# THE ONLINE SAFETY BILL – SECTOR SUPPORT ROUNDTABLE

Summary Report

October 2022

1. On 27 September 2022, Open Rights Group convened a roundtable to discuss how the Online Safety Bill’s draft to date provides more cause for concern and fails to fulfill its mission of reducing harm experienced online.

2. Two particular issues – content moderation and encryption – were raised during the event.

3. Dr. Edina Harbinger, a Reader in law at Aston University, Birmingham and Alec Muffett, a network security expert, addressed attendees to provide an overview of the content moderation and surveillance of private messaging and encryption and the proposed Bill’s impact on these two issues.

4. Dr. Monica Horton, Open Rights Group’s freedom of expression policy manager, chaired the discussion with a range of civil society representatives in attendance, including:

   Halaleh Taheri, Middle East Women and Society Organisation
   Hera Hassan, Chayn
   Hilary Watson, Glitch
   Azfar Shafi, CAGE
   Jun Pang, Liberty
   Matthew Johnson, Runnymede Trust
   Penny Wangari-J, Racial Justice Network
   Carys, Stop the Scan Campaign, Racial Justice Network
   Mallika Balakrishna, Migrants Organise
   Zehrah Hasan, Joint Council for the Welfare of Immigrants
   Diana Gheorghiu, Child Rights International Network
   Jen Persson, Defend Digital Me
   Ayesha Saran, Barrow Cadbury
   Simmi Bhagri, Borkowski
   Participant 15 (Anon)

## The balance in the Bill

5. The chair summarised how since the dawn of the Internet, the sudden potential of increased access to content – which comes with many benefits – has caused commercial disputes, such as copyright infringements and social harms from content that steps into what’s deemed illegal or that steps beyond a social norm, causing friction with the right to freedom of expression. Rules have had to catch up with the internet age and with the balancing exercise needed to address the tension continuing today, one thing Open Rights Groups hopes to find out is whether civil society perceives the balance in the Bill to be right.

6. The caveat to the discussion is that the recent change in leadership over the country leaves some of the proposals’ longevity in question and uncertainty over the final version of the Bill.

### Illegal content

7. Harbinger explained that Illegal content in the current draft bill is divided into priority illegal content, such as child sexual exploitation and terrorist content as well as other priority content listed in Schedule 7 of the Bill.

8. The moderation of such content will necessitate the use of technology to help with filtering, which is very problematic as seen via the EU example and its e-commerce directive, which prohibits the use of technology to monitor user content. This will likely occur, particularly in relation to child sexual abuse and terrorist content. Doing so will raise concerns over free speech and removal of that content or the exercise of restraint before people post online.

9. Privacy and data protection concerns also arise where these technologies may may include violations of users' privacy with implications for encryption and encrypted services as even private messaging channels will be subject to monitoring.

### Legal but harmful

10. While there is a list of recognised illegal content, less clear is the proposed management of legal but harmful content. The Secretary of State may designate content as harmful in a Regulation that can be approved by Parliament, and the list of content to be included is growing.

11. In this category there is also malicious communications, harmful communications and false communications that are new offences included in the bill.

12. Platforms will be required to judge whether there is a reasonable excuse for users to post a harmful communication that might offend and cause harm to another, including to their psychological health as well as physical harm. If the user has no defence, the platform will conclude that it is harmful communication, so will need to take the content down.

13. The question is how well will platforms convert the understanding that offence and serious distress has been caused into a content moderation algorithm. Do they have the required nuances? How will they consider external factors they are not aware of and that the law requires? What about free speech concerns?

14. The other question is if something is harmful enough to be taken down from services, why is that thing not criminalised and deemed illegal?

### Technology capability notices (TCN)

15. It is proposed that OfCom will be given the power to issue a TCN, which forces a platform to facilitate an interception warrant and effectively acting as an injunction for end-to-end encryption, and they may issue them when faced with illegal or priority content. That duty raises the prospect of monitoring and breaking encryption on private messaging platforms.

16. On being issued with a Technology notice, the platform is obliged to co-operate with the regulator, and provide the information in way that can be comprehended rather than trying to be clever and provide a "binary blob." Trying to trick the regulator out of providing such information would be an offence.

### Judging opposition

17. Muffet opened his session noting that a look around global events sheds light on the challenges present when deciding what type of content is harmful or even terrorist content. A timely example is the protests occurring in Iran, which to the Iranian regime are deemed threatening, yet, to a reasonable outside observer, they are legitimate protests and the example could be applied to climate change activists or any other type of dissent. However, how you deduce what terrorism is depends on intentions, impact and consequences.

18. The Bill is proposing that to uncover potential terrorism and child abuse crimes, the government and society should be able to spy on what people are saying in messaging apps: WhatsApp, Signal, Facebook and so forth.

### End-to-end encryption

19. Encryption, in particular, end-to-end encryption, has elicited a less-than-cohesive view from among civil society, between being seen as a tool of the big platforms to enable surveillance capitalism and simply as a means to protect the privacy of two consenting adults wanting to send a message to each other.

20. Those in favour of the Bill have tended to focus on bad people doing bad things. It follows, through their argument, that breaking encryption [see Point 36 below] would allow victims to regain control but there are risks to ending encryption too.

21. Corrupting end-to-end encryption would also making messages and sometimes personal details vulnerable to hacking.

22. Having the ability to look in on messages imposes a risk that those messages may be taken out of context and erroneously deemed nefarious, such as when a father in New York sent a picture of his son's genitalia to a doctor to diagnose a condition. As a result, was flagged by Google's child abuse algorithms and arrested, resulting in a ban from Google.

### Privacy-protection solutions

23. Having an auto-check in place on all images just in case abuse is being perpetrated, could be likened to having CCTV fitted in our houses, just in case abuse or violence occurred or if Alexa or Echo had to keep microphones running all the time and recording everything, just in case anything happened. Is it to say that people should not be permitted to have secrets amongst themselves or privacy, just in case they get up to harm. Encryption is an enabler of privacy.

24. A popular question, particularly from groups advocating against violence against women and girls and for child protection, was how you get the right balance between the right to privacy and the right to protect and any particular models that have been seen to work well, or the minimum reporting requirements

25. There is parallel work being conducted among the participants around encryption, grappling with how tech companies will deal with the volume of fake (and disturbing) imagery, which looks real, and how they will manage to intervene. The question arose as to whether a pathway may exist in the monitoring and reporting of the content in public and private messaging forums, should there be no success in stopping the Bill.

26. The potential answer is to ask for some transparency. In the bill there is an obligation on service providers to explain what technology they will be using for monitoring. There is a lot of power on Ofcom to develop and make this transparency meaningful around the types of technology and explanation of what they are as well as the policy around the types of content they would take down.

27. One contribution from participants asserted that the rights community has been discussing hashing technology for some time and how to apply that to image-based abuse as another way of looking at tech-enabled harm.

28. It was also suggested that there is a need for a holistic response and resources needs to be put into prevention, social services and healing justice, for example.

29. There is a sense that there must be more than a response to the harms being committed, including a public health and criminal justice response to violence against women and girls as well as education to aid prevention.

### Cost-benefit analysis

30. A cost benefit analysis was put forward as a reasonable thing to push for in terms of public policy and requesting data to back up the choices. For example, earlier this year, the Home Office spent £534,000 on a campaign called No Place to Hide, which was deployed via several charitable organizations aimed at trying to dampen popular support for end-to-end encryption with large numbers touted in terms of child abuse images that could be lost – 15 million reports according to the campaign. However, on further scrutiny, those numbers are much smaller considering 40-50% of images are recirculated; a GCHQ report earlier this year placed the number of children who are put into safeguarding due to reports at 8,800. On the other side there is the risk to those whose privacy is being lost; the businesses that aren't being built because end-to-end-encryption doesn't exist; the fines that are being levied; the fallout where NHS doctors are sharing medical imagery on WhatsApp; and what would be the cost of a leak.

31. A response to the proposition of a cost benefit analysis was to also think about what giving more power to the State to dictate the terms of engagement of our public communications and dialogue would cost. In addition, the disproportionate power that social media companies have is also not a good situation. Between these two actors that hold significant power, it's important to go back to the individual and their rights and reduce the potential restrictions and limitations on speech that each side is is trying to impose. The question that leaves is how to build in safeguards and refocus the conversation on private messaging.

32. Within the women's rights space there is a concern about framing the cost benefit in terms of there only being a small number of victims and survivors at stake but rather reflecting on the impact on those people and where there is a disproportionate impact on Black and minoritised communities who are much more likely to be victims of online abuse. Who experiences these abuses and their experiences matter and aligns with the oppressive systems at play, yet in the same breath it's clearly a concern that privacy infringements could lend itself to the over-policing of Black and minoritised communities.

### Disproportionate impact

33. The implications once the norm changes from encryption to decryption is likely to fall on people that are made vulnerable and marginalised: people's undocumented status could be revealed; the crackdown on dissent that is creeping through could applied increasingly online.

34. In the women's rights space, the conversation has been about how to move things offline because, as an example, talking about abortion in certain States and trying to convince can cause a lot more harm and put you on the radar. Victims and survivors are impacted as are the people helping them and each group are citizens in their own rights and experiencing the impact to their whole life and not just an exacerbation of existing harm.

35. Further thought from civil society was that the Bill is a "paradoxical situation" with a government that purports to support freedom of speech, yet are overseeing a drastic expansion of surveillance technologies that disproportionately affect marginalized communities like the Prevent matrix and other kinds of targeted interventions that do already surveil Black and Muslim communities and other people of colour in severe ways. Eroding the digital rights to privacy and free expression will therefore, hit communities that are already over surveilled even more.

36. There are forms of speech that are legal that do create harm, but giving the state backing of social media companies to potentially delete it does nothing to approach the issue down the road and creates a perception from government that they've tackled anti-racism when in reality, they are one of the most hostile governments in the UK's recent history.

### Weakening encryption

37. Interest was shown in understanding the way encryption is weakened with monitoring. Muffet explained that the government is proposing two separate approaches. If you consider what wire tapping or breaking early encryption involved – tapping onto the connection of a phone exchange or breaking the "shell" found around a conversation – breaking what evolved into public key encryption became more

challenging because encryption became part of the conversation itself and not just an armour plate around it that could be broken. Spying can only take place now by installing software that copies everything to GCHQ, for example, if it matches certain characteristics, also known as client-side scanning. It can utilise hashing, software or databases to identify child abuse imagery, terroristic content or Edward Snowden's leaked PDF document and send to law enforcement to be investigated. In essence, everyone's phone would need to be connected to GCHQ or the local security services. The other method would be to have what's called a ghost, a third party, spliced into the communication to receive a copy of everything so that it's all secure, but so that the man in the middle is permanently present and "listening in" on everything that is said.

### Criminalising content

38. Another viewpoint from around the table made reference to the intersecting impact of the Bill alongside dangerous counter extremism policies, raising the concern that the Bill securitises the space online when there could be legitimate discourse.

39. Many grassroots groups are using the online space to tackle sensitive issues, including virginity tests, misogyny, racism, cultural taboos, LGBTQ+ and migrant issues. The question the social media teams are finding themselves asking repeatedly is 'will posting about certain issues lead to them being shut down.'

### Digital justice

40. Migrant rights groups have been sensitised to digital justice for migrants recently, which seems completely pervasive at the moment and there is a real concern for the over-policing of racialised communities and particularly migrant communities. There is therefore a keenness to learn more about how the Bill could impact migrant communities and people in solidarity with migrant communities being criminalised through the Policing Act.

41. Various groups in civil society are challenging invasive tech used by authorities, such as the scanning of biometrics and facial recognition, therefore, are keen to see the interconnectedness because impacts tend to be the same.

42. For groups working on race and migration, it can be seen that new rules could lead to dire consequences from detention to deportations to the impact on people's citizenship and it all comes in after a series of assaults by the government.

### Opposing rights paradigm

43. There was some consensus that the privacy versus protection paradigm and pitting rights against each other was unhelpful, where you are not recognising that privacy, has a protection element to it. That could be relevant for children from disadvantaged communities carrying out political activism in settings where that poses a risk, so protecting their privacy means also protecting them from violence.

44. The privacy versus protection paradigm also means positioning children versus adults with all children seen as victims of rights infringement rather than actual subjects of rights and people able to have agency and decide for themselves to what extent they want to be involved in private conversations.

### Individual versus collective rights

45. In concluding, the chair proposed that the Bill seems to be balancing everybody's privacy against generic harms, seen as a collective harm, but from a legal perspective the balance should address the individual. Therefore, should a government infringe upon individual rights and freedoms without any indication of criminal activity or harm? In an offline environment, a warrant would be sought if there has been evidence of wrongdoing.

46. Being able to balance individual human rights against certain interests of society also proves difficult from the lack of numbers available to demonstrate harm, which makes it difficult for platforms to judge

those interests and to build limitations into their algorithms and  human moderation.

**Links for further information:**

Open Rights Group Online Safety Bill Campaign Hub

Alec Muffet's Primer on End-to-End Encryption

Privacy International's Encryption Report

Liberty's Report on the Online Safety Bill

Child Rights International Report on Encryption

Open Rights Group Blog on Immigration Impact of Online Safety Bill

End Violence Against Women Coalition Code of Practice

Racial Justice Network Campaign – Stop the Scan

Racial Justice Network Report on Biodata Collection