

HOPE NOT HATE ONLINE SAFETY BILL REPORT STAGE BRIEFING – DON'T LET THE FAR RIGHT OFF THE HOOK

HOPE NOT HATE'S KEY ASKS AT REPORT STAGE:

- Remove loophole clause on journalistic content
- Categorise platforms by risk, not number of users

OTHER AREAS OF INTEREST IN THE BILL:

- Remove loophole clause on content of democratic importance
- Keep harmful content to adults
- Remove excessive Secretary of State powers to direct OFCOM
- Enshrine the right to anonymity in the Bill

Remove the Lennon Loophole - Support Amendment 10

Clause 16 of the Bill adds a duty on social media companies to protect journalistic content on their platforms. The current definition of 'journalistic content' is unspecific and could lead to a wide interpretation of the term.

Many of the key far-right figures HOPE not hate monitors self-define as journalists and could seek to exploit this loophole in the Bill and propagate hate online. Some of the most high profile and dangerous far-right figures in the UK, including Stephen Yaxley-Lennon (AKA Tommy Robinson) now class themselves as journalists. There are also far right and conspiracy theory "News Companies" such as Rebel Media and Urban Scoop.

These both replicate mainstream news publishers but are used to spread misinformation and discriminatory content. Many of these individuals and organisations have been deplatformed for consistently breaking the terms of service of major social media platforms, and this exemption could see them demand their return.

The Bill as it currently stands could provide a loophole for the far right and other harmful actors to use this exemption to propagate harm online. The loophole must be closed.

Categorise platforms by risk, not number of users – support New Clause 1

Schedule 11 of the Bill determines the categorisation of platforms, split between Category 1 services and Category 2a (for search services) and Category 2b (for user-to-user services) by size and functionality. Category 1 platforms would have extra duties placed on them to deal with content that is harmful but not illegal.

At present, the definition of functionality in the legislation does not include risk, which is a huge problem. At HOPE not hate we know that social media space has changed and a huge amount of the most harmful behaviour happens on small and alt-tech platforms. If this bill is to properly address the issue of online harms it has to reflect the real nature of the online space and take into account not just the size of a platform, but also the risk that it poses.

In recent years, far-right figures have begun to migrate to alternative and usually smaller platforms. The result is that there are now broadly three categories of social media platforms used by the far right:

The first are **mainstream platforms**: those that are widely used by all across society, such as Facebook, Twitter, Instagram, Youtube and TikTok. While these platforms all have an extremism problem, they generally have terms and policies that prohibit extreme and discriminatory behaviour, even if they don't always enact them as consistently as necessary. Where possible the far right want to remain on these platforms, as they afford huge audiences beyond existing supporter bases. This is where they want to propagandise and recruit.

Next are **co-opted platforms**: those not created for or by the far right, but which have become widely used by them, either because of loose policies, a lack of moderation, or a libertarian attitude towards deplatforming and content removal. Most notable is Telegram, which is an enormous social media app with over one billion downloads globally. Due to its consistent failure

to remove extremist activity, it has become a crucial hub for the contemporary far right. The danger for the far right with these platforms is that they may eventually choose to clean up their act and remove illegal or harmful content, making them insecure homes in the long term.

The final category is BESPOKE PLATFORMS: a growing group of platforms, created by the far right or by people consciously courting extremists. Many of these are essentially clones of major platforms, but featuring little or no moderation. The best known are Gab, BitChute and most recently, GETTR.

For OFCOMs' priorities to be directed at the highest harm platforms, this schedule must be amended.

Remove the democratic content loophole

Similarly to journalistic content, one of HOPE not hate's major concerns about the Bill is that, at present, the vague protections of "democratically important" content could again open up the opportunity for abuse by far-right activists and organisations. This duty could enable a far-right activist who is either standing in an election, or potentially even for just supporting candidates in elections, to use all social media platforms. This could again mean far right figures being 'replatformed' onto social media sites where they would be free to continue spreading hate.

It is vital that any discussion about how this Bill protects democratic speech goes beyond limiting censorship, and includes the promotion of a genuinely pluralistic online space. This demands an analysis of the voices that are so often missing or marginalised online, namely the voices of minority and persecuted communities. We will only create a genuinely democratic online space by broadening out the definition of "democratically important" to include not just content that is often removed, but also content that is missing in the first place. It cannot just protect existing "democratically important" speech, it must also create a safe and pluralistic online space that encourages and empowers diverse and marginalised voices, enabling them to be heard.

Currently, the Bill indicates that content will be protected if created by a political party ahead of a vote in Parliament, election or referendum, or campaigning on a live political issue. Will this clause mean that far-right figures who have already been deplatformed for hate speech must be reinstated if they stand in an election? Does

this include far-right or even neo-Nazi political parties?

Again, content and accounts that have been deplatformed from mainstream platforms for breaking terms of service should not be allowed to return to these platforms via this potential loophole.

HARMFUL BUT LEGAL CASE STUDIES

Below are a series of short case studies produced by people concerned about ensuring the inclusion of legal but harmful content within the bill. As noted above, it is important to state that the inclusion of legal but harmful content within the scope of this legislation does in no way make the harmful speech, outlined in the case studies below, illegal. In most instances, it will remain legal to deny the Holocaust, to push Covid-19 conspiracy theories, and to be racist and misogynist. However, by demanding better platforms, with design goals in line with our shared values and processes that empower governments, civil society and a platform's users, this legislation will reduce the spread of, volume of and rewards for harmful behaviour.

COVID-19 CONSPIRACY THEORIES AND ANTISEMITISM

Danny Stone, Antisemitism Policy Trust

Anti-Jewish racism is not static. Over millennia, it has evolved and adapted to societal circumstances. Whether it was bubonic plague in the 14th century, tuberculosis in the 19th century or typhus in the 20th century, Jews have been scapegoated by conspiracists and extremists for disease. It was therefore sadly inevitable that Jews would be blamed in some circles for COVID-19 in the 21st century. Antisemitic scapegoating surfaced and spread during the global pandemic, providing ammunition to those seeking someone to blame. Online, memes have been circulating espousing antisemitism, whilst offline, several public figures and others in the public eye have alluded to Jews being the cause of the pandemic. Much of the material, comment and discussion is legal but harmful.

Conspiracy theories that the virus is real and a way for Jews to expand an imaginary web of global influence and profit, or the accusations that the virus is in fact a fake Jewish conspiracy are not necessarily illegal, but they can lead, and have led, to significant isolation of a minority

community and real-world harm. Engagement with the aforementioned conspiracy theories easily relates in online spaces to other sinister narratives, for example that the Jews are the 'real virus' and its primary spreaders, the celebration of Jewish death, or even the so-called 'Holocough', a call to spread the virus to Jews.

Whether it be [a caricature](#) of a former French and Jewish health minister pouring poison into a well receiving tens of thousands of social media shares, protesters in Ohio holding up a sign during an April 2020 rally depicting Jews as rats and claiming they are 'the real plague', the refusal of services to visibly identifiable Jewish people for [supposed Covid culpability](#), or the sale of Holocaust [memorabilia](#), the consequential harms of antisemitic conspiratorial narratives are manifold. The space for Jews or other minorities targeted by such conspiracy theories is reduced, the tremendous efforts to help people during the pandemic become a greater task for fear of personal safety, the potential threat and requirement to meet it become the burden of Jewish people worldwide, and the online space becomes more toxic and hateful. Legal but harmful material does not exist in a silo, it has consequences, and we have a societal responsibility to protect against this.

FOOTBALL AND RACIST AND MISOGYNIST ABUSE

Hilary Watson, Glitch

The increased prevalence of online abuse in football is receiving unprecedented media coverage, culminating to date in the racist abuse targeting several Black England footballers after the Euro 2020 Final. Women's football is also blighted by online abuse, where Black and minoritised players are racialised, subjected to faith-based and homophobic online abuse, alongside an exorbitant level of sexist abuse that is commonly experienced by women in public life and especially prevalent for sports women. Removing the 'legal but harmful' obligations currently in the draft Online Safety Bill does not move us forward from our experiences now.

The status quo, where online hate is passed off as 'banter' and only a small proportion of incidents of online abuse are followed up by law enforcement has meant that social media platform self-regulation not only allows legal but harmful content but there is also very little consequence for content that does reach the legal threshold for hate crime and harassment,

with so few incidents pursued by the criminal justice system or brought to justice.

The level of online hate experienced by footballers in the UK is not going to go away without increased accountability, transparency and support from social media companies, as well as governments. Several social media boycotts to raise awareness of online abuse, including the four-day social media boycott across professional football in the UK in May 2021, has brought little change.

The current system is not sustainable, where social media companies put the onus on people in the public eye to use features on their platforms to shield themselves from seeing the worst of the abuse directed at them. It should not remain on the platform for every other user to see either.

As we have seen from the fallout of the success of the England football team in the Euro 2020 tournament during a hugely successful campaign where the team is more racially diverse than ever before, social media amplifies hatred but does not provide adequate tools for anti-racism work or allow for meaningful active online bystander intervention – unnuanced automated moderation systems currently in place can mean that talking about abusive behaviour gets would-be activists censored or blocked from the platform.

Offline, Marcus Rashford's wall mural on the streets of Manchester was defaced with racial hatred, yet community response led to a powerful and unifying act of anti-racist action. We must have the opportunities and tools to clean up our online spaces as we do our offline spaces.

Far too many people are being harmed online and platforms are becoming increasingly unsafe. Football boycotts have not brought meaningful, lasting change. The Online Abuse Bill must use a multi-pronged approach to improving the safety of all users through regulations on both illegal content and legal but harmful content. We must ensure social media companies are obliged to make their platforms safe for everyone.

ELECTION DISINFORMATION

Poppy Wood, Reset

There are already a number of examples of state backed disinformation campaigns inauthentically amplifying partisan views on Scottish referendum. Twitter and Facebook have identified swathes of fake accounts linked to the governments of Russia and Iran which

[amplify messages from bogus pro-independence campaigners](#). Facebook has also identified, and deleted, a page called [Free Scotland 2014 which was traced back to Iran](#) and linked to fake news sites. This is just the tip of the iceberg and it is completely plausible that such efforts to skew electoral processes could result in the ramifications (both offline and online) that we witnessed in the US following the recent Presidential election.

The content shared by these groups and pages was either intentionally misleading (false stories about the Royal family) or intentionally divisive (calling for a change of Conservative party leadership). It is likely that all of the content would have been legal. At present it remains unclear whether election misinformation would be included as a 'harm' in this legislation, but the removal of legal but harmful from this legislation would make it even less likely that it would be dealt with.

ABUSE OF PUBLIC FIGURES

Lexie Kirkconnell-Kawana, IMPRESS

Beyond the scale of abuse experienced by public figures; evidence has been collated on the impact system design has on legal but harmful speech directed at public figures (particularly female politicians) in the Committee for Standards in Public Life 2018 [report](#):

“The first thing we do in the morning is to block and delete online abuse, usually whilst having breakfast. Porridge with one hand, deleting abuse with the other.”
Office of Rt Hon Diane Abbott MP; Staff of Diane Abbott MP, Individual Oral Evidence, 1 November 2017.

“Social media also bleeds into your 24 hours home life, at night the tweets come in when you're cooking your kids' tea or going to bed. There is little place to hide.”
Lisa Robillard Webb

That speech may be insulting, sexualised, racialised, imply threat or reveal personal and sensitive information, and while individual instances of this speech may amount to a) malicious communication (criminally liable under MCA 1988) or b) harassment (criminally liable or actionable under the PHA 1997), most of this speech does not reach the threshold for prosecution set out in CPS guidance, and is therefore legal. The purpose of this speech is not to engage in legitimate democratic debate

or engage policy decision makers on issues affecting individuals and communities, but to punish, silence and broadly discourage those likely to be subject to discrimination and abuse from participating in public life. While platforms can engage in moderation and takedown and individual accounts can report instances of this speech, the system design of platforms actively encourages speech patterns which target public figures:

“Extreme positions whether political or moral or abusive, you will get a rise in followers. There is an incentive to go to the extreme.” Lionel Barber, Editor of the Financial Times, Individual Oral Evidence, 30 October 2017.

The democratic speech exemption gives the 'bad actor' and the 'poor system design' a legal shield, as all speech - no matter how insulting, offensive, abusive, or objectionable - directed at a public figure could be construed as civic participation, and therefore as democratic content. Should the online safety bill withdraw legal but harmful categories of speech from scope, alongside carving out a protection for democratic content, this could lead to a perverse consequence where platforms are legally disincentivised from moderating speech directed at public figures (particularly political figures), or otherwise designing their systems in such a way that limits or mitigates its impact.

ANTI-VAX AND DEEP STATE CONSPIRACY THEORY MISINFORMATION

Lexie Kirkconnell-Kawana, IMPRESS

There is no law in the UK that criminalises the publication of inaccurate information (outside of civil liability for reputational inaccuracy, i.e., defamation or malicious falsehood), whether the publication is intentional or accidental. Inaccuracies about innocuous subjects are unlikely to require regulatory intervention, but inaccuracy about matters integral to civic participation and public health and safety have profound effects on social cohesion and daily life. There is plenty of documented evidence of harm that stems from inaccurate information, particularly at scale and enabled by system design. The Institute for Strategic Dialogue in a recent [study](#) concerning Germany and Covid-19 information, were able to find 400,000 pieces of false Covid-19 information from more than 1,000 social media users and found that

targeted influencing by these actors of the public vaccination debate has been successful in parts of the population.

Similarly, NewsGuard has [documented](#) the rise in growth of QAnon groups in the UK, who seek to influence public debate by sowing distrust, targeting politicians and public figures and encouraging the public to withdraw from democratic processes. A series of QAnon Facebook groups appeared in the UK and Ireland starting in April 2020: The Great Awakening – the History of Everything (Cabal. Q. Qanon), a group with 18,200 members was created in April 2020, followed by Q-UK, a group with 2,500 members, created in May 2020. One of the largest UK QAnon Facebook groups (over 9,600), UK Patriot Alliance, was launched in June 2018; its About page stated: “This group is a place to come and learn, wake up the British People and spread the message of #Q.” A Twitter account, run by Martin Geddes a London-based technology consultant and photographer who joined Twitter in December 2007, boasted 188,000 followers and since 2018, Geddes has regularly tweeted and retweeted QAnon content.

There is no legal mechanism by which to prevent these actors operating or prevent them from publishing this false information in the UK. Therefore, all information published across these accounts is legal. If the legal but harmful category were to be removed from the Bill, there would be no regulatory incentive for platforms to mitigate the impact of this content as part of their system design.

MEDICAL MISINFORMATION

David Lawrence, HOPE not hate

Few circumstances impress the need for accurate medical information as much as a global pandemic. However, since the onset of COVID-19 there has been an explosion of conspiracy theory-driven medical misinformation online, ranging from claims that 5G technology is causing the health crisis, to the allegation that vaccines are being used to insert microchips into, or to simply kill, recipients. Such scare stories carry potential health implications for both individuals and for society as a whole.

Of course, there is much reasonable discussion to be had about the potential health implications of vaccines, or disputing the efficacy of the current advice from health bodies. There have been many times when a medical product

that has been allowed onto the market has subsequently been shown to be harmful, due in part to a lack of transparency.

However, flagrant, conspiracy theory-driven medical misinformation, such as propaganda likening vaccination efforts to Nazi death camps, or alleging that NHS nurses should stand trial for genocide, clearly falls into the category of legal but harmful.

HOLOCAUST DENIAL

Joe Mulhall, HOPE not hate

Since Germany’s military defeat in 1945, Holocaust denial has been an attempt by surviving unreconstructed Nazis and their postwar acolytes to whitewash the monstrous crimes of the Third Reich in the hope of rehabilitating the Nazi regime.

Recognising the internet’s potential for reaching people at an unprecedented scale, Holocaust deniers were early adopters of online platforms, some as early as the 1980s. Since the 2000s, the spread of social media has had a profound impact, not just on the ability of the denial community to spread their ideas, but more fundamentally on the idea of, and motivation for, Holocaust denial itself.

Holocaust denial is a form of antisemitism and is used to attack and cause harm towards Jews. The idea that Jews have lied about the Holocaust, often, it is argued, to legitimise the State of Israel, has resulted in abuse and attacks against the Jewish community.

While illegal in some parts of Europe, Holocaust denial remains legal in the UK. The inclusion of legal but harmful content within the scope of this legislation would not change that in any way. It would merely ensure that Holocaust denial is not spread on social media platforms where it can cause harm.

GENDERED DISINFORMATION

Ellen Judson, Demos

Gendered disinformation campaigns manipulate information and weaponise gendered stereotypes to discredit people – particularly women in public life – on the basis of their gender and to undermine the fight for gender justice. It is often used as part of a broader political strategy, manipulating existing gendered narratives to silence critics and consolidate power.

Gendered disinformation commonly weaponises stereotypes such as women being devious, stupid, weak, or immoral; or sexualising them to paint them as untrustworthy and unfit to hold positions of power, a public profile or influence within society. Gendered disinformation is a global phenomenon - it is a pattern observed across the world (including in the UK and Europe).¹ These campaigns can start online or offline, on social media or in mainstream media, but the possibility for amplification and scale of gendered disinformation campaigns on online platforms makes them a particular cause for concern. Examples of classic gendered disinformation tactics include faked nude images, videos and sexualised rumours being shared online about women parliamentarians;² harassment, abuse and threats targeting women journalists³; and campaigns playing on tropes of the role of 'mothers' and the 'family' in society to justify curtailing the rights of women and LGBT+ people.⁴

These campaigns not only take a profound psychological toll on their targets: but threaten their personal safety, with campaigns seeking to justify detention and violence against women in public life and activists for women's and LGBT+ rights.⁵

They also pose a wider democratic threat - gendered disinformation undermines equal participation in democratic life: seeking to silence existing political critique, trying to prevent opponents being elected, and reducing the space in which it is safe for women and LGBT+ people to be involved in public life.⁶ This in turn undermines the effectiveness, the equity, and the representativeness of democratic institutions.

There are several factors, however, which mean that gendered disinformation online is likely to fall into the category of 'legal but harmful' rather than illegal speech:

1. It is gendered: in the UK, gender is not a category that can define illegal hate speech.
2. It operates at scale: individual posts which would not meet a criminal threshold can together form a gendered disinformation campaign. e:
3. Though gendered disinformation often weaponises highly abusive language, it is not always so clearly identifiable: it can appear similar to many different forms of legitimate and protected speech: it masquerades as political critique, invokes political beliefs, common views and stereotypes, or can appear similar to counterspeech or satire.⁷

4. Gendered disinformation evolves to take account of the context and environment in which it operates: currently evading algorithmic detection through deliberate misspellings or coded imagery.⁸ Even if some forms of gendered disinformation became illegal, those who engage in these campaigns would identify where the line was and seek to avoid it, and be able to exploit any ambiguities

As such, relying only on methods to combat individual pieces of illegal content online will be unlikely to greatly reduce the risks associated with gendered disinformation.

Endnotes

- 1 <https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat>
- 2 <https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/>; <https://www.brookings.edu/techstream/gendered-disinformation-is-a-national-security-problem/>; <https://demos.co.uk/project/engendering-hate-the-contours-of-state-aligned-gendered-disinformation-online/>
- 3 <https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat>
- 4 <https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat>
- 5 <https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat>; <https://demos.co.uk/project/engendering-hate-the-contours-of-state-aligned-gendered-disinformation-online/>
- 6 <https://www.brookings.edu/techstream/gendered-disinformation-is-a-national-security-problem/>
- 7 <https://demos.co.uk/project/engendering-hate-the-contours-of-state-aligned-gendered-disinformation-online/>
- 8 <https://www.wilsoncenter.org/publication/malign-creativity-how-gender-sex-and-lies-are-weaponized-against-women-online>