ORG OPEN RIGHTS GROUP

POLICY BRIEF

WHO'S CHECKING ON YOUR CHATS IN PRIVATE ONLINE SPACES?

Published by Open Rights Group, 12 Dukes Road, London, WC1H 9AD.
A UK-based organisation that works to preserve digital rights and freedoms by campaigning on digital rights issues.
Open Rights is a non-profit company limited by Guarantee, registered in England and Wales no. 05581537.

Who's checking on your chats in private online spaces?

A policy brief on the inclusion of private communications in scope of the Online Safety Bill and the impact on end-to-end encryption.

Open Rights Group, February 2023
Author: Dr Monica Horten

This briefing outlines how private chat services have been surreptitiously included in the scope of the Online Safety Bill. It examines how compliance will result in the monitoring and interception of chat messages, with the consequence that the security provided by end-to-end encryption could be compromised. The outcome could entrench a form of mass chat surveillance involving checks on videos, images and text with ensuing risks to confidentiality of communications and a chilling effect on speech.

Parliament is being asked to legislate for these disproportionately intrusive measures, affecting our privacy and freedom of expression, without any specific information about the impact on either users or providers. People and companies have a right to know what the measures are and how they can take action to avoid penalty, before the Bill goes on the Statute.

Given the large numbers of users of chat services in the UK, these measures require proper scrutiny and due process. Our elected representatives should not be asked to vote on an invisible measure with such enormous consequences.

Executive Summary: What's the Issue?

At the heart of this issue are private, encrypted messaging services, chat platforms like WhatsApp and Telegram, and whether they should scan their users' messages to support government policy aims. What we are talking about is a form of chat surveillance that is being slipped in through a back door measure in the Online Safety Bill.

Chat services allow us to send short messages instantaneously, but they also let us post images and videos, and make voice and video calls to UK and international numbers. They are generally accessed via apps than run on users' smartphones. The end-to-end encryption of messages protects confidentiality of communication for the whole of society. It is a safety mechanism against hackers trying to steal personal data or intercept calls.

Millions of British people use these services every day for personal, business and social communications. Indeed, for some people they have become indispensable. For many, they have stepped into the role of the old fashioned wired telephone.

The policy question that the government is seeking to address is how to tackle child sexual abuse online. A proposed solution is that encrypted chat services should scan the messages of all users to identify any material shared by offenders. However, this would put at risk the end-to-end encryption that protects confidentiality and ensures that conversations remain secure. In other words, what appears to be a silver-bullet technical solution will create further harms to wider society.

This paper presents a two-fold argument, Firstly, that the Bill itself obfuscates whether or not chat platforms are meant to comply with its requirements, and what they could be asked to do. And secondly, that the government has overlooked the size of the chat services user base in the UK, representing 60 per cent of whole population, and larger than the base of old-fashioned telephone lines. Both are deeply problematic because of the potential for the measures to interfere with privacy and the likely chilling effect on free speech. Under those circumstances, is it right that Parliament should vote on the Bill without the necessary knowledge to assess its implications?

It is not obvious to anyone reading the Online Safety Bill that it addresses private chat services. It does not mention chat platforms specifically so one could be forgiven for thinking they are not in scope. However, the Impact Assessment confirms that they are, and this meets with the understanding among stakeholders and Ministers that they are[1].

Compliance with the Bill will have implications for the operation of their services. The concern is that the UK regulator, Ofcom, who will be overseeing the implementation, could force chat platforms to use government-accredited Technology. This is confirmed in Clause 187(11) as "an example of content moderation technology". The Bill is silent as to the precise implementation, but it is generally understood to mean a form of client-side scanning, where the software would reside on users'

1 Damian Hinds, giving evidence to the Draft Bill Committee, 4 November 2021 See Draft Online Safety Bill (Joint Committee) Oral Evidence transcript 4 November 2021 Q287 https://committees.parliament.uk/oralevidence/2949/pdf/

smartphones. This approach is similar to that of authoritarian regimes such as Russia and China, and sets a poor example for other countries to follow. It raises concerns in regard to the UK's international standing.

Chat platforms operate across borders, and they support billions of people around the world. The required measures would have a significant impact on security of the communications infrastructure, not only in Britain but worldwide. The UK has demonstrated with the recent funeral of the late Queen Elizabeth that it remains a leading player on the world stage. It must not do anything to jeopardise security of communications for people in the UK, nor set a precedent on the global stage for breaches of fundamental privacy rights. Ideally the government would provide an undertaking that surveillance measures will not be imposed on encrypted chat services in future.

Moreover, it will affect the user experience. The size of the user base and the popularity of chat services has significant implications for the implementation of this policy. More than 40 million people in the UK use chat services. They have largely replaced the old-fashioned telephone in many people's lives. The dominant provider is WhatsApp, with 40.23 million monthly active UK users, according to the statistics portal Statista[2]. It has more than 2 billion users worldwide. It's smaller rival Telegram has 550 million users worldwide, and is growing its base in the UK with 1 million downloads of its app in the first quarter of 2022. Another rival chat service Signal has 40 million users worldwide.

The government is targeting these services because of their end-to-end encryption. It is also gunning for Facebook Messenger which has 31.78 million active monthly users. In the UK[3], and a global active monthly user base of 988 million. It is not yet encrypted although there are plans to do so.

We are therefore looking at measures that will result in mass surveillance of communications services used by more than two-thirds of the UK population for private messaging, including video and voice calls. They will interfere with UK citizens' privacy and freedom of expression.

According to an expert legal Opinion [4], this Bill would create the power to mandate some of broadest surveillance powers in any Western democracy. It goes much further than the Investigatory Powers Act, without any safeguards or oversight to protect privacy rights of individuals.

As such, there is a deep flaw in the Bill. Parliament is being asked to legislate for disproportionately intrusive measures, affecting our privacy and freedom of expression, without any specific information about the impact on either users or providers. People and companies have a right to know what the measures are and how they can take action to avoid penalty, before the Bill goes on the Statute. They should not be smuggled in via this Bill, whose primary aim is to tackle content posted on public social media platforms. Our elected representatives should not be asked to vote on an invisible measure with such enormous consequences.

2 Statista, Mobile apps in the United Kingdom (UK) - Statistics & Facts https://www.statista.com/topics/7344/mobile-apps-in-the-uk/#dossierKeyfigures

3 Statista, Number of monthly active users (MAU) of leading smartphone and tablet apps for users in the United Kingdom (UK) in September 2021 – https://www.statista.com/statistics/1284692/top-apps-uk-mau/

4 Opinion by Matthew Ryder KC, and Aidan Wills of Matrix Chambers for Index on Censorship https://www.indexoncensorship.org/wp-content/uploads/2022/11/Surveilled-Exposed-Index-on-Censorship-report-Nov-2022.pdf

Chat Surveillance: An Introduction

Chat surveillance is about the scanning of private messages, either in transit or prior to transmission to look for prohibited content, which would then be removed without the user's knowledge and potentially referred to a law enforcement agency. In today's environment where chat platforms can process millions of messages a day, it would be carried out by large scale content moderation systems.

The scanning could be done on the platform server, or on the user's own phone. It works by intercepting chat messages to check out uploads of photos, videos, graphics and text. Images would be compared against databases looking for prohibited material. As the Bill currently stands, this will be child sexual abuse material, but it's worth noting that the technology is not limited to any particular type of image, if the law were to be amended in future.

Chat platforms will be doing this on behalf of the government and law enforcement agencies. If a match for an images is found in the database, it could be reported to the National Crime Agency. If chat services don't implement scanning voluntarily, they could be forced to do so by Ofcom. It would create a form of mass surveillance, operating 24/7, with obvious implications for individual privacy.

A key concern is about the way the Bill impacts on the underlying systems. Notably, compliance would compromise the end-to-end encryption that has been put in place to ensure the security messages for everyone who uses their services. This is because content that is end-to end encrypted cannot be read by anyone except the sender and recipient. That includes the chat platforms themselves and so the only way they can comply with the requirement in the Bill is to either break into the encryption (via a "back door") or work around it by scanning the content on upload – on the users' phone – before it is encrypted.

Back doors create security risks. They create vulnerabilities that can be exploited by malicious actors and hostile states to corrupt the system. This is not good for users. It breaks the security that they currently enjoy. It may not just affect users in the UK, but potentially also those outside the UK, as the chat systems are global and cross-border. Scanning on the phone increases the "attack surface" for bad actors to exploit. Hence, the measures in the Online Safety Bill risk weaken the security of chat services for all.

How Private Messaging is in the Bill

The strange thing is that there is no mention in the Bill of encryption or encrypted services, nor any clear definition that would signify private messaging platforms or chat services. The three categories of services outlined in the Bill don't include anything that suggests private chat services are in scope. The general assumption is that the Bill addresses social media platforms, which are open, public services, where content is generally accessible to other users on a platform.

So why then, can we make a claim that this Bill will have the effect of compromising private messaging services, and their underlying end-to-end encryption, and put at risk the security of UK and global users?

The Bill does it in a devious way. It defines a new type of online service that did not previously exist and does not exist as far as we know in other jurisdictions. This is "user-to-user services," which are defined in Clause 2 as a service where "content is uploaded, shared or generated on the service by a user and may be encountered by another user". It could be anything with a share button. The logic of the Bill is that messaging or chat services meet these criteria, as they typically do have share functions, and allow content to be shared both within the platform and outside it.

Private, encrypted chat is included only because the Bill mentions the word 'privately'. This is done by means of the definition of "content". The specific reference is any content communicated 'publicly or privately' in Clause 207, where content may consist of text, image, graphic, video, or a recording. This language appears to be sufficient to bring these private messaging platforms in scope. Of course, it still says nothing about whether they use encryption. Indeed, it never recognises that as an issue.

Technically chat services are quite different from a public social media platform. On social media platforms, users posts are available to all users, and they may engage with a wide range of users, whom they may or may not know personally. On a private chat service, users only communicate with those they choose to communicate with, whether it is a personal friend or a group with a collective interest. Their content is only communicated to those they choose, and is not visibly publicly. Hence, messaging or chat services are "private" even if the definition is stretched to a few hundred people in a group.

The position regarding voice and video calls over private chat services is not clear. One-to-one calls should be excluded, according to our reading of the Bill as it currently stands, which excludes one-to-one live aural communications, but the question is open around group calls.

The Bill makes one key differentiation between public and private content in the area of child sexual abuse material and terrorism content. Both of these are illegal content under the terms of the Bill. It is only concerned with public terrorism content, but with public and private child sexual abuse material. It gives Ofcom an enforcement power whereby it can require – under threat of large fines – platforms to seek out this specific content that is communicated privately, as well as publicly.

Hence we find that the scope of the Bill is automatically expanded from public social media services, to private messaging platforms, under the auspices of contrived definitions for services and content. It lumps together two very different types of communications service, whilst failing to recognise their technical differences and specificities.

Scope of duties

The scope of the duties imposed on chat platforms will depend on how they are categorised. This remains an open question and is another example of the loose drafting and failure to attend to detail. The Bill establishes three categories of Internet service which are merely labelled Category 1, Category 2a and Category 2a. The thresholds – in terms of user numbers or usage rates – have not been made public, nor have any other criteria. This information will only be known after the Bill is on the Statute.

It's widely understood that Category 1 services will include the very large online platforms. Category 2a will apply to search engines only. Category 2b will be all other user-to-user services. Messaging services therefore could either be Category 1 or Category 2b. The categorisation will have implications for the services and their users. The giant WhatsApp, with over 40 million users in the UK, could meet the threshold for Category 1. This could mean a raft of compliance requirements that are out of kilter with private chat. It would mean collecting vast pools of data that they do not currently collect, and seriously undermining the privacy of their users, who would effectively be put under constant surveillance.

As regards to the content they should monitor, it is likely to be limited to child sexual abuse material (as defined in Schedule 6). The Bill goes around in circles on this issue. However, our interpretation is that (under Clause 124) platforms cannot be required to use proactive technology (content moderation systems) on privately communicated content for the purposes of complying with the illegal content safety duties in Clause 9. This would preclude them being asked to monitor for any other type of content. This convoluted drafting ideally requires clarification.

How chat surveillance would be enforced

Given the general understanding is that chat platforms will be asked to monitor for child sexual abuse material only, the Bill incorporates a deeply cynical manoeuvre to force them to do this.

The regulator, Ofcom, may require private chat platforms to install government "accredited technology" that would identify and take down this material, or to "prevent users encountering" it. Chat platforms who don't want to use the government technology may

be required to use "best endeavours" to developed their own systems which must meet government standards[5].

It's the only specific requirement with regard to private communications. Even then, it is not obvious what the meaning is. The actual text states only that the provider could be required to "use accredited technology to identify CSEA content, whether communicated publicly or privately[…]". Alternatively, they must use "best endeavours to develop or source technology" that achieves this purpose and "meets the standards published by the Secretary of State".

The Bill confirms in Clause 202(11) that "accredited technology" refers to a content moderation system which are also classified for the purposed of this Bill as "proactive technology". However, it does not specify any further what these accredited systems would do, nor what the government's standards would be. It is not at all clear what the accreditation process would mean, nor what standards would be expected. It is also not clear what is meant by "prevent users encountering" the content in the context of chat platforms, other than sweeping it away and not allowing it to be transmitted.

There are large question marks around Ofcom's power to impose these measures. They are highly intrusive and are effectively a mandate for bulk surveillance of users' communications on behalf of the State. There are further question marks around the necessity of the policy aim, and whether it could be addressed by alternative, less intrusive measures.

The Bill does not provide for any independent oversight of Ofcom's exercise of these powers. In light of this, there would be seem to be serious flaw in the Bill that allows Ofcom to make a decision of this nature on its own, without any further scrutiny. According to the barrister Matthew Ryder KC, Ofcom would have more powers than GCHQ, whose powers for bulk surveillance measures are limited by the Investigatory Powers Act.[6]

It's worth recalling that the idea that Internet services should implement systems and software approved by the security services for the purpose of monitoring peoples private communications has strong roots in authoritarian regimes. For example, in 2014, Russia brought in a law requiring Internet platforms to link directly to servers run by their security services. How different is the UK government's thinking in the Online Safety Bill?

5 Online Safety Bill, Clause 110, government amendment of 25 November 2022

6 Opinion by Matthew Ryder KC, and Aidan Wills of Matrix Chambers for Index on Censorship https://www.indexoncensorship.org/wp-content/uploads/2022/11/Surveilled-Exposed-Index-on-Censorship-report-Nov-2022.pdf

What the Bill means for private messaging

Having established the private chat platforms would be in scope of the Bill, and potentially forced to comply by means of State-approved systems, there is a question about what it would mean for users. We assess this by examining the type of systems that would have to be implemented and the ensuing implications for encrypted messaging systems.

Scanning messages on chat services

The language in the Bill requires chat platforms to either identify and take down prohibited content, or to "prevent users encountering" it.[7] The Bill gives the platforms an enormous amount of leeway in making assessments of illegality. They are not required to have evidence – merely to "reasonably consider" it is illegal[8]. A government amendment, not yet passed a the time of writing, asks them to consider the "mens rea" - the mental elements in making their "judgements" (sic) of illegality[9].

Preventing users encountering content implies drastic measures. What's not specifically stated is "how" they should do this. They could use human or automated methods. Human moderation involves individuals looking at each piece of content that has been flagged as potentially illegal and taking a decision. This is impossible on the very large chat platforms that operate on a massive scale, with millions of pieces of content being uploaded every day.

The Bill confirms in Clause 187(11) that chat platforms would be required to use automated content moderation systems. They could either use the government's "accredited technology" or make "best endeavours" to meet the government standards. These content moderation systems would automatically seek out, detect and identify the prohibited content. The systems would then make an assessment as to the illegality, and determine an action The requirement to "prevent" users from encountering this content on a chat service would be handled by intercepting the content and sweeping it off the platform before it could be read by the intended recipient. It is a de facto general monitoring requirement. In the case of child sexual abuse material, there is an additional requirement in the Bill to report it to the National Crime Agency.

Content moderation systems operate in two different ways. When seeking known images, they look for matches of those images uploaded by users. The technique is known as perceptual hashing. Algorithms generate a digital fingerprint for the images known as the hash value – and this is checked against a database of hashes of known content that meets the prohibited criteria.[10] When they find a match, they take a decision to remove it.

7 Online Safety Bill, Clause 110
8 Online Safety Bill, Clause 9
9 Clause 170 from 12 July 2022
10 It's Not What It Looks Like: Manipulating Perceptual Hashing-based applications https://gangw.cs.illinois.edu/PHashing.pdf

An alternative method uses machine learning to look for unknown content. Neither of these methods is bullet-proof and both can generate false flags, as discussed below.

Even if messaging platforms "only" have to scan for a single type of content, they have to intercept and check every message being uploaded, resulting in a "de facto" general monitoring regime. This was illegal in the UK until recently. A general monitoring regime creates interference with freedom of expression, and the scale of it in the UK will effectively mean mass surveillance.

How the Bill undermines encryption

The privacy of users on private messaging services is protected by encryption – scrambling them so that they can only be read by the people at either end of a conversation and not be any third-party in transmission. Messages are guaranteed against eavesdropping or interference. This is known as "end-to-end encryption" also written as e2ee.

However, we have just said that chat services would have to intercept messages and scan the content in order to comply with the Bill. The effect would be to compromise encryption, and with it compromise the security and integrity of their systems. This has serious implications for users' privacy, and freedom of expression as was noted by the United Nations Special Rapporteurs in their report of May 2022[11].

End-to-end encrypted messages are scrambled in such a way that the content cannot be read by the platform[12]. In other words, the chat platform does not know the content of its users' posts. That also means the content cannot be accessed by content moderation systems unless something is done to peel away the encryption and reveal the clear text of the message.

The scanning can be carried out on the server but this requires the creation of a back-door into the encryption. Security experts are reluctant to do this because of the likelihood that the back-door becomes a hole that can be exploited by bad actors. Mandated back doors expose the system to unlawful interference by cyber-criminals and hackers, who may seek to access personal information like bank details or photos. The recent hacking of former Prime Minister Liz Truss' smartphone was a salient case in point about the risks of back-doors[13]. There are concerns among cyber-security experts that such back-doors, if insisted on by the UK, would be legitimised for use by in suppressing political dissent by authoritarian regimes in other parts of the world.

The alternative is to scan the content on the smartphone whilst the user is uploading it. The transmission is intercepted before the content is encrypted. If the prohibited information were found, it would be transmitted to law enforcement agencies, but otherwise, no information would be sent to third-parties. This is known as client-side scanning. Proponents of this method claim that it is a perfect solution because it enables checks to be made without compromising the encrypted transmission.

11 United Nations, The right to privacy in the digital age, A/HRC/51/17
12 For a full explanation of end-to-end encryption, see Alec Muffett's e2e Primer https://alecmuffett.com/alecm/e2e-primer/e2e-primer-web.html#surveillance-you-can-t-be-a-little-bit-pregnant
13 Liz Truss phone hack claim prompts calls for investigation https://www.bbc.co.uk/news/uk-politics-63442813
14 Abelson, Anderson, et al 2021 Bugs in our Pockets: The Risks of Client-Side Scanning https://arxiv.org/abs/2110.07450

However, it is a breach of confidentiality. Moreover, cyber-security experts highlight that it creates serious risks for individual security and privacy, and there is little proof of its usefulness for law enforcement.[14] Not least of these is that the software and database would be on millions of phones with varying levels of security protection. Or as the cyber-security professionals would say, it increases the "attack surface".

There is no technical reason that limits chat surveillance systems to a specific type of image. The database itself can be corrupted – where additional images are inserted by stealth – to look for documents and images other than what is officially required. In this way, the system has the potential to be corrupted for political purposes, for example.

False flags
The risk of wrongful incrimination

It would be wrong to consider content moderation systems a panacea to solve the problem of illegal content on chat platforms. Their robustness in identifying targeted content has been challenged by cybersecurity experts worldwide[15]. Data scientists have shown how image detection can be corrupted or manipulated. For example, it's possible for the same image to have two different hash values (digital fingerprints). It has been demonstrated how two images may look similar to the eye, but with the addition of "noise" they may generate a different hash, so their won't show up as a match when scanned. This highlights a way for illegal content to evade detection[16].

It has also been demonstrated how two quite different images, can have same hash value. To take a simplistic example, researchers have shown how an image of a dog and a cat can generate the same hash value if one image is corrupted with "noise". This technique could be used to disguise illegal content.[17]

All content detections systems are understood to result in false positives. The Swiss Federal Police have indicated that as much as 90 per cent of child sexual abuse material that is reported by automated systems is not illegal.[18]

Where illegal content is required to be reported to the National Crime Agency, there is real risk that people could be wrongfully incriminated. The possibility is illustrated by a recent case in the US where a father was flagged as a criminal over a photograph of his baby son's genitals that he forwarded to a doctor[19].
The dangers of false flags should be taken seriously by Parliament when it considers the application to private chat services.

15 Hal Abelson, Ross Anderson, et al. Bugs in our pockets: The risks of client-side scanning https://arxiv.org/abs/2110.07450

16 https://towardsdatascience.com/apples-neuralhash-how-it-works-and-ways-to-break-it-577d1edc9838 [see also https://gangw.cs.illinois.edu/PHashing.pdf] 577d1edc9838

17 https://towardsdatascience.com/apples-neuralhash-how-it-works-and-ways-to-break-it-577d1edc9838 [see also https://gangw.cs.illinois.edu/PHashing.pdf]

18 Breyer: Chat Control: the end of privacy of digital correspondence https://www.patrick-breyer.de/en/posts/messaging-and-chat-control/

19 New York Times, A Dad Took Photos of His Naked Toddler for the Doctor. Google Flagged Him as a Criminal. https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html

A paradigm shift in surveillance

Chat surveillance and client-side scanning represents a paradigm shift in online surveillance. To put it in context, it's worth taking a very brief look at the history of surveillance. Interception of communications by government and intelligence services is not new. Indeed, it's an age-old practice that goes back through the centuries when governments intercepted letters in the post. Interception of phone calls has been going on since the early days of the telephone in the first half of the last century. The first explicit interception law in the UK was Section 4 of the 1920 Official Secrets Act.[20] The interception was done with the co-operation of the phone company, then a State-owned monopoly. It was also known as wire-tapping, and it required a warrant.

When the GSM mobile phone system – that underpins the smartphone – was first introduced, it was criticised by law enforcement because it could not be intercepted. Changes were subsequently made in the technological design and interception of GSM phones was made possible.

In 2006, the intelligence services began lobbying in the UK and in the EU for access to communications traffic data. This took interception in a new direction. It wasn't about listening to the calls but about knowing the calling patterns of users, who they called and when, creating a permanent digital record for law enforcement to search. These methods escalated, as was revealed in 2013 by the US whistleblower Edward Snowden, who showed how intelligence services could bulk-search millions of messages, not only on phones, but also Internet content.

Powers exist to require communications service providers to assist the security services, via the Investigatory Powers Act 2016 (Section 253). These powers are however, protected by a double lock, including judicial review. The government has recently admitted in the case of Robinson v United Kingdom[21] that the bulk interception powers in Section 8(4) of the Investigatory Powers Act fall short of compliance with the European Convention on Human Rights Articles 8 (Privacy) and 10 (Freedom of Expression).

What we are seeing here in the Online Safety Bill is a covert piece of law to smuggle a new form of mass surveillance onto the statute without scrutiny. It relies on the bulk scanning of messages of over 40 million users of chat services in the UK, with serious implications for the encryption that protects chat messages and keeps them secure.

20 Scott, P. F. (2022) The first interception provision: Section 4 of the Official Secrets Act 1920. Journal of Legal History, (doi: 10.1080/01440365.2022.2140487

21 Robinson v UK Application no. 65487/16 European Court of Human Rights

Chat surveillance and individual rights

Systematic surveillance through bulk scanning without suspicion of any involvement in criminal activity, does constitute an interference with privacy and freedom of expression of chat platform users. It would be likely to have a chilling effect on freedom of expression.

The government is therefore under a legal obligation to assess the human rights impacts of these systems. This is a wider requirement than considering data protection, and, because the surveillance is being mandated in the context of a law enforcement requirement, the law should be very clearly defined, with strict limitations on, for example, access to government databases.

In this case, the government is legislating to regulate a communications system with over 40 million users. It has a duty to balance the rights of those users against the rights and interests of providers and other stakeholders.

Measures likely to interfere with privacy or freedom of expression should be prescribed by law, but it is hard to see how this Bill meet that test, when the type of service is not specified in the law, and there is no description of how the measures should operate.

Moreover, the bill does not enshrine the principle of freedom of expression and privacy as rights, as per the Human Rights Act and European Convention on Human Rights. Indeed, the text of the Bill has downgraded speech and privacy rights which are considered little more than a contractual matter[22].

The Memorandum on the European Convention on Human Rights that accompanies the Bill[23] does acknowledge that Article 10 and Article 8 rights are engaged and seeks to justify interference with them. It accepts that requiring the use of scanning systems on chat platforms may constitute a restriction on freedom of expression because "it will involve the removal of content". This interpretation is in line with human rights standards, since any form of removal or take down, blocking or filtering, is considered to be interference.

The Memorandum also acknowledges that scanning and analysing users' content is an interference with privacy rights, but swiftly dismisses it. However, it does not attempt a balancing act between the rights of the vast majority of users who are acting lawfully, and the interests of the State. Notably, it cites the Technology Notices in Clause 110, but fails to recognise the intrusiveness of the measures.

22 Online Safety Bill, Clause 18
23 Online Safety Bill ECHR Memorandum

Interference with freedom of expression

In the scenario described above, how would chat users know that their photos or message had been removed? The error rates and false flags of the scanners are a cause for concern. Usually, on a chat service, the sent message appears in the senders chat account, linked to the name of the sender. There is an indicator to say if it was delivered, and if it was read. So either the user would see no message and wonder what had happened, or they'd see the message but no read receipt.

Any removal of content would constitute an interference with freedom of expression under Article 10 of the European Convention on Human Rights, enshrined in UK law within the Human Rights Act. The Bill makes no provision for users to be told that their message has been restricted, or how they may appeal. There is a basic complaints procedure (Clause 19) that places a very low bar for the provider. It would be a relatively simple amendment to incorporate stronger procedural safeguards into the Bill. This would at least offer some possibility for both users and providers to be assured that the law had been applied correctly.

Violation of privacy

Children, as well as adults, need these safe spaces to talk, engage and interact with trusted friends, family and colleagues. It's not that people are doing anything wrong, they just don't feel comfortable when someone else is listening in. Especially if that person is the government or a private company acting on behalf of the government.

Moreover, people who rely on secure, confidential communications would be put at risk. They include journalists, victims of abuse and other crimes, and also children who need to be able to speak safely to others, without risk of their calls being hacked or messages compromised.

These measures in the Online Safety Bill represent an unprecedented attack on individual privacy. The sheer scale of them puts everyone under suspicion. Individual communications will be monitored without a warrant.

The Bill will ask chat platforms to monitor vast swathes of private communications which is not limited to child sexual abuse material. The potential for error is large.

Such surveillance of more than 40 million UK users would clearly represent a human rights violation under Article 8 of the European Convention[24]. Moreover, the template set by this Bill is extremely dangerous if it is transferred to other countries, with regimes that do not respect rule of law and human rights.

Necessary and proportionate measures

Tackling child sexual abuse is quite clearly a legitimate policy aim. However, the question we ask here, is whether these measures are necessary and proportionate to achieve that aim, and whether the necessity and proportionality has been addressed on the face of the Bill.

What's at stake is the desire to tackle the a very serious criminal offence on the one

24 See Ot van Daalen, Does monitoring your phone affect the essence of privacy? https://europeanlawblog.eu/2022/06/07/does-monitoring-your-phone-affect-the-essence-of-privacy/

hand, balanced against an uncertain technical solution that may deliver flawed outcomes, on the other. It is seeking a technical silver bullet to solve a problem that has its roots in societal issues. The legislation calls for the removal of images, based on hashes in a database, but does not address the societal causes of these crimes, nor does it address the need for law enforcement resources to tackle the perpetrators. Ross Anderson, Cambridge Professor of Security Engineering, argues that the child safety debate should be addressed "from the perspective of children at risk of violence, rather than from that of the security and intelligence agencies and the firms that sell surveillance software."[25]

An assessment of necessity and proportionality would have to therefore take account of these factors. This should be balanced by an assessment of the potential systemic risk that this proposal undermines the security of everyone on the system and will create new harms affecting all sectors of society. The policy aims should be spelled out clearly in the Bill by the government, as well as the actions that online chat providers are expected to take.

The Online Safety Bill states over and again that the measures should be proportionate, but it leaves it to the service providers to determine proportionality. As the measures are never clearly stated in the Online Safety Bill, the objective is also never revealed. It is therefore difficult for providers to define how the measures could be proportionate to address an unspecified objective. This itself indicates a lack of clarity in the thinking that went into the drafting of the Bill.

The law that governs freedom of expression and privacy insists that the least intrusive method should be used. That has been interpreted by courts to mean specifying the content to be restricted, such as providing a URL, and obtaining a warrant for intrusion into private communications. These measures cannot meet the test of being the least intrusive[26]. As blanket measures that seek to intercept and check the chats of over 40 million people, they are far and away the most intrusive. Moreover, there is no accountability on the part of the provider.

What is very curious is how the ECHR Memorandum attempts to relegate the necessity and proportionality test to Ofcom and seems to rely on the accuracy of the accredited technology[27]. This does not seem appropriate and indeed seems to be shirking the State's responsibilities. The tools are not yet developed and their accuracy has not been assessed, and given the level of intrusiveness, it seems inappropriate for Ofcom to exercise this power without any oversight. It should be clear on the face of the Bill how the measures would be necessary and proportionate, and state precisely the circumstances in which they could be authorised. It would also specify what the measures would be. Delegating it to a regulator, or a private actor, would seem to be a derogation of duty.

Better still, private messaging services that are end-to-end encrypted should be out of scope of the Bill entirely.

25 See: Ross Anderson, Chat control or child protection? https://arxiv.org/abs/2210.08958

26 Opinion by Matthew Ryder KC, and Aidan Wills of Matrix Chambers for Index on Censorship https://www.indexoncensorship.org/wp-content/uploads/2022/11/Surveilled-Exposed-Index-on-Censorship-report-Nov-2022.pdf

27 Online Safety Bill, ECHR Memorandum, paragraphs 61-64.

Conclusions and Recommendations

- **Parliament should not be asked to vote on an invisible measure.** The scale of the user base that will be affected, combined with the likely interference with privacy and freedom of expression, means that the government and Parliament should conduct due diligence before passing these provisions into law.

- **The Bill should incorporate the right to privacy and freedom of expression as an over-arching principle.**

- **The Bill should mean what it says and say what it means.** It should be spelled out precisely on the face of the Bill what providers are to do and what users can expect.

- **There should be a dedicated oversight mechanism for any powers granted to Ofcom to impose mass surveillance measures.**

- **Ideally, end-to-end encrypted services should be outside the scope of this Bill.**

openrightsgroup.org
info@openrightsgroup.org
+44 (0)207 096 1079

ORG OPEN RIGHTS GROUP