DSD) at Civitas.

Victoria Hewson is Head of Regulatory Affairs at the Institute of Economic Affairs. A practising solicitor, she has specialised in commercial, technology and data protection matters across a range of sectors. She writes regularly for The Telegraph and City AM and appears on television and radio to discuss trade and regulatory policy and matters affecting free speech and civil liberties.

Matthew Lesh is Head of Research at the Adam Smith Institute. He regularly appears across radio, televison and print to discuss technology policy issues and freedom of expression. He previously founded a mobile application development startup. Matthew has a Bachelor of Arts (Degree with Honours) from the University of Melbourne and a Masters in Public Policy and Administration from London School of Economics.

Dr Bryn Harris is Chief Legal Counsel of the Free Speech Union. He received a doctorate in Classics from the University of Oxford, and was called to the Bar in 2018.

All views expressed in this document are those of the authors. The Free Speech Union and its directors have no corporate view.

Introduction: what are the Government's plans?

The Government published the Online Harms White Paper in April 2019.[1] In December 2020[2], it affirmed its intention to put an Online Safety Bill, previously called the Online Harms Bill, before Parliament in 2021.

The Government's stated aim is to make the UK "the safest place in the world to go online", but its plans will restrict our freedom of speech online to a degree that is almost unprecedented in any democracy.

Outlines

The Government's plans include:

- Regulations that will apply to companies hosting user-generated content that is accessed in the UK.

Imposing a "duty of care" to make companies "take responsibility for the safety of their users". This will cover "harmful content" which risks "significant adverse physical or psychological impact on individuals", including content "that is harmful to adults". Internet service providers and smaller "low-risk" companies will be exempt from the duty of care.

- Legislation will set out a general definition of the "harmful content" the duty of care covers. Failure to fulfil it may see the regulator "taking robust enforcement action." Search engines will also need to "assess the risk of harm" across "their entire service".

1 Department for Digital, Culture, Media and Sport and the Home Office. Consultation outcome, Online Harms White Paper. Updated 15 December 2020: https://www.gov.uk/Government/consultations/online-harms-white-paper/online-harms-white-paper#conclusion-and-next-steps

2 HM Government. Online Harms White Paper. Full Government Response to the consultation. December 2020. (Direct quotes refer to this document unless otherwise stated) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/944310/Online_Harms_White_Paper_Full_Government_Response_to_the_consultation_CP_354_CCS001_CCS1220695430-001__V2.pdf

- The Government confirms that the regulator will be Ofcom, which "will issue codes of practice" outlining the systems and processes companies will have to use to fulfil their duty of care. Companies must either comply with the codes or demonstrate another effective approach. There will be distinct codes of practice for search engines.

- Parliament will "have a role" approving codes of practice, but it is unclear what this will be. The Home Secretary will be able to order reviews into the codes as far as they deal with child abuse and terrorist content, but it is not clear he or she will be able to order reviews for freedom of speech reasons. The Government has not produced an exhaustive list of the systems companies may be forced to adopt.

- Ofcom's imposition of codes of practice will force companies to make rules on what is permitted on their platforms, giving them a role comparable to law enforcement.

- The plans state that only companies "providing Category 1 services" (larger social media companies with content-sharing) must deal with "legal but harmful content and activity that is accessed by adults". The Government claims this will not require removal of specific pieces of legal content, simply that companies "must consider the impacts of their decisions regarding moderation and design choices on user safety".

- Codes of practice will include some kind of process "to allow users to report harmful content or activity and to appeal the takedown of their content", implying that legal content is liable to be removed. The process for redress is unclear.

- The Government will set out priority categories of legal but harmful material including hate content, pending regulator Ofcom's advice. Companies will have to address these in terms and conditions "at a minimum".

- The duty of care will also cover "Disinformation and misinformation" which "could cause significant harm to an individual", with secondary legislation outlining categories that "companies must address in their terms and conditions." This is also liable to silence contrarian opinion: the definition says misinformation is "inadvertently spreading false information", which could simply mean accidentally stating something that is not factually correct.accidentally stating something that is not factually correct.

- The regulation will encourage self-censorship to mitigate the risk of massive fines. Ofcom will have a duty to support innovation and reduce burdens on business, but will have the power to issue fines of up to £18m. The Online Safety Bill will stifle new firms by imposing large regulatory costs.

Undermining the marketplace of ideas

Some of the harms the forthcoming Online Safety Bill is intended to address genuinely require state action, such as preventing the distribution of images of child abuse or terrorist activity online. However, these would be better addressed through a combination of simpler and better targeted legislation and more resources for law enforcement.

Some of the harms the White Paper outlines are vague, such as "disinformation" and "misinformation". The Government previously highlighted "offensive material", as if giving offense is a harm the public should be protected from by the state.[3] The inclusion of these vague "harms" risks considerable censorship by internet companies to avoid breaking the law.

The Government has stated that the proposals are partly inspired by Germany's 2017 "NetzDG" internet law. Notably, President Lukashenko of Belarus, Vladimir Putin's United Russia Party and the Venezuelan Government have also cited NetzDG as the model for their repressive online laws. The UK Government's plans also show disturbing similarities to Beijing's policies. In China, "rumours" are censored because they cause "social harms"; the UK Government has said "disinformation" which may be "harmful" will be blocked by making "content which has been disputed by reputable fact-checking services less visible to users", encouraging companies to promote "authoritative news sources".

In December the Government published its final consultation response. While most of the concerns the FSU has expressed before still apply, there have been some changes in the Government's plans, which in our view have improved the plans in some respects, but made them more concerning, incoherent and confusing in others. The Bill remains a serious threat to freedom of speech.

3 Department for Digital, Culture, Media and Sport and the Home Office. Consultation outcome, Online Harms White Paper. Updated 15 December 2020: https://www.gov.uk/Government/consultations/online-harms-white-paper/online-harms-white-paper#conclusion-and-next-steps

Chapter 1: The threat to freedom of speech online

Harmful content, expectations on companies and exemptions

The Government confirms that the regulator will be Ofcom, previously stating "Ofcom is a well-established independent regulator with a strong reputation internationally and deep experience of balancing prevention of harm with freedom of speech considerations". Aside from the irrelevance of its "international reputation", Ofcom has already demonstrated a dangerous attitude to freedom of speech.

On 1st January, Ofcom extended its hate speech definition to include intolerance of gender reassignment, social origin and "political or any other opinion" (required under the revised Audiovisual Media Services (AVMS) Regulation 2020). There are now 18 protected characteristics in its broadcasting code: previously there were four. Broadcasters must now ensure programmes avoid so-called "hate speech", defined as:

> all forms of expression which spread, incite, promote or justify hatred based on intolerance on the grounds of disability, ethnicity, social origin, sex, gender, gender reassignment, nationality, race, religion or belief, sexual orientation, colour, genetic features, language, political or any other opinion, membership of a national minority, property, birth or age.

This very broad definition of "hate speech" undermines the free expression of ideas and demonstrates Ofcom's poor track record in defending free speech.

The new regulatory framework will apply to companies that host user-generated content which can be accessed by users in the UK and/or facilitate public or private online interaction between service users, one or more of whom is in the UK. Businesses which are considered low-risk and therefore exempt include online services used internally by organisations (such as intranets), online services managed by educational institutions (already subject to "sufficient safeguarding duties or

expectations") and email and telephony. For the "dark web", the Government is employing a "law enforcement response to tackle criminal activity" instead of a regulatory approach.

Harmful content which is covered by the refined "duty of care" includes "online content and activity [that] give[s] rise to a reasonably foreseeable risk of a significant adverse physical or psychological impact on individuals…"

There will be "differentiated expectations on companies", determined by "high level factors which lead to significant risk of harm occurring to adults through legal but harmful content". These factors will be the size of a service's audience and the functionalities it offers: some of these, like sharing content or contacting users anonymously, "are more likely to give rise to harm".

There is no objective standard of which harms cause a sufficient level of "adverse… psychological impact". This high level of regulatory discretion risks an arbitrary and heavily politicised approach to online regulation, with the regulator chasing after issues or platforms that happen to be unpopular at any given time. This model will also lead to inevitable lobbying, with larger companies able to secure a place at the table to design regulations in their favour and to the detriment of competitors.

Ofcom will oversee the duty of care and issue codes of practice that companies must follow to be compliant, forcing companies to decide what is legal and illegal on their platforms, giving them the role of law enforcement and the courts. This duty of care and these codes of practice could apply to all companies and content, including private channels. This is an extraordinary power for Ofcom to have, making it an online judge, jury and executioner, with limited public accountability or protection for freedom of expression.

"Legal but harmful" content

The Government has stated that only companies "providing Category 1 services" (mainly the major social media networks) will need to "take steps in respect of legal but harmful content and activity that is accessed by adults". This will not directly require removal of specific pieces of legal content, but companies "must consider the impacts of their decisions regarding moderation and design choices on user safety", which the Government believes will "empower adult users to keep themselves safe online and protect freedom of expression, by preventing companies from arbitrarily removing content".

Yet Government also says that as-yet unwritten codes of practice will include some process "to allow users to report harmful content or activity and appeal the takedown of their content", implying that *legal content is in fact liable to be removed*, with any process for redress still unclear, meaning it is also unclear what the Government sees as "arbitrary" removal. There is a contradiction at the heart of these proposals: the Government claims there will be protection from arbitrary removal and no obligation to remove legal content, but at the same time expects internet company adherence to codes of practice about "legal but harmful" speech, backed by huge fines. In practice, a company will only be sanctioned for a failure to remove content, or perhaps a failure to operate a process that allows for demands to remove content or take other censorious steps like "down-ranking" content.

The Government will set out priority categories of legal but harmful material in secondary legislation, and this will include hate content. Alarmingly, this is pending Ofcom's advice. Companies will then be expected to address these through their terms and conditions "at a minimum". Terms and conditions will state how they will handle priority categories of legal but harmful material established in legislation, "and any others identified by them through their risk assessment [while they] will be expected to consult with civil society and expert groups when developing their terms and conditions".

It is extremely concerning that the state, through a regulator, is involving itself in the management of speech that is otherwise legal. The practical impact of the "legal but harmful" provisions is that major "Category 1" platforms will be required to censor substantial quantities of legal speech. The framework will encourage platforms to engage in excessive censorship to avoid the threat of substantial fines from the regulator.

Even if online censorship is not the immediate intent of the current Government, there would remain an ongoing risk of further encroachments of legal speech once this framework is established. Once the precedent is set for a state role in "legal but harmful" speech, the political pressure will always be for expanding the codes of practice to cover more areas.

The intention to initially include fewer companies in the scope of the legislation is an improvement, but the scope of what is believed "harmful" is still broad. For example, one type of "harmful" content is "content about… suicide", but like much else this is not defined. Will discussing *Madame*

Butterfly or *Romeo and Juliet* be banned? This may seem like a joke, but it shows how vague these ideas are, and that vast swathes of expression might be called harmful.

Disinformation and misinformation

The Government's approach to "disinformation and misinformation" remains a threat to free speech. These are vague terms that are often used to justify silencing contrarian opinions. The concept will fall under the "duty of care" if it could, in theory, "cause significant harm to an individual" such as "relating to public health." Secondary legislation will propose specific types as categories of "priority harm" that companies will have to address in their terms and conditions and introduce further provisions "to address the evolving threat", like "an expert working group".

But even as they stand, the Government's plans are a danger to free speech. The duty of care will apply to content which could cause "significant physical or psychological harm to an individual". Ofcom will be given "the power to act" where it regards any "disinformation and misinformation" as a "significant threat to public safety, public health or national security". These are extremely broad categories that could be interpreted widely by the regulator.

The plans define disinformation as the "deliberate creation and dissemination of false and/or manipulated information that is intended to deceive and mislead audiences, either for the purposes of causing harm, or for political, personal or financial gain". We suggest that politicians might be especially careful about introducing a definition like this. Who will judge whether something is false or deliberate?

The definition calls misinformation "inadvertently spreading false information". But this could simply mean accidentally stating something that is not factually correct. It is extraordinary, indeed chilling, that the Government believes its job is to prevent people from making unintentional untrue statements. More deeply, it is frequently impossible to determine in advance what is true and what is false. This is precisely why we have freedom of speech: to encourage debates about controversial issues, including the expression of unorthodox ideas that challenge what people currently believe to be true. The Catholic church believed that the proposal by Copernicus and Galileo that the earth orbited the sun was misinformation: had Ofcom existed at the time, would it have expected the publications of the day to censor the idea?

Ofcom will be able to fine any "in-scope company worldwide that provides services to UK users" up to £18m (or 10% of global annual turnover, whichever is higher), or take "business disruption measures". This power to pass judgment on misinformation and disinformation is bound to stifle free debate.

The examples the Government provides to justify its position actually undermine it. It claims:

> COVID-19 has brought these dangers into sharp focus. Ofcom data suggests that in week one of the UK lockdown, nearly 50% of respondents reported seeing information they thought to be false or misleading about the pandemic, with this figure at almost 60% for 18-34 year old respondents.

But this could simply mean that half the population read something they disagreed with. So what? In fact, this figure demonstrates the opposite of what the authors think. It shows that people are already capable of judging for themselves, because they can spot "misinformation". It goes on:

> While Ofcom has recorded a gradual decrease in self-reported exposure to narratives considered false or misleading, navigating a COVID-19 online environment can be challenging and at times confusing for many people in the UK.

To say that many people find information about COVID-19 online "confusing" is not simply patronising or paternalistic, but speaks to an authoritarian attitude that suggests the public, unlike the Government, is incapable of distinguishing truth from fiction. It must be said that the quotes above suggest that the opposite is sometimes the case. We often do not know the truth in advance of public debates that set out each case, but rather, through that very debate, which depends on the ability of all parties to express ideas.

When the Government promises that the "duty of care will apply to disinformation and misinformation that could cause harm to individuals, such as anti-vaccination content", it is worth asking who they intend this to apply to. In recent months the German and French Governments have made apparently false claims about the Oxford-AstraZeneca Coronavirus vaccine, which may have been made for political reasons and without thorough scientific evidence. These claims may even fall under part of

the Government's proposed definition of disinformation, the "deliberate creation and dissemination of false and/or manipulated information… for political… gain". We do not suggest that these Governments' claims should be censored in the UK, but this example demonstrates how many claims by politicians are liable to fall under the definition.

Yet the Government has not called for the social media accounts of the German or French Government to be censored (just as Twitter has not censored the comments on President Emmanuel Macron's account). Are we to expect different censorship rules for the powerful and the powerless? It is hard to believe that Ofcom will expect the censorship of German and French Government social media accounts. But if it doesn't, who will it target, and why? But it is not in the public interest even for the false claims of foreign Governments to be censored: people who are only allowed to see a sanitised version of the world and its politics cannot build an informed picture of reality.

The Government also boasts that its "cross-Whitehall Counter Disinformation Unit" has unearthed "false narratives" which have "caused significant harm to individuals and society". These include "conspiracy theories" about "junk cures". The Government also notes approvingly that YouTube censors content that "contradicts the World Health Organization". It even links to the Google page outlining the YouTube censorship rules, a strange conflagration of the opinions of the UK state and a global corporation. YouTube's censorship of talkRADIO earlier this year followed this exact policy. This suggests that the Government regards this as best practice, even though Ministers claimed to be against it.

The WHO has changed its mind about many aspects of the Covid crisis, including whether human to human transmission was occurring (repeating the Chinese Government line that there was no sustained human to human transmission until 22nd January 2020 and implicitly supporting China's claim that the Chinese doctors claiming otherwise were posting online misinformation), and whether to wear face masks (like the British Government, the WHO reversed its position on face masks, having claimed until 6th June that masks for uninfected people would only provide a false sense of security). Approving of the censorship of content that "contradicts the World Health Organization" is a dangerous position to take and an example of why many of these proposals are so badly thought through. These examples show why open disagreement with those in power does not constitute dangerous misinformation, but is what allows policy to improve.

The WHO denied that human to human transmission was occurring after China's President Xi Jinping reportedly personally urged its Director General Dr Tedros Adhanom Ghebreyesus to "hold back information about human-to-human transmission", according to Germany's Federal Intelligence Service (BND).[4] Tedros also praised China's "transparency" after it cracked down on the doctors who tried to alert it to the disease and has called China's President Xi Jinping a "visionary". Since 2019, the WHO's International Statistical Classifications will include traditional Chinese medicine as well as evidence-based medicine,[5] despite there being little evidence for the efficacy of many of these remedies, some of which use endangered wildlife products.[6] Therefore the Government boasts of its work preventing the promotion of quack cures, but in the same section of its White Paper approves of censoring contradiction of an international organisation that is effectively promoting a system of medicine whose remedies must be said to include quack cures.

In the link provided, YouTube also states that it also "doesn't allow content that spreads medical misinformation that contradicts *local health authorities*' [medical] information about COVID-19". Here, the Government is implying that it also wants to censor opinions that contradict its own.

In its most recent White Paper response, the Government notes: "Facebook has expanded its work with fact-checkers to continue addressing misinformation. In March 2020, Facebook displayed warnings on roughly 40 million posts related to COVID-19, based on 4,000 articles reviewed by independent fact checkers. When users saw the warning labels, 95% of the time these users did not go on to visit the original content." Facebook employs the Centre For Combatting Digital Hate for fact-checking services, which regards the theory that the virus originated in the Wuhan Institute of Virology as a "conspiracy theory". But this theory has since been called at the very least plausible and worth exploring further by the US State Department and a large number of scientists.

4 Mark Moore. "China pressured WHO to delay global coronavirus warning: report." *NY Post.* https://nypost.com/2020/05/10/china-pressured-who-to-delay-global-coronavirus-warning/

5 Lam et al (2019). "Impact on Traditional Chinese Medicine and World Healthcare Systems" *Pharmaceutical Medicine*, 2019 Oct; 33(5): 373-377 ICD-11. https://pubmed.ncbi.nlm.nih.gov/31933225/

6 Cees N. M. Rencken and Thomas P.C. Dorlo. "Quackery At WHO: A Chinese Affair." Skeptical Inquirer, Vol. 43, No. 5, September/October 2019. https://skepticalinquirer.org/2019/09/quackery-at-who-a-chinese-affair/

The Government's own text links to a statement by Mark Zuckerberg boasting that once "misinformation" is "rated false by fact-checkers, we reduce its distribution". It is worrying that the Government is almost cutting and pasting policy from the internet corporates that are now mired in censorship scandals of their own. After the WHO changed its position on masks, Facebook flagged as "false" an article in the Spectator by Oxford's Professor Carl Heneghan and Professor Tom Jefferson about a Danish study of 6,000 people which suggested that wearing them had a negligible effect in protecting the user.

This concern also applies to companies' active manipulation of algorithms on what content people can see. This is often proposed as a way to counter disinformation and misinformation but is liable to infringe freedom of expression. Allowing protesters only to express their opinions in an empty hall with no onlookers is an anaemic form of free expression. Limiting the audience is limiting access to speech.

Clearly, the Government approves of corporations' success in reducing the opportunity of the reading public to make decisions for themselves (without regard to the possible side effect if people believe that information online has been made "reliable" and needs no further critical thought). But fact-checkers are no less free from bias than anyone else, and frequently make mistakes (or, following the Government's own definition, spread "misinformation").

The Government claims to want to rein in the "harmful" power of social media giants but also believes they should act as the arbiters of scientific truth. These two beliefs are not compatible.

Given its views on the WHO, for example, the Government should also provide more clarity on its plans to "strengthen the integrity of UK elections and promote fact-based and open discourse" under its Defending Democracy programme.

Free speech for the press but not for the public?

The Government now states that "to protect media freedom, legislation will include robust protections for journalistic content shared on in-scope services" and that "[t]he Government is committed to defending the invaluable role of a free media".

While at first glance, greater protections for journalists may appear welcome, this approach in fact raises more problems than it solves. Freedom of speech requires not just a free press but free speech for the public at large. Journalists should benefit from broader protection of free expression; it is not clear that they are entailed to special protection. The White Paper response text goes on to say:

> Journalistic content is shared across the internet, on social media, forums and other websites. Journalists use social media services to report directly to their audiences. This content is subject to in-scope services' existing content moderation processes. This can result in journalistic content being removed for vague reasons, with limited opportunities for appeal. Media stakeholders have raised concerns that regulation may result in increased takedowns of journalistic content… In order to protect media freedom, legislation will include robust protections for journalistic content shared on in-scope services.

This new approach is nonsensical. It puts journalistic content on social media into a special class, no matter whether it contravenes other policies from social media sites. It is also an effective admission that the proposed framework will encourage social media companies to remove content for vague reasons with limited opportunities for appeal. It also raises questions about who is entitled to these special journalist protections.

The Government also states that:

> Content and articles produced and published by news services on their own sites do not constitute user-generated content and therefore fall outside the scope of legislation… comments on articles on news publishers' sites will be explicitly exempted from scope… to protect media freedom, legislation will include robust protections for journalistic content shared on in-scope services.

What about when individuals who are not journalists want to write and publish independently of news websites? Will freedom of speech apply only to card-carrying members of the National Union of Journalists? (Even journalists will still be bound online and offline by the tightening strictures of other speech laws, not to mention the speech codes of the National Union of Journalists, for instance.) Free speech should not require a special pass. These proposals run against the principle of equality before the law. The White Paper also says it supports most recommendations in the 2019

Cairncross Review (on “A sustainable future for journalism”). At least one of these is a concern for a free press. Recommendation 9 of the Review (“Establish an Institute for Public Interest News”) proposes “a dedicated body … working in partnership with [platforms including] Nesta, Ofcom, the BBC and academic institutions [which] might become a rough equivalent to the Arts Council, channelling a combination of public and private finance into those parts of the industry it deemed most worthy of support”.

It would be unwise to establish yet another quango that will channel taxpayers’ money to the favourite media outlets of Nesta, Ofcom and the BBC. This appears to be an attempt by these state-sponsored organisations to tilt the playing field even further towards output that propagates their worldview. This would harm press freedom and incentivise editors to produce content that could gain them these financial rewards. The Government should not give these bodies even more power over what we watch and read.

The Government’s proposals have emerged in an environment where the UK authorities are increasingly committing resources to policing speech online – resources which could be used to prevent and prosecute genuine harms such as child abuse and terrorism. No UK Government or Opposition should support proposals which give internet censors, whether this be a state regulator or “fact-checkers” employed by social media companies, the power to censor the sometimes-offensive free speech which is part of any democracy. Political parties should also note that this will inevitably result in the censorship of their own activists.

The internet is not a problem to be solved, but one of the greatest tools ever created for the free flow of information, research and debate – in considerable part due to the efforts of British scientists. The freedoms it creates and enhances are sometimes abused. But that can be said of any of the freedoms of a democratic society. It is no reason to roll them back.

In the next section, we propose the outlines of an alternative White Paper that sets out ways in which regulation could protect the vulnerable without jeopardising free speech.

Chapter 2: Our approach

The premise of our proposals is that the Government's proposals and its internet regulator:

1. are serious threats to freedom of expression;
2. will undermine innovation, online competition, and the digital economy; and
3. will do little to tackle actual online harm.

The two principles

Instead, our proposal depends on two principles:

1. Narrow and focused.
   1. Remove all references to "legal but harmful" material from the proposal and focus instead on more education and digital literacy;
   2. Limit the scope of a regulator to areas approved by Parliament so that it would have very narrow powers to oversee specific rules laid down by Parliament and proper accountability.

2. Law and order.
   1. Provide more resources to the police and law enforcement and investigate improving sentencing;
   2. Make specific, tightly defined and demonstrably harmful activities – like inciting minors to self-harm – unlawful (while avoiding possible unintended consequences).

The categories of harm

The categorisation of specific harms by the original White Paper is outlined in the following table (in its latest response the Government says it will set out a limited number of categories of harm in secondary legislation):

Table 1: Categories of specific harms in the Online Harms White Paper

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 innocent 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).

We propose a re-categorisation, then ways these categories might be addressed, below.

New categories

1. Illegal content and activity and its promotion

- Child sexual exploitation and abuse
- Terrorist content and activity
- Organised immigration crime
- Modern slavery
- Extreme pornography
- Revenge pornography (involving the posting of content without consent)
- Harassment and cyberstalking (where proscribed by the Malicious Communications Act 1998 and the Communications Act 2003)
- Incitement of violence (where proscribed)
- Sale of illegal goods/services
- Content illegally uploaded from prisons
- Sexting of indecent images by under-18s
- Encouraging or assisting suicide

2. Children accessing adult content

- Children accessing pornography
- Children accessing inappropriate material

3. Promotion of suicide and self-harm

- Advocacy of self-harm

4. Other harms

- Cyberbullying and trolling
- Extremist content and activity
- Coercive behaviour (unless blackmail, already defined by the Theft Act 1968)
- Intimidation
- Disinformation
- Violent content (unless “snuff films” or in the context of extreme illegal pornography, both above)
- Promotion of Female Genital Mutilation

Outline of an approach to these categories in a Bill and wider policy

For **category 1**, the harms listed are already illegal. The Government White Paper’s approach is essentially to add a form of intermediary or “messenger” liability for these activities separate from any liability of the perpetrator (i.e. to make online companies to some extent responsible for the actions of private citizens). But it is not clear what this will achieve, beyond creating the appearance of action by the Government. The preferable approach is to understand where improved resources for law enforcement agencies can help fight these crimes and to increase the sentences for them.

For **category 2**, the ultimate responsibility and decisions about the type of content accessible by children, *other than unlawful material*, should sit with parents. The state should avoid attempting to raise children. However, it can inform parents better about the potential risks of online content for children and the tools available to protect children from inappropriate content, such as age-filtering by ISPs and parental control software. Of course children should be prevented from accessing pornography, but for the Government to take responsibility away from parents may appear to be justified in the short term, but it is always liable to cause long-term harm, not least by giving parents the impression that the internet has been made safe so they no longer need to worry about keeping their children safe online.

For **category 3**, the promotion of suicide or self-harm, we acknowledge that there is a case for legislation to help prevent the physical harm of others, especially minors. However, the area is complex and it is difficult to legislate in this area without, as a side effect, making mere conversations about these subjects illegal. In this and other areas, the Government might consider the proper application of JS Mill's harm principle, which states that the only actions that should be legally prevented are those which cause harm to others: individuals should be free to do as they choose providing their actions do not harm other people. This does not include offense, or hurting people's feelings. As JS Mill put it in On Liberty:

> "That principle is, that the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection. That the only purpose for which power can be rightfully exercised over any member of a civilised community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant."

Mill explained in the same essay that the harm principle cannot justify infringements on free speech, most famously saying:

> "The peculiar evil of silencing the expression of an opinion is, that it is robbing the human race; posterity as well as the existing generation; those who dissent from the opinion, still more than those who hold it. If the opinion is right, they are deprived of the opportunity of exchanging error for truth: if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error."

This means that any legislation to prevent these physical harms should avoid, first, proscribing internet companies hosting discussions which do not see people actively encouraging others to self-harm or commit suicide. However, there is a case for proscribing the active encouragement of suicide as a form of incitement to violence, although the Government should consult widely before taking any such action.

For **category 4** (other harms), some of the activities on this list are already illegal, namely some forms of coercive behaviour such as blackmail, some intimidation and disinformation (including via the Malicious Communications Act and Communications Act 2003), and some violent content (such as extreme pornography). Other instances here are covered by the fact that encouraging or assisting in an offence is already itself an offence.

The question of "hate crime" in the categories

While the Government's White Paper included "Hate crime" under its category of "Harms with a clear definition", we do not include this in any of the categories above. We do not believe that any additional forms of speech should be categorised as hate crimes, and we are in principle against the suppression of free speech that is implied by the concept of "hate speech", which we believe should be rolled back, not extended (racially-aggravated violent crimes, for example, are a different matter). Because this refers to forms of expression that cause offence, this can also be defended by appealing to JS Mill's harm principle.

Sentencing and resources for law enforcement

Beyond legislation, improving the sentencing regime – in particular, tougher sentences – for individuals convicted of already-illegal acts above should be a priority for Government. It is not clear why these have not been considered as part of the solution to online harms. Tougher sentencing could be applied to a range of offences, including paedophilia.

Other areas of online crime where sentencing may be toughened up include creating and distributing terrorist content, and using the internet to organise modern slavery or immigration crime. For organised immigration crime, the Government should urgently review the state of current law, and a review would ask why the growing numbers of boat crossings on the English Channel are not being matched by a comparable increase in convictions of the people smugglers who operate in the UK, as well as how law enforcement agencies can be better resourced to monitor the use of the internet by these gangs.

To bolster the investigation of illegal online activity in general, better resourcing law enforcement agencies should be a priority, including regional police forces and the National Crime Agency (NCA), below.

One of the most important reasons why more resources are needed for law enforcement – and why the Government White Paper's attempt to suppress things that are already illegal will likely be ineffective – is that most genuine online harm, especially the distribution of paedophilic material and terrorist activity, occurs on the so-called "dark web", where the proposed regulation will have little effect. However, there are already agencies of law enforcement capable of investigating activity on the dark web, leading to prosecutions based on existing law. Again, it is unclear why the

Government is not keener to properly equip law enforcement agencies to do their jobs.

Reforms in 2013 merged the Police Central e-Crime Unit (PCeU), the Child Exploitation and Online Protection Centre (CEOP), and the Serious Organised Crime Agency into the National Crime Agency (NCA); law-breaking on the internet is also dealt with by specialised units within bodies such as the Medicines and Healthcare Products Regulatory Agency (MHRA) and HM Revenue and Customs. The NCA estimates that there may be 50,000 people in the UK involved in downloading and viewing indecent images online, yet the Government's proposals would impose bureaucratic responsibilities on intermediaries instead of giving the NCA more resources to tackle the problem.

Disinformation

The White Paper makes a number of claims about other online harms that are not backed up by proper evidence. The Free Speech Union's previous briefing outlined why the Government's claims about disinformation were supported by research that wasn't based on evidence. This included research that implied that North Koreans, allowed virtually no internet access, are safer online than Britons.

Of course there is "disinformation" online, in the sense that in any forum for debate there are things that are untrue, indeed mendacious. But one of the historic principles of our democracy is that contentious claims are best tested in the marketplace of ideas, not assessed by so-called experts. As Winston Churchill put it in 1902: "Nothing would be more fatal than for the Government of States to get into the hands of experts. Expert knowledge is limited knowledge... from the dominion of all specialists, good Lord deliver us."

There is a case for Government action on online disinformation, just not the action the White Paper envisages. When states such as China and Russia launch disinformation campaigns against the United Kingdom and the West, our Government might state what is taking place, and draw attention to the techniques and sources that are being used to spread disinformation, instead of attempting to vet or censor news and debate that is generated by its own citizens. This should be undertaken in an open and transparent manner. The Government already has the resources to help it do this, including GCHQ and smaller agencies like the Army's 77th Brigade, created in 2015 under the Army 2020 concept.

Instead of censoring online content that it regards as disinformation, Government should flag state-sponsored falsehoods and argue against them.

Another possibility would be a UK body taking *some* of the roles of the Reagan administration's Active Measures Working Group (AMWG) of 1981-1992. The main aim of the AMWG was to counter Soviet disinformation; it also alerted the US media and citizens to the activities of front groups and other Soviet measures.

AMWG exposed Moscow's *Operation Infektion* disinformation campaign, which claimed the US Government created the AIDS virus. An equivalent today would alert Britons to Russian and Chinese disinformation online, as well as to the groups and activities that adversaries like these are supporting in the UK. But the need is for *more* information, not less. It is dangerous to give actors within the state or sponsored by it the power to remove or block access to information. However well intended this might be, it is likely to have the side-effect of increasing the appeal of autocracies and conspiracy theories.

Government should stand up for free speech online, not undermine it

One of the most serious online harms our society faces is the increasing censorship of debate by internet giants such as Facebook, Twitter and YouTube. These are private firms and as such they have a right to decide what is on their platforms. Nevertheless, the removal of content does impoverish public debate. The Government's plans would make this situation worse.

For instance, YouTube might feel under an even greater obligation to remove any content challenging the official narrative of the WHO if it risks being fined if that content remains in place. The irony is that in censoring this alleged "misinformation", YouTube is contributing to its spread, because the WHO is sometimes simply wrong. Implementing the polices our Government will demand is therefore itself likely to create online harms. Nobody should be beyond challenge, whatever their credentials, and YouTube's approach is not anti-disinformation but ultimately anti-science.

China's cyber-war against western countries, including the UK, is having a corrosive effect on our democracy: Beijing has been described as waging "a battle for the future of the internet" as it attempts to normalise state

surveillance and deletion online, and turn the internet from a tool for free debate and information to a system for thought-control and the monitoring of individuals.

For example, Beijing sees the technological standards that determine how the internet is administered as "strategic weapons" and is attempting to propagate internationally its own standards to allow "granular control over citizens' net use" and a new "authoritarian web architecture"[7]. The US Government has warned employees of its armed forces not to use Chinese-owned video-sharing app *TikTok* due to risks of surveillance by Beijing (*TikTok* has also censored content that displeases China). Again, our Government's view on this genuine online harm is not yet clear, but a Government that is serious about reducing online harms would include a plan to address those directly perpetrated by China, Russia and other autocratic states.

The question of harassment

The Government is keen to deal with the issue of harassment online. One option that should be studied is an online version of the Injunction to Prevent Nuisance or Annoyance (IPNA), which has existed since 2015 and is similar to the previous Antisocial Behaviour Order (ASBO).

A court may impose an IPNA if it is satisfied that "a person has engaged, or threatens to engage, in antisocial behaviour" and "it is just and convenient to grant the injunction for the purpose of preventing that person from engaging in antisocial behaviour". IPNAs define antisocial behaviour as conduct that has caused, or is likely to cause, harassment, alarm or distress to any person; is capable of causing nuisance or annoyance to a person in relation to that person's occupation of residential premises; or is capable of causing housing-related nuisance or annoyance to any person. Importantly, IPNAs target individuals who are undertaking concerning behaviour, rather than putting problematic sweeping requirements onto all social media companies and their users.

7 *A Long March through the Institutions: Understanding and responding to China's influence in international organisations*. Radomir Tylecote and Robert Clark, Civitas, September 2020: https://www.civitas.org.uk/publications/a-long-march-through-the-institutions/

The Home Office could issue guidance on their use in digital environments. However, it will be important that any online version must only be issued for genuine and direct threats of harm, not provocative or offensive language, and should be deployed against serial offenders or abusers. Online IPNAs would require further analysis by the Government to prevent their being used to infringe free speech.

Should there be a specific regulator?

The Government has confirmed that it intends Ofcom to be the regulator under the proposed Online Safety Bill.

As we have stated previously, Ofcom has already demonstrated why it would be a poor choice. Ofcom recently sanctioned a TV channel for "risk[ing] undermining viewers' trust in advice from public authorities" by broadcasting an interview which it said might result in viewers harming themselves by ignoring social distancing rules. Its reaction demonstrates an implicitly authoritarian attitude. That Ofcom appears to believe that the media should not challenge the opinion of the state on public health is a serious concern. Furthermore, the head of Ofcom, Dame Melanie Dawes, recently said that on the issue of gender and transgender people, broadcasters should "steer their way through these debates without causing offence and without bringing inappropriate voices to the table",[8] suggesting that Ofcom believes the media should not cause offence and that only a small range of opinions should be broadcast.

In fact, there are already a number of regulators dealing with the internet. Ofcom has a number of online responsibilities, such as granting licences to online platforms that broadcast live TV. The Advertising Standards Authority (ASA) is the regulator of advertising across all media; the Information Commissioner's Office (ICO) deals with the General Data Protection Regulation (GDPR) and the use of personal data through cross-device tracking and profiling; and the Competition and Markets Authority (CMA) deals with online companies' approaches to competition. The existing legal system, through the police and courts, also deal with crimes that take place in the online context in the same way that they deal with crimes offline.

8 "Ofcom threatens diversity of opinion." *The Critic*, 23rd December 2020. Andrew Tettenborn. https://thecritic.co.uk/ofcom-threatens-diversity-of-opinion/

A specific regulator is not necessary. Our proposals would mean that online harms would be either criminal matters (such as use of the internet by terrorists or the distribution of indecent images) which would be the responsibility of the police and law enforcement agencies, or would be areas where the emphasis is on more research.

The "duty of care"

The White Paper proposes a "duty of care" for private companies to prevent harms that happen as a result of other people's conduct. This would in effect make companies responsible for how members of the public treat each other. There is no legal precedent for this; the way the Government uses the expression "duty of care" has little to do with the existing legal concept.

The duty of care would also mean the online environment, which has been able to free information and liberate debate, would in the UK become less free than the offline world. The lawyer Graham Smith has noted that this would also lead to a *de facto* parallel legal system for online content, and could mean the suppression of material the UK Supreme Court has held[9] should be not be suppressed. Online content that is perfectly legal in books or newspapers could be a breach of an internet company's duty of care.

The concept of the duty of care as it stands in the White Paper should be scrapped. There should also be a clear restatement of "safe harbour" under the e-commerce directive[10] as incorporated into UK domestic law under the European Union (Withdrawal) Act 2018.

In the offline world, companies can be liable if they create a specific risk to others which then causes real and measurable harm or if they break the law: publishing a book does not convey on the author or publisher a duty to ensure that its contents do not distress people, just as beyond obeying the licensing laws, a pub is not responsible for preventing its customers harassing each other.

9 In Rhodes v OPO [2015] UKSC 32.

10 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ("Directive on electronic commerce")

Digital platforms are already subject to the criminal law and have a duty of care under the common law, which could give rise to liability if they do not remove illegal images and content as soon as they have knowledge of them. There should not be a general legal duty to proactively and directly prevent other people from harming each other, especially when harm is defined as vaguely as it is in the Government's White Paper. Although well-meaning, alongside the risks to free expression outlined elsewhere in this paper, such a duty would undercut the concept of personal responsibility and de-emphasise and diminish criminals' responsibility for their crimes.

Terms and Conditions (T&Cs)

Terms and conditions (T&Cs) are not a matter for Government to either prescribe or enforce, as unlike community content guidelines they exist for the protection of the platform. To the extent that T&Cs are a consumer protection issue, there may be a role for existing consumer protection authorities (e.g. the Competition and Markets Authority), as well as the courts where T&Cs denote contractual responsibilities which are not upheld.

Safeguards

For the reasons set out above, we oppose the approach described in the Government's White Paper, including the imposition of both a "duty of care" and Codes of Practice on platform operators. However, if the Government is minded to proceed with this legislation, the protections for freedom of expression outlined in its proposals need to be strengthened significantly. In particular:

- The Online Safety Bill **should include protection of free speech as an overriding objective**, and should expressly seek to preserve freedom of expression, with **no obligation imposed upon internet companies not to publish views some might find offensive**.
- Alongside the legal duty on Ofcom to pay due regard to innovation, there should be a **general duty for Ofcom to pay due regard to freedom of expression** and association in all of its activities, including the production of codes of practice and enforcement action. There should be a **specific prohibition of any code of practice, enforcement activity, or other intervention** that would result in, or would be likely to encourage, the **removal by platforms of content that is protected under common law** or the ECHR.

- The economic impact assessments to be produced by Ofcom in respect of statutory codes of practice should be accompanied by **free expression impact assessments** to demonstrate specifically how freedom of expression and association have been protected and how the content of the code will not lead (intentionally or otherwise) to removal of content that is protected by common law or the ECHR.
- The **objectives of the regulator** and the **codes of practice** to be set out in primary legislation should **specifically refer to protection of freedom of expression** and association and direct the regulator not to seek to have platforms remove material that is protected by common law or the ECHR.
- **Disinformation and misinformation should be removed** from the scope of the duty of care, and of the legislation generally.
- Specific provision should be included in the legislation to the effect that if a platform has rules or community standards that are consistent with laws that protect speech, as stated above, the regulator will not enforce removal of content that is within those house rules, or otherwise penalise the platform for carrying it.
- There should be a mechanism whereby individuals whose appeals to platforms against the removal of content are rejected can **escalate their appeal to Ofcom** (similar to the role of the Information Commissioner in resolving data protection, subject access and freedom of information complaints), where the reason for the removal was the purported fulfilment by the platform of its statutory duty of care or observance of a code of practice.

Ofcom should be accountable for its work in this area to Parliamentary committees that are concerned with human rights in general and free speech in particular. In light of the increasing encroachment on free speech by regulation that has been actively supported by some Parliamentary committees, there may be a case for the creation of a Parliamentary committee that is specifically tasked with examining matters relating to freedom of expression, and in particular to hold public bodies and authorities to account in this context.[11]

11 We note that the Joint Committee on Human Rights is currently undertaking an inquiry into freedom of expression.

Given the duty imposed on platforms to provide ways to report content that breaches platform rules and standards, we welcome the fact that this will include a duty to provide for appeals against removal of content. However, we note that in the fast-moving context of social media and internet searches, reinstatement following a successful appeal will be of limited value, so it is vital that the duty of care and associated codes of practice do not incentivise removal of content by platforms that take a risk-averse approach. We also note that down-rating, deprioritising and de-monetising content, or otherwise limiting its reach, can also have the effect of stifling online speech, and should not be endorsed or mandated by any code of practice.

Of course, defending free speech online also requires the repeal of other laws not directly related to the internet but which will be the subjects of forthcoming Free Speech Union publications.


AUDI ALTERAM PARTEM
FSU
FREE SPEECH UNION