Free Speech Union briefing

Why the Government's plans to regulate the internet are a threat to free speech

Dr. Radomir Tylecote

September 2020

Contents

FSU research papers are designed to promote discussion of free speech issues. As with all FSU publications, the views expressed are those of the author(s) and not those of the FSU, its directors, Advisory Councils or other senior staff.

Summary: The Government's proposed new internet regulator will infringe free speech

- The Government published the Online Harms White Paper in April 2019, followed by an initial consultation response in February 2020. It intends to put a Bill before Parliament next year.
- Launching the White Paper, then Home Secretary Sajid Javid said he wanted the UK to have the "toughest" internet laws in the world.
- The proposed legislation's overall aim is to make the UK "the safest place in the world to go online, and the best place to start and grow a digital business". It states that "the digital economy urgently needs a new regulatory framework to improve our citizens' safety online".
- Some of the harms the White Paper identifies are real and merit new legislation, including the distribution of images of child abuse and online activities by terrorists. But these would be better dealt with by simpler legislation alongside better resources for law enforcement.
- However, some of the harms the White Paper describes are vague, such as "unacceptable content" and "disinformation". These are not fixed but would be determined by a future regulator. This will lead to sweeping censorship. Online Harms never properly defines "harm", which risks outsourcing the definition to activists and lobby groups.
- The White Paper says the new regulatory framework should prohibit material "that may directly or indirectly cause harm" even if it is "not necessarily illegal". In other words, the new regulator will be empowered to censor lawful content. As an example of what the Government has in mind, the White Paper singled out "offensive material", as if giving offense is a type of harm the public should be protected from by the state.

- The Government's legislative proposals claim to be partly inspired by Germany's 2017 "NetzDG" internet law, but Human Rights Watch has called for Germany to scrap the law, saying it "turns internet companies into censors". President Lukashenko of Belarus, Vladimir Putin's United Russia Party and the Venezuelan government have also cited NetzDG as the model for their internet laws.
- The authors of the White Paper take a similar view to the Chinese authorities on "disinformation", "social harms", and "authoritative" news sources, stating that the regulator should make "content which has been disputed by reputable fact-checking services less visible to users" and make companies promote "authoritative news sources", contradicting their claim that "the regulator will not be responsible for policing truth and accuracy online".
- While the authors of the White Paper believe their proposals will lead to more "tolerance" and less "hate", they will likely have the opposite effect, as people respond angrily to censorship and conspiracy theorists and hate-peddlers will enjoy the cachet of being banned by the state.
- In this briefing we explain why the Government's Online Harms plans are a danger to free speech and would mean a level of state control over the internet that most British people would find unacceptable. Later this year, the Free Speech Union will propose an alternative form of online regulation that will protect the vulnerable without jeopardising free speech.

The dangers of the Online Harms White Paper

The White Paper does not define harm

The White Paper never defines "harm", although it makes it clear that the intention is for the new regulatory framework to prohibit speech that is not illegal. But what some people regard as harmful is liable to change and is influenced by their ideological views, with some opinions being regarded as unacceptable simply because they challenge the views of progressive lobby groups. For instance, a recent petition on Change.org demanded that the Oxford English Dictionary add "transgender woman" to its definition of "woman". When a trans-critical feminist posted another petition on Change.org asking for the OED to preserve its current definition of woman the social media company removed it on the grounds that describing a woman as an "adult human female" is "hate speech". This illustrates that a failure to define "harm" or "hate speech" in the proposed new legislation risks outsourcing the definition to activists and lobby groups and introducing political censorship through the back door.

Many of the harms the new framework seeks to protect us from are already prohibited by existing legislation

There is already a wide body of law that prohibits many of the harms the White Paper is seeking to protect us from. Legislation covers: grossly offensive, purposefully annoying, or distressing speech (Malicious Communications Act 1988); speech intended to cause alarm, distress, or fear (Public Order Act 1986); speech intended to harass (Protection from Harassment Act 1997); and inciting hatred on the basis of race, religion or sexual orientation (Crime and Disorder Act 1998; Race and Religious Hatred Act 2006). The Law Commission has previously stated that the legislation covering communications is already expansive, that its breadth makes it flexible enough to cover offensive and abusive online communications and adapt to future developments, and that speech that is criminalised offline is also criminalised online.

The internet has provided another platform for people to commit these offenses, but in most cases it is not clear that it has increased the overall number of such crimes being committed.

There may be an argument for bringing this body of law into a single piece of legislation, and adding some additional harms not already prohibited by law, and even empowering a regulator to enforce these prohibitions. But that's not what the White Paper is proposing.

The White Paper will lead to the censorship of legal speech

The White Paper's authors aim to protect people from online material they regard as harmful. Some of the "harms" should be prohibited (and are), such as online terrorist activity, while others should not be the business of government, such as the expression of views the authors regard as not being sufficiently "authoritative". The failure to provide a precise legal definition of "harm" will inevitably lead to political censorship. White Papers generally deal with the law and propose ways in which it should be changed, but Online Harms is proposing that a regulator be empowered to prohibit "legal harms" (which are left for to the regulator to define). For a regulator to compel a company to prevent its customers saying things *that are legal* would be an extraordinary infringement on our freedoms.

The White Paper states that the "harms" the new regulator should prohibit should include "unacceptable content and activity", because "beyond illegal activity, other behaviour online also causes harm". This content need not even cause harm directly, but may do so "indirectly".

The White Paper gives a future regulator the theoretical flexibility to define almost anything as harm, because it provides a taxonomy of harmful content which is "neither exhaustive nor fixed" so as to allow "swift regulatory action to address new forms of online harm". The consultation states: "the list of harms provided was not... exhaustive… organisations suggested specific harms, for example misogyny".

The list of all harms is divided into three categories: a) "Harms with a clear definition", b) "Harms with a less clear definition", and c) "Underage exposure to legal content". Some of these harms should be prohibited – and are – while some should not be.

The first category includes drug dealing, harassment, and slavery. Most of these activities are already illegal and if the government wants to prohibit these collectively under a new law this could be done in a much simpler way (see *Areas of the White Paper that would be better addressed elsewhere*, below, for a discussion of how the Government could deal with the genuine harms mentioned in the White Paper).

The second category – "Harms with a less clear definition" – is more troubling. Although this includes areas that could also be addressed in simplified legislation (such as the facilitation of female genital mutilation), it also includes "Disinformation", "Intimidation" and "Extremist content".

In its discussion of "disinformation", the White Paper states that providers should: "help users understand the nature and reliability of the information they are receiving", to "minimise the spread of misleading and harmful disinformation". They should "[make] content which has been disputed by *reputable* fact-checking services less visible to users" and "[promote] *authoritative* news sources" (our italics).[1] It provides no clear definition of "extremist content" or "intimidation", meaning online commentary that is simply hostile towards public figures could easily be censored. It also overlooks the problem that "*reputable* fact-checking" often means fact-checking done by think tanks with a political agenda like the Institute of Strategic Dialogue.

The third category, "Underage exposure to legal content", includes serious concerns such as "Children accessing pornography", which are generally already unlawful or can be addressed in simplified legislation without infringing free speech.

No effective right of appeal for the victims of online censorship

The proposal in the White Paper is for a regulator – either an existing one, or a new one – to be given sweeping new powers to make sure tech companies are removing harmful contents from their online platforms. These will include being able to levy substantial fines, including on small companies, and will impose liability on senior management.

The White Paper discusses a "right of appeal", but in reality this will apply to the tech companies only – it will be hard for internet users to

1 The concept of "fact checking" literally means censorship insofar as the Latin term "censor" (censēre) means "to estimate, rate, assess, be of opinion" (OED).

appeal to the regulator if they don't think their content should have been removed, which may necessitate applying for a judicial review. Not only is that a lengthy and cumbersome procedure, but it would be almost impossible for such a legal challenge to succeed. How could Twitter or Facebook demonstrate to the High Court that a particular viewpoint won't under any circumstances cause harm, directly or indirectly, particularly when "harm" isn't defined?

Merely showing that the content in question hasn't caused the complainant any tangible harm won't be sufficient, since all the regulator will need to show is that it *may* cause them *indirect* harm. More or less anything falls into that category, particularly if the regulator can claim that causing offence is a form of harm.

Indeed, appealing any of the regulator's enforcement action will be rendered virtually impossible because of the plethora of undefined terms in the White Paper and which, as things stand, will find their way into the bill. Other undefined terms include "unacceptable content", "trolling", "intimidation" and "cyberbullying". Judgments about what forms of lawful speech are "unacceptable" or constitute "trolling", "intimidation" and "cyberbullying" are almost wholly subjective, yet these are all activities that the executives of social media companies will be expected to ban on pain of massive fines. It's a safe bet they will err on the side of caution – extreme caution, given how high the stakes are.

In light of the resources an online company will need to devote to complying with the new regulations if it is to avoid being penalised, the regulatory framework proposed in the White Paper will entrench the existing monopoly of tech giants like Google and Facebook and make it even more difficult for new internet companies to compete with them.

Online Harms would move the UK towards the internet laws of Russia, Belarus and China

It is especially worrying that the White Paper refers positively to Germany's Network Enforcement Act ("NetzDG"), which Human Rights Watch has called on Germany to reverse. Human Rights Watch says of the NetzDG that it "turns private companies into overzealous censors to avoid steep fines, leaving users with no judicial oversight or right to appeal", and is being copied by autocratic countries.

NetzDG imposes fines up to €50m for social media companies that fail to remove illegal content within 24 hours, incentivising them to censor legal content out of caution. Facebook alone has hired over 1,000 German-language censors to review content in "deletion centres" in Berlin and Essen.

In 2017, President of Belarus Alexander Lukashenko invoked the precedent of NetzDG to justify a crackdown on dissent and opposition, before passing a law against fake news that orders social media companies to moderate comments or face fines or geo-blocking.

The same year, Vladimir Putin's United Russia Party submitted a bill to the Russian Duma designed to force social media companies to remove unlawful content. Reporters Without Borders described the bill as a "copy-and-paste" of NetzDG and the explanatory report that accompanied the bill explicitly refers to it. Putin signed two laws in March 2019 outlawing "unreliable information online" which, like the Online Harms proposals, authorise an official watchdog to force online publications to remove content and order service providers to block access to sites if they fail to do so. Its definition of "false information" contains similar concepts to the White Paper, including information that supposedly endangers people's health.

In late 2017, Venezuela also passed a law called the "Law Against Hatred, for Tolerance and Peaceful Coexistence", imposing fines on social networks that fail to remove "hate speech". Before the law's adoption, then Vice-president of the National Constitutional Assembly, Elvis Amoroso, also referred explicitly to NetzDG.

Requiring a regulator, as an agency of state, to promote "authoritative news sources" (and by implication consider itself able to determine what these are) is similar to China's approach to internet regulation. The Cyberspace Administration of China justifies its internet censorship on the grounds that it is steering China's 1.1 billion citizens towards more reliable, authoritative sources of information. The White Paper often describes content the authors disapprove of as "harmful", without demonstrating why such content is actually harmful, a similar approach to China's censorship of "rumours" on the grounds that they could cause "social harms". The Chinese Government has also pushed online influencers to counter "disinformation" by promoting content it regards as more authoritative, holding education sessions such as the "Responsibility Forum for Online Personalities".

We would not dispute that some conspiracy theories that are disseminated online can cause tangible harm. But instead of taking Beijing's approach, the approach recommended by Justice Louis Brandeis is preferable: "If there be time to expose through discussion, the falsehoods and fallacies, to avert the evil by processes of education, the remedy to be applied is more speech, not enforced silence."

The proposed interventions against "disinformation" are not based on evidence

Some of the studies the White Paper cites to justify the claim that disinformation is a "threat" to the UK's civic institutions provide little or no evidence for this assertion. The White Paper gives pride of place to "A recent study from the University of Oxford's Computational Propaganda Project [which] has found evidence of organised social media manipulation campaigns in 48 countries in 2018". However, the five claims that study made about the UK weren't backed up by evidence and were, for the most part, unfalsifiable. The study also consistently found that the people of North Korea, who are allowed virtually no internet access, are less vulnerable to online manipulation (implying that in this respect North Korea has a "safer" internet than South Korea).

First, it claimed that: "formal organization between industry and political parties appears to have occurred… political parties and campaign managers have directly hired PR or consulting firms to help spread computational propaganda during elections". What "computational propaganda" means in the UK's case is not explained, although it appears to mean political campaigning, a necessary part of any democracy.

Second, in a table entitled "Organizational Form and Prevalence of Social Media Manipulation", the report claims that the categories of these in the UK are "Government agencies, politicians and parties, and private contractors", without defining what it means.

Third, in the table "Social Media Manipulation Strategies: Messaging and Valence", the UK is said to have "Pro-government or party messages, attacks on the opposition, distracting or neutral messages, and trolling or harassment", without evidence, or explanation as to why "pro-government or party messages and attacks on the opposition" do not fall under the heading of robust democratic debate. It found that Iran, North Korea and

Zimbabwe did not suffer “trolling or harassment” online and the country with the fewest of these harmful online activities is Nigeria.

Fourth, in the table “Observed Strategies for Social Media Manipulation” the UK was subjected to “Targeted ads, counter-info ops, and search engine optimisation”, again without evidence. It found no such strategies in North Korea, Saudi Arabia or Egypt.

Finally, it claimed that the UK has medium-rated “cyber troop capacity”, but its definition of these apparently sinister “cyber troops” included political parties’ PR staff. Egypt, Pakistan and Zimbabwe, by contrast, had few “cyber troops”. This is part of the evidence the government has compiled to justify interventions against “disinformation”.

The White Paper recommends that AI be used to censor legal speech

The White Paper states that: “We will work further with research organisations to understand how AI can best be used to detect, measure and counter online harms.” It describes the removal of “hateful content” as being within “the role of AI”, but states that a lack of “adequate data” has limited “our ability to develop more sophisticated and effective responses” and describes a project “setting out to address this issue”, specifically to “identify and categorise different types and strengths of hate speech”. (Will that include defining a woman as an “adult human female”?) Its ultimate aim is “for [AI] to be used to support a broad range of commercial and public sector providers to detect and address harmful and undesirable content”. That the White Paper fails to explain what it means by harmful or undesirable makes this a dangerous proposal.

More immediately, Online Harms recommends regulating the recommendations websites make to their users, meaning what information they are allowed to promote. The Government wants companies to take “reasonable steps” to ensure that users “will not receive recommendations to hateful or inappropriate content”, without explaining what “hateful” or “inappropriate” mean. Although this covers efforts to remove material that promotes self-harm, it is also likely to mean lawful content is steadily removed or ”shadow-banned”.

The “duty of care” in the White Paper is too broad

The White Paper imposes a “duty of care” on private companies to prevent harm that happens as a result of other people’s conduct, effectively making companies responsible for the way the public treat each other. This is an

onerous “duty”, particularly when internet companies can be penalised if the public engages in behaviour that is “not necessarily illegal” as a result of something they’ve seen online.

The White Paper threatens online privacy

Although the White Paper says the proposed new duty of care will not apply to “private channels”, e.g. direct messages, it also says “users should be protected from harmful content or behaviour wherever it occurs online” and the Government is currently consulting on how a “private channel” should be defined. This means the duty of care might be so broad as to require internet companies to monitor private conversations on platforms like WhatsApp.

The White Paper would punish lawful behaviour

The White Paper recommends that the new regulator should be able to impose swingeing fines on companies that fail to prohibit online harms – and that includes censoring lawful content. Punishments will include the capacity for the regulator, without legal process, to “disrupt the business activities” of internet companies and “force third party companies to withdraw any service [that] directly or indirectly facilitates access to the services of the first company”. Although this is intended to stop online terrorism, the meaning of the phrase “indirect facilitation” is unclear. The threat of these penalties will make the companies over-censorious as they err on the side of caution to avoid running afoul of their duty of care.

The White Paper wants the regulator to make “content which has been disputed by reputable fact-checking services less visible to users”. This contradicts the White Paper’s claim that “the regulator will not be responsible for policing truth and accuracy online”.

The regulator should not enforce companies’ terms and conditions

One of the ways the regulator will get companies to discharge their duty of care, as set out in the White Paper, is to insist that they “[uphold their] terms and conditions” with “enforcement action” if they do not. Community and company standards have long been regarded as a private matter by law. Companies’ standards on acceptable content also often go well beyond prohibiting unlawful speech, so this would be another way in which the regulator would compel social media companies to prohibit legal speech.

A new regulator risks giving the impression that children can be totally safe online

The White Paper's proposed new regulatory framework risks transferring primary responsibility for protecting children online from parents to the state. Despite the good intentions of government in this respect, it is important parents do not get the impression they can stop monitoring what their children are viewing on the internet.

The new regulatory framework seeks to protect children, but it will infantilise adults

Much of the discussion in the White Paper is about an online regulator's well-intended capacity to protect children, but while attempting to protect children, the regulator is likely to infantilise adults. The assumption that the state is a better judge than individual citizens of what is harmful or acceptable, what constitutes disinformation, or what should be said and not said, is authoritarian. As J.S. Mill said in *On Liberty*: "The peculiar evil of silencing the expression of an opinion is [robbing] 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."

Online Harms copies the EU's new model of censorship

The 2018 EU Code of Practice on Disinformation, agreed to by firms like Google and Facebook, describes disinformation as "verifiably false or misleading information" which (a) "Is created, presented and disseminated for economic gain or to intentionally deceive the public"; and (b) "May cause public harm", which means "threats to democratic, political and policymaking processes as well as public goods [such as] health". Companies must "dilute the visibility of disinformation by improving the findability of trustworthy content" and must invest in "technological means to prioritise relevant, authentic and *authoritative* information…" (our italics). The Code of Practice requires that "fact checkers" identify this authoritative content. However, fact-checking has been outsourced to activist think tanks and lobbyists who have proved much more likely in practice to flag right-wing views as "disinformation" rather than left-wing ones.

Many of the current Government's supporters will be surprised to see it embrace EU-style red tape, given that we have now left the EU. They might

also see an attempt by the civil servants who drafted the White Paper to stop the spread of opinions many are known to disagree with (like the "populism" they believe caused Brexit). The EU's Code of Practice is rooted in the false belief that many anti-EU electoral trends are the result of online "manipulation" rather than genuine discontent. If we replicate the EU's approach, it will allow governments to suppress challenges to their authority by designating them "extreme" or "hateful" instead of responding to them.

US and other foreign content could be censored or geo-blocked

The proposals are likely to lead to content from the US and elsewhere being blocked, censored, or made harder to find, as content providers attempt to shape content for the requirements of a UK regulator. Anyone who tries to follow US politics from the UK already encounters access problems due to the EU's General Data Protection Regulation (GDPR). The Government should remove these barriers, not erect more.

The Online Harms regulator would police thought

The Government wants the broadcast media regulator Ofcom to become the online regulator. Recent events have shown that Ofcom has too much power, not too little.

Ofcom recently sanctioned a television channel for "risk[ing] undermining viewers' trust in advice from public authorities", following its broadcast of an interview with David Icke which it said might result in viewers harming themselves by ignoring social distancing rules. This reaction is implicitly authoritarian. The vast majority of viewers are capable of deciding for themselves whether Icke's views are more trustworthy than the advice of public health authorities, yet Ofcom apparently believes that giving a platform to anyone who challenges the state's public health advice (which has often been wrong during the coronavirus crisis) is potentially harmful.

The White Paper is emotive and unprofessional

The White Paper often uses emotive, vague and unqualified terms that suggest its proposals have not been properly considered. For example, it claims that "There are too many stories of public figures closing their social media accounts following waves of abuse", without disclosing what would be an acceptable number. It states that "abuse… dissuades *good* people from going into public life" (our italics); and that "Adult [internet] users should act in an acceptable manner", without defining "good" or "acceptable".

Areas of the White Paper better addressed elsewhere

In a forthcoming paper we will outline how the sensible parts of the White Paper should be incorporated in new legislation that begins by carefully defining “harm”, and would protect the vulnerable without infringing free speech.

The current White Paper’s harms include, under the category “Harms with a clear definition”: child sexual exploitation and abuse, extreme pornography, encouraging or assisting suicide, sexting of indecent images by under-18s, and content illegally uploaded from prisons. Beneath the heading “Underage exposure to legal content” it includes promotion of female genital mutilation, and below “Underage exposure to legal content” it includes children accessing pornography. Some of these genuine online harms are already covered by legislation, but some are not, like the sale of opioids. These are serious concerns which are the proper subjects of simplified legislation.

Conclusions

The aim of the regulatory framework proposed in the White Paper is to make the UK “the safest place in the world to go online”, yet the White Paper never fully defines the harm the framework will purportedly prevent, which risks outsourcing the definition to activists. It proposes the effective censorship of legal speech in the name of addressing unacceptable content and preventing disinformation, requiring a regulator to block material that may “indirectly cause harm” even when that content is “not necessarily illegal”.

The framework proposed in the White Paper is partly inspired by Germany’s NetzDG internet law which has already inspired the internet security laws of Russia, Belarus, and Venezuela. The authors take a similar view to China on the subject of “disinformation”, “social harms”, and the encouragement of “authoritative” information, despite their claim that “the regulator will not be responsible for policing truth and accuracy online”.

The authors of the White Paper believe their proposals will lead to more “tolerance”, but they will likely have the opposite effect by allowing extremists to claim “free speech martyr” status. While some of the harms the White Paper describes are real, protecting internet users from these requires simpler legislation. In due course, the Free Speech Union will propose an alternative way of regulating the internet that will protect the vulnerable without infringing free speech.


AUDI ALTERAM PARTEM
FSU
FREE SPEECH UNION