



## The Online Safety Bill – What's Changed?

By Matthew Feeney

### Introduction

Earlier this week, the Department for Digital, Culture, Media and Sport (DCMS) announced a series of changes to the Online Safety Bill. The changes address some of the most prominent criticisms of the Bill and introduce new provisions.

Some of the reforms to the Bill are welcome, but fundamental flaws remain in the legislation. While it is reassuring that the government has implemented changes to the Bill, the legislation continues to pose a threat to free speech, privacy, and competition.

This bulletin analyses some of the most noteworthy changes and notes what problems remain with the Online Safety Bill.

### What is changing?

#### *'Legal but harmful'*

Perhaps the most notable change to the Bill is the removal of the so-called 'legal but harmful' content clause. Free speech groups and many others (including the Centre for Policy Studies) have been critical of this provision, describing it as an unintentional threat to free speech.

Under the previous version of the Bill, large online firms would have had to disclose how they planned to treat 'legal but harmful' content. The Bill did not require that firms remove such content. Rather, the Bill required firms to commit to removing such content; restricting access to it; limiting the recommendation of such content; and/or recommending or promoting the content.

As we noted in the CPS paper *A Censor's Charter? The case against the Online Safety Bill*, 'given that the largest social media firms already take steps to moderate speech that is likely to be included in the "legal but harmful" category (such as misogynistic abuse or content associated with eating disorders), it is unlikely that they will choose to recommend or promote such content. They will therefore be put in a position of pledging to moderate such content and face significant fines if they fail to act on it swiftly.'

If passed into law, the clause – and the multi-billion-pound fines that accompanied it – would have encouraged large online firms to remove troves of content, erring on the side of caution in order to avoid the Bill's significant penalties. So, removing the clause is a welcome change.



### *Harmful communications*

The previous version of the Bill introduced a new criminal offence, the ‘Harmful Communication Offence’. New Culture Secretary Michelle Donelan has chosen to remove this offence, which would have criminalised ‘sending a message if at the time of the sending “there was a real and substantial risk that it would cause harm to a likely audience” and the sender of the message “intended to cause harm to a likely audience”.’ The Bill defined ‘harm’ as ‘psychological harm amounting to at least serious distress.’

This is a welcome development. As we noted before, it was always wrong to treat online and offline speech in different ways. And crucially, the threshold for breaching this new law was entirely about the audience rather than the sender or their intention: if ‘several’ or ‘many’ people were in the ‘likely audience’, then it would not have mattered whether the original composer of the message intended harm. As we noted, this would have allowed for the prosecution of the message’s composer in cases where a social media post went viral, even if they never intended to cause distress.

So again, the removal of this offence is a good first step. Yet the Secretary of State also promised that the updated version of the Bill will abandon the plan to scrap Section 127 of the Communications Act. This is concerning.

Thanks to Section 127, those who upload social media posts deemed ‘grossly offensive or of an indecent, obscene or menacing character’ (such as racist tweets) can face jail time. Those convicted under Section 127 include a man who sent an offensive tweet about NHS fundraiser Captain Sir Tom Moore and a YouTuber who filmed himself urging his girlfriend’s dog to perform Nazi salutes in response to commands such as ‘Sieg Heil!’ and ‘Gas the Jews!’. These are not things that any one of us would do: they are shocking and objectionable actions. Yet a commitment to free speech can require tolerance of objectionable, offensive and upsetting speech.

The Government has sought time and again to insist that the Bill is not a threat to free speech. Yet by abandoning its plans to repeal Section 127, it weakens its own claims about the Bill’s effects on free speech, and returns us to a situation in which different legislation applies to the same speech online and off. As we noted previously, the legislation also means that it is a regulator, Ofcom, which assumes the role of warden of acceptable and lawful speech rather than these issues being properly scrutinised by Parliament.

### *The ‘triple shield’*

The new version of the Online Safety Bill will include what the Government describes as a ‘triple shield’, made up of obligations on social media firms to 1) remove illegal content, 2) take down material that violates their own terms of service, and 3) provide adults with greater control over the content they access.

The obligation to remove illegal content is not objectionable on its face, but the details are important. If removals must take place within a tight time window, then we should expect online



firms to embrace false positives and take down thousands if not millions of pieces of legal content, out of an abundance of caution.

The terms of service provision is also worrying. Given the plethora of user-generated content uploaded to social media platforms every minute, it is impossible for even the best-funded firms to apply their terms of service consistently. In addition, firms change their content moderation policies in light of discovered nuance, news events, and emerging threats. Because of the nature of content moderation at scale, any obligation to remove content that is associated with massive fines will result in firms inadvertently removing content that does not violate their terms of service.

Popular social media platforms already allow users to mute words, block and unfollow users, and provide warnings for sexually explicit content. Forcing online firms to provide users with more content filters creates a risk that such a requirement will be overinclusive and that online firms will harm user experience while seeking to be compliant. For example, a filter for 'suicide' or 'terrorism' content might prevent a user from seeing upsetting graphic images, but might also reduce exposure to valuable news items, debate and discussion. Firms may decide to implement something like 'safe mode' by default, thereby requiring users to opt into particular content, which will inevitably shrink the universe of legal speech available to them.

#### *Age assurance*

The updated version of the Bill will require platforms that have a minimum age policy to outline how they will enforce this policy. The government cited age verification as one such method.

Age assurance online is often discussed in the context of pornographic content providers, but prominent social media firms such as Meta also require users to reach a particular age before creating an account. For example, Facebook bans those younger than 13 from joining.

Such a requirement may sound extremely sensible, but it also poses risks to user privacy, because the flipside of age assurance is an end to anonymity. Pseudonymous and anonymous speech can provide valuable information to the public, especially when it informs users about data released by whistleblowers and updates from war zones such as eastern Ukraine. Many people running anonymous or pseudonymous accounts will be understandably hesitant to volunteer personal information to an online platform. And even if the platforms themselves have no intention of exposing whistleblowers, activists, and journalists, they risk being increasingly targeted by foreign adversaries and criminals who are interested in learning more about such users.



## What Remains the Same?

Despite the changes to the Bill outlined above, most of the concerns outlined in our two previous CPS papers, *A Censor's Charter?* and *Safety Without Censorship*, still remain.

### *Speech*

The announcement of the 'legal but harmful' provision being scrapped for adults is welcome, but the Government apparently does not intend to remove the obligations on large firms to prevent children from accessing 'content that is harmful to children'. Because no one needs to prove their age to surf the web, it is impossible for online firms to know which of their users is a child and which is an adult. Firms will therefore likely continue to reduce user access to 'legal but harmful' content in order to comply with the Bill's child protection provisions. The age assurance provisions referred to above may ameliorate this, but there is a vast universe of content that is available without being logged in to a particular site or service, all of which might fall under the provisions of the Bill.

### *Privacy*

The Government's press release does not mention encryption, which is concerning. The CPS and many other groups have noted that the Bill poses a significant threat to encrypted messaging services such as WhatsApp, iMessage, Signal and others. The Government should change the Bill to make clear that it does not mandate the weakening or the banning of end-to-end encryption. If the Bill results in online services scrapping encrypted messaging services, the privacy and security of millions of law-abiding British citizens will be put at risk. There could also be a significant cost to the economy.

### *Competition*

The changes to the Bill fail to significantly reduce the burdens on online firms, which will have to dedicate significant resources to compliance. Household name firms such as Meta, Alphabet, and others in the 'Big Tech' club will, thanks to their powerful market positions, be best placed to comply with the Bill's obligations. As the Institute for Economic Affairs has noted, the government's estimated £2.5 billion cost of implementation over 10 years is likely a significant underestimation. Unfortunately, the announced changes to the Bill will do little to mitigate its anti-competitive effects.

## Conclusion

Despite some welcome changes, the Bill continues to pose a threat to free speech, privacy and competition. We are grateful that the Government has listened to its critics, but urge it to do more to ensure that free speech is at the heart of the Bill.

*Matthew Feeney is Head of Tech and Innovation at the Centre for Policy Studies*