Ofcom Illegal Harms Consultation

Online Safety Act

Internet Society Response to Ofcom Consultation: "Protecting people from illegal harms online"
February 2024

Summary

The Internet Society is a global charitable organization that advocates for an open, globally connected, secure, and trustworthy Internet—an Internet that works for everyone. In this submission we share our concerns over the impact of Ofcom's proposed measures on end-to-end encrypted (E2EE) services.

This Consultation does not address the process for issuing Technology Notices in S.121 of the Online Safety Act, the measure that would most directly impact E2EE. The Consultation likewise does not address client-side scanning, widely discussed by policymakers as a possible 'accredited technology' for Child Sexual Exploitation and Abuse (CSEA) detection in E2EE environments. The Consultation does, however, state that proactive measures using automated content moderation technologies[1] will not apply to encrypted services. This is a positive and welcome message.

On the other hand, this Consultation and associated Guidance does raise new issues that are of concern to encrypted services. Namely, it expresses the viewpoint that E2EE is a risk factor for numerous illegal activities, implying a possible expansion of the scope of content that encrypted services would have to tackle beyond CSEA. The framing of E2EE as a risk factor discounts the many ways that strong encryption reduces risk for users, places indirect pressure on providers not to roll-out encryption on their services and goes beyond what is required in the Online Safety Act with regards to the type of content that can be moderated in encrypted environments.

[1] A.9.10; "Proactive technology is defined in section 231 of the Act. Broadly speaking, this refers to: (i) 'content identification technology', except where this is used in response to a report from a user or other person about particular content; (ii) 'user profiling technology' (which excludes technology deployed in the circumstances referred to in section 231(5) of the Act); and (iii) 'behaviour identification technology', except where this is used in response to concerns identified by another person or an automated tool about a particular user."

Recommendations

The Internet Society recommends that Ofcom redraft the Guidance in the following ways:

1.1	to resolve the conflict between the positioning of E2EE as a risk factor for the 12 categories of harmful content (Volume 2) and Ofcom's legal limitations for mitigating this 'risk' in encrypted environments (Volume 4).
1.2	to remove E2EE as a risk factor in risk assessments, recognizing that providers use strong encryption to mitigate the risk of various harms that would not be properly captured in assessments. Ofcom should unambiguously maintain its position that encrypted services cannot be subjected to proactive content moderation measures.
2	to provide a clear definition of "content communicated privately", where the method of communication (for example the use of E2EE) is evidence that the communication was private.
3	to clearly state Ofcom's opposition to proactive content moderation measures for the 12 categories of harmful content (Volume 2) in encrypted environments. Given the intrusive nature of proactive content moderation, the expansion of these powers would require new legislation, public consultation, and impact assessments, including an Internet impact assessment.
4	to clarify whether Ofcom's intention is that providers should monitor for criminal actions or behaviour, in addition to content.
5	to address the balance of rights, and the necessity and proportionality of the measures, in line with human rights standards.

Volume 2, Question 1 - Ofcom's Register of Risks

"Do you think we have missed anything important in our analysis? Please provide evidence to support your answer."

We are concerned that Ofcom has missed the link between the risks of online harms (Volume 2) and the technical reality for mitigating those risks (Volume 4).

Q1.1 The Guidance approach

There is a contradiction in Ofcom's approach, presenting two diverging positions on encrypted services.

The Guidance calls for a risk assessment to be carried out for each of the 12 categories of harmful content, with the implication that measures should be taken to address these risks. The Online Safety Act, however, explicitly states that S.10 measures do not apply for encrypted services.[2] The technologies to address these risks in encrypted environments are not feasible and therefore cannot be recommended in the Guidance. Ofcom acknowledges this in Volume 4[3] which states that proactive measures will only be recommended for publicly communicated CSEA material.

Q1.2 The Online Safety Act requirement

In the Online Safety Act, private encrypted platforms are exempted from scanning for the 12 listed categories of illegal content (S.10). The Act also tells Ofcom that it may not "recommend the use of the technology to analyse user-generated content communicated privately."[4] On this basis, Ofcom cannot require E2EE platforms to moderate S.10 illegal content.

The requirement for encrypted platforms to scan for CSEA can only be enforced by means of a special legal mechanism—the Technology Notice (S.121 of the Act). The Technology Notice is specifically designed to require services to scan for CSEA content, including content "communicated privately".[5]

In short, under the Online Safety Act, Ofcom can require an E2EE encrypted service to detect CSEA content through a Technology Notice, but not other illegal content (S.10 measures).

---

[2] On one hand, Volume 2 states that encrypted services pose a risk to a dozen categories of harmful content (S.10 measures in the Online Safety Act). On the other hand, the Guidance ignores language in the Act that limits Ofcom's ability to require that service providers address such risks.

[3] for example, in Volume 4, 14. 14-16.

[4] Schedule (4) 13(4): *"A proactive technology measure may relate to the use of a kind of technology on or in relation to any Part 3 service or any part of such a service, but if the technology operates (or may operate) by analysing user generated content or metadata relating to such content, the measure may not recommend the use of the technology to analyse user-generated content communicated privately, or metadata relating to user-generated content communicated privately."*

[5] See our response to Question 21.

Q1.3 Our response

From our reading of the Consultation documents, we identify two conflicting Ofcom positions on E2EE:

Position 1: Ofcom cannot require encrypted services to implement measures:

Ofcom acknowledges that proactive measures under S.10 do not apply to E2EE services. Ofcom also acknowledges that proactive measures would not apply where they are not technically feasible without compromising the security of a service, and states that this means E2EE services.[6]

We interpret this as meaning that Ofcom will not recommend accredited technology for an E2EE service under the S.121 Technology Notices because there is no technology available that meets the requirement. As such, it confirms that Ofcom would not use the Technology Notice mechanism to enforce a requirement for CSEA scanning on an encrypted service. This is in line with Lord Parkinson's statement in Parliament in Lords Report stage.[7] Without the existence of a proactive measure that is technically feasible, Ofcom could not use the powers outlined in the Online Safety Act.

Position 2: Encrypted services must assess multiple risks of illegal content:

The Guidance establishes E2EE as a "risk factor" for a dozen types of illegal content (S.10 measures).[8] At the same time, Ofcom confirms that a Technology Notice (S.121) cannot be issued for the purpose of scanning any of the illegal content included in this list, besides CSEA. This means that the law would need amendment before Ofcom could require the scanning of encrypted services for any of the other 11 types of content mentioned in Volume 2.

This provides services with a mixed message as they are asked to consider how encryption on their platform is a risk factor, while at the same are told that they will not be required to moderate this content if they are using encryption. This begs the question of how they are expected to mitigate the risk identified. It is important that Ofcom address the contradiction between these two positions. Otherwise, indirect pressure on providers would effectively circumvent exceptions for E2EE laid out in the Online Safety Act, implicitly pushing service providers not to roll-out encryption on their services.

If, on the other hand, Ofcom's position is that services should be proactively detecting illegal content beyond CSEA in encrypted environments, it is our view that the Secretary of State would need to place new legislation before Parliament. New legislation should include a public consultation and impact assessments, including an Internet impact assessment.[9] The Internet Society would strongly oppose such an expansion of powers.

[6] Volume 4, Section 14, notably in Section 14.16

[7] Lord Parkinson, House of Lords Hansard, Column 2363, 19 July 2023.

[8] Volume 2 Section 6.

[9] "Internet Impact Assessment Toolkit." Internet Society, 15 Dec. 2023, www.internetsociety.org/issues/internet-way-of-networking/internet impact-assessment-toolkit/.

Volume 2, Question 2 - Risk factors

"Do you have any views about our interpretation of the links between risk factors and different kinds of illegal harm? Please provide evidence to support your answer."

The Guidance identifies E2EE as a risk factor for multiple offences listed in the Online Safety Act. We are concerned that this framing of the risk assessment results in a broadening of the requirements imposed on encrypted services in contradiction to the limits found in the Act. Such a regime would effectively introduce new systemic weaknesses and vulnerabilities, putting users at risk and causing economic harm.

Q2.1 Our understanding of the Consultation guidance

Ofcom states that end-to-end encryption is a functionality that "stands out as posing a particular risk" to 12 different types of illegal content (Volume 2). In addition to child sexual abuse material, these offences are hate crime, terrorism, drugs, immigration, sexual offences, extreme pornography, intimate image abuse, proceeds of crime, fraud, foreign interference, and false communications.[10]

Ofcom's Guidance, however, goes further than what the Online Safety Act requires, suggesting an expansion of the scope. The Guidance requires providers to assess risk first by identifying the type of harm (such as image-based CSEA) and then by assessing possible risk factors including the presence of child users; social media services; private messaging services; discussion forums and chat rooms; user groups or group messaging; livestreaming; direct messaging; and encrypted messaging.[11] Providers must consider not only the presence of illegal content but also the possibility of an offence being committed or facilitated. Providers are asked to assess the likelihood of these harms taking place, and the potential severity of the harm to users.[12] This means that for each category of harmful content, they must outline what services they have which could be facilitating it.

This point is of concern to us because it suggests a broadening of the requirements for encrypted services. We understand that content moderation in encrypted environment would be limited to CSEA.[13] There is nothing in the Online Safety Act to say that the type of content could be expanded for encrypted platforms, indeed it says the opposite. The framing of encryption within the risk assessments suggests that Ofcom is considering expanding the scope of content requiring moderation beyond CSEA.

Our second concern is that language in Volume 2 suggests that the risk is not about the content itself but about the actions of criminals and perpetrators of these offences. For example, Ofcom states:

[10] The targeted content is defined as criminal offences under English and Scottish law. The full list can be found in Volume 2 of the Consultation documents, Sections 6: A to S. This goes further than what the Online Safety Act could requires, which would be limited to child sexual abuse material, (see below) and the inference is that the scope could be expanded.

[11] Vol 4, p111 footnote 201.

[12] Annex 5 A5.23 – A5.24.

[13] As per S.121 working together with Schedule 4.

> "Encryption and ephemerality make messaging particularly attractive to terrorist actors as they can reduce the chance of detection [...] end-to-end encryption can enable perpetrators to circulate CSEA, engage in fraud, and spread terrorist content with a reduced risk of detection [...] Messaging services are commonly used to offer and facilitate the supply of drugs, as they offer a closed channel of communication that reduces the risk of detection [...] suppliers often redirect prospective buyers to private messaging services, particularly those with encryption."[14]

The common theme is that encryption helps criminals avoid detection and that E2EE is therefore an enabler or facilitator of these crimes. If implemented, this approach would vastly expand the scope of the Online Safety Act. Providers that offer encryption would not only face requirements for the detection of CSEA images,[15] but also a dozen additional categories of content to be analysed for images and criminal behaviour. Providers would have to proactively seek out and report criminal activity that is both subjective and difficult to identify. The issue of private sector providers assuming an adjudicatory role was discussed widely during the Online Safety Bill's Parliamentary debates and was one factor that limited the scope of content moderation in encrypted environments to CSEA.

Although the Online Safety Act does discuss risk assessments for 'priority offences',[16] we believe that Ofcom's Guidance represents a significant shift in the policy aim of this law, that was not fully reflected during the legislative passage through Parliament. The implication is that providers should monitor the activities of criminals, rather than—as per the original aim of the law—to remove harmful content from their platforms. It would turn providers into enforcement officers for criminal activity, rather than moderators of content. We would appreciate clarification as to whether that is the intention.

Q2.2 Our response

Ofcom's conflicted position suggests that proactive measures[17] could be on the table. This scenario raises serious issues about systemic risks to the global Internet ecosystem. These are risks that have been called out by many cybersecurity experts, as well as former members of the UK intelligence community.[18]

Ofcom's approach undervalues the fundamental premise of E2EE in providing safety and security for all users on the platform. That is over 40 million users in the UK and some two billion users globally. E2EE messages can only be read by the sender and recipient. They cannot even be read by the service provider. This ensures a guarantee of privacy, security, confidentiality, as well as authenticity that it has come from the sender who it says it has come from (and not a spoof) and that the message cannot have been changed or altered by anyone. With billions of people reliant on digital communications to

[14] 6B: 46

[15] As per S.121 of the Online Safety Act.

[16] S.9 of the Act calls for providers to assess "*the level of risk of the service being used for the commission or facilitation of a priority offence*" and the "*risk of functionalities of the service facilitating the presence or dissemination of illegal content*".

[17] A.9.10; "Proactive technology is defined in section 231 of the Act. Broadly speaking, this refers to: (i) 'content identification technology', except where this is used in response to a report from a user or other person about particular content; (ii) 'user profiling technology' (which excludes technology deployed in the circumstances referred to in section 231(5) of the Act); and (iii) 'behaviour identification technology', except where this is used in response to concerns identified by another person or an automated tool about a particular user."

[18] Anderson, Levy and Robinson, Ciaran Martin.

speak not only to friends but also to government bodies, their health provider, and their bank, this level of security is important.

Q2.3 Introducing systemic risk

Any measure to screen the content of messages on an E2EE platform would introduce systemic risk, compromising devices and systems and leading to unauthorised access to data. It would increase risk for service providers and users. The outcome would be an unsafe and untrustworthy online environment and a new canvas for criminals to exploit. We have identified both technological and economic risks that arise.

A systemic weakness or vulnerability is one that extends beyond the targeted device or service that an individual user is using and is implemented such that any other user could be affected.[19] How might this happen?

Ofcom would require providers to use proactive measures such as perceptual hashing techniques.[20] Perceptual hashing creates a digital fingerprint of images uploaded and checks them against a database of images classified as illegal. There are workarounds, and the risk is that perpetrators of CSEA and other illegal images will seek to evade the scanning software by flooding the system with false positives, hence overloading law enforcement authorities with material that is not illegal. Alternatively, they would generate false negatives—images that "match" as CSEA material that is not—that can slip through the scanner. We would urge Ofcom to take full account of these risks in its Guidance in Volume 4.

Perceptual hashing could be implemented on the provider's server by creating a backdoor into the system to decrypt the message for scanning.[21] A backdoor is a form of exceptional access to allow for interception of messages.[22] Importantly, a backdoor represents a vulnerability point that not only would be used by law enforcement but also criminals and hostile foreign governments that seek unauthorised access.

The alternative proposal is to implement the perceptual hashing system on the users' devices. The sales pitch is that it does not require a backdoor to decrypt the messages because it will intercept the user's communications as they are being uploaded, before the encryption process begins. This is commonly known as "client-side scanning". We believe this is the technology that Ofcom acknowledges does not exist.

[19] Parliament of the Commonwealth of Australia, Review of the amendments made by the Telecommunications and Other Legislation Amendment (Assistance and Access) Act 2018, 4.56

[20] As outlined by Ofcom in Volume 4.

[21] For example, encrypting each message with a key known to the provider, rather than the keys on the devices of the communicating parties

[22] Jeff Wilbur, Ryan Polk. "A Backdoor Is a Backdoor Is a Backdoor." Internet Society, 24 Mar. 2020, www.internetsociety.org/blog/2020/03/a-backdoor-is-a-backdoor-is-a-backdoor/.

Q2.4 Client-side scanning

Client-side scanning is not a viable solution for content moderation in encrypted environments due to issues of inherent systemic risk and the violation of user trust. For example, putting the hash algorithm onto the client device would open it up to reverse engineering. The average user's expectations of privacy would be violated while criminals and hostile state actors would encounter little more than a speed bump that they would quickly develop techniques to circumvent.

Client-side scanning risks other unintended consequences. It creates new opportunities for attackers to target the database, for example, by inserting unauthorised material for scanning. Academic researchers in the UK suggest that facial recognition could be surreptitiously inserted.[23] The update mechanism could be subverted to install malicious software like what happened in the Solar Winds cyber-attack. Putting the database on the device increases the "attack surface" that bad actors can exploit,[24] and exposes millions of people's phones to bugging by unauthorised entities. These could include foreign states. Interference with privacy would be collateral damage.

Multiple databases to address different content categories would increase the complexity of enforcement, generate extra network traffic, and require extra processing on the device, leading to complex issues around scalability, testing, consistency of data, and governance.

Q2.5 Economic harm

Compliance with any measures to screen encrypted content would therefore introduce systemic weaknesses that would endanger users' security. Providers could be compelled to introduce systemic vulnerabilities or weaknesses that would extend beyond any individual targeted user or device. These vulnerabilities and weaknesses would create opportunities for other malicious actors that could in turn impact the privacy of others on the platform.

The effect would be reduced trust in digital systems which in turn could fuel economic loss. When trust is lost in an Internet service this depresses demand for the service and imposes higher costs for the business as it seeks to offset the harm caused—for example, by moving data centres abroad.[25] This effect was identified in Australia when the government brought in similar measures. Australian broadband providers said the adverse impact of the legislation on the trust in their brand meant they abandoned expansion into international markets and lost existing customers.[26] Another Australian

[23] Jain, S., Cretu, A., Cully, A., and de Montjoye, Y., 2023. Deep perceptual hashing algorithms with hidden dual purpose: when client-side scanning does facial recognition.

[24] Joint Statement by Europol and the European Union Agency for Cybersecurity (ENISA) of 20 May 2016 on lawful criminal investigation that respects 21st Century data protection in the case of Podchasov v Russia in the European Court of Human Rights, (Application no. 33696/19) Judgment 13 February 2024.

[25] "The Economic Impact of Laws That Weaken Encryption." Internet Society, 17 Mar. 2023, www.internetsociety.org/resources/doc/2021/the-economic-impact-of-laws-that-weaken-encryption/.

[26] P61 / P3 Exec Summary. "The Economic Impact of Laws That Weaken Encryption." Internet Society, 17 Mar. 2023, www.internetsociety.org/resources/doc/2021/the-economic-impact-of-laws-that-weaken-encryption/.

report suggested that strong cybersecurity policies are economically beneficial.[27] Trusted services which provide strong security are a basis for growth.

Another illustration of this point is how the Schrems II case in European courts raised concerns in Australia upon introduction of the new law.[28] Schrems II was the 2020 decision to strike down the safe harbour provision for international data transfers. Australian Internet businesses were concerned that the new law, known as TOLA, would put at risk international data flows to and from Australia. The UK, now outside the EU, would be in a similar position.[29]

The risk with encrypted systems, is not all one-sided. It is the role of Parliament rather than Ofcom (the regulator) to weigh the balance between economic harms and the way that the State protects the security of its citizens. However, Parliament has left it to Ofcom. We urge Ofcom to maintain its position that encrypted services cannot be subjected to proactive measures.

Volume 6, Question 53 - Enforcement

"Do you have any comments on our draft Online Safety Enforcement Guidance? [...] Please provide the underlying arguments and evidence that support your views."

The Online Safety Act provides guidance on how a Technology Notice could be enforced against non-compliant providers.[30] This is about penalties for failure to comply with a notice which would apply to end-to-end encrypted service providers. [31] They could be fined or blocked for failure to comply with Technology Notice[32] but as things stand, they would not know what they have to do in order to comply.

Yet the Online Safety Act provides no guidance on processes and procedures for issuing a Technology Notice to a provider of an encrypted service.[33] Encrypted service providers would not be able to foresee what they should be complying with. Therefore, they would not be able to take action to avoid the sanction.

This raises a pressing question about how a Technology Notice would be issued. We recommend that Ofcom kindly provide clarification about the process and procedures for Technology Notices. This should include clarification as to whether, under a Technology Notice, Ofcom would enforce a requirement to scan for criminal activity being facilitated by the encrypted service.

---

[27] "Mitigating Cyber Risk Could Make a Difference of USD 120 Trillion to Global Economy by 2030." Zurich, Atlantic Council, 10 Sept. 2015, www.zurich.com/en/media/news-releases/2015/2015-0910-01.

[28] Telecommunications and Other Legislation Amendment (Assistance and Access) Act 2018 (TOLA).

[29] P62.

[30] S.140 of the Online Safety Act.

[31] Section 121(1).

[32] [Annex 11A6.57]. [Annex 11, A 2.8 (d) and Footnote 10 (a)].

[33] S.121. Confirmed in Annex 11, p5, Footnote 10].

Volume 4 Question 21 - Content communicated 'publicly' or 'privately'

"Do you have any comments on the draft guidance set out in Annex 9 regarding whether content is communicated 'publicly' or 'privately'?"

We are concerned by a lack of a definition for content communicated privately and we disagree that the communication of this content would not engage Article 8 ECHR.

Q21.1 Interpreting content communicated privately

E2EE is not specifically mentioned in the Online Safety Act. It is understood to be an intended interpretation of the phrase "content communicated privately,"[34] and is implied in Ofcom's statements that proactive measures in S.10 cannot be applied to E2EE services.[35]

The Guidance on this point lacks clarity.[36] Ofcom has sought to define content communicated "publicly" but has not defined content communicated "privately." It is unclear as to whether Ofcom is looking at the technical delivery of the service or the path of an individual piece of content. The Consultation states:

> *"[…] the factors to be considered are how many people can access it, whether it is subject to access restrictions, and how easily it can be shared."*[37]

Q21.2 ECHR Alignment

We note with interest that Ofcom believes its interpretation of private communication may not align with the European Convention on Human Rights. The Guidance states:

> *"The question is whether the communication of the content is public or private, rather than whether the content itself is of a 'private' nature. As a result, whether content is communicated 'publicly' or 'privately' for the purposes of the Act will not necessarily align with whether that content engages users' (or other individuals') rights to privacy under Article 8 of the European Convention on Human Rights."*[38]

Q21.3 Our response

We disagree with this interpretation. The question of whether the *communication* of the content is private, does indeed engage Article 8 ECHR. It applies not only to the individual pieces of content, but also to the service used for communication.

There is case law to this effect. In *Liberty and others v United Kingdom*, the ECtHR established that interference with the means of communication engages Article 8. In other cases, the ECtHR has held that interference with the means of communication is interference with Article 8. In *Big Brother Watch and others v United Kingdom*, the ECtHR ruled that bulk interception and processing of

[34] Which is essential to the meaning of S.121 of the Online Safety Act.

[35] Volume 4, 4.14-4.16.

[36] Annex 9.

[37] Annex 9, 9.12.

[38] Annex 9, 9.15.

communications data by the State interferes with Article 8. A recent judgement, *Podchasov v. Russia* on 13 February 2024, provides an important guarantee under Article 8 for the privacy of users of telecommunications and Internet services:

> *"...confidentiality of communications is an essential element of the right to respect for private life and correspondence, as enshrined in Article 8. Users of telecommunications and Internet services must have a guarantee that their own privacy and freedom of expression will be respected."* [39]

E2EE is one of the most privacy preserving designs for encrypted services, used billions of times daily to protect information flows online (most webpages are encrypted end-to-end from the server to the browser).[40] Service providers of E2EE services cannot read messages shared on their platform—they merely transmit them. Only the sender and recipient can read them. The encryption keys are used to encode the message into a random series of numbers and letters at the "endpoint" (a device or piece of software) of the sender and then, on the other side, they are used to decode the message at the endpoint of the receiver. Neither the private key nor the original message is available to the operator. In this way, end-to-end encryption preserves and protects the integrity, authenticity, and confidentiality of people's messages.

If one were looking to describe a service that enables private communication, one would probably describe something very like an encrypted service, such as that it could only be read by the sender and recipient. We assume that *"content communicated privately"* means encrypted services, as has generally been assumed.

We recommend that Ofcom provide a positive confirmation of the meaning the 'privately' in the context of the Online Safety Act, where the method of communication (for example the use of E2EE) is evidence that the communication was private.

Volume 4, Question 48 - Statutory tests

"Do you agree that Ofcom's proposed recommendations for the Codes are appropriate in the light of the matters to which Ofcom must have regard? If not, why not?"

Our response to this question addresses the issues of technical feasibility and the proportionality test for end-to-end encrypted services. A requirement for E2EE services to scan content would be likely to fail a proportionality test, in light the judgement of 13 February in the European Court of Human Rights, *Podchasov v. Russia*.

[39] European Court of Human Rights, Podchasov v. Russia. Judgement 13 February 2024 https://hudoc.echr.coe.int/eng/#{%22itemid%22:[%22001-230854%22]}

[40] Husovec, Martin. "Podchasov v. Russia App. No. 33696/19 - Martin Husovec." European Information Society Institute, 28 Sept. 2021, husovec.eu/wp-content/uploads/2021/10/Podchasov-v-Russia-Brief.pdf.

Q48.1 The statutory tests

The statutory tests applicable to Ofcom's duties under the Online Safety Act are the test of necessity and proportionality under human rights law. The Online Safety Act states that measures implementing the duty in S.10 of the Act (Illegal content safety duties) must be designed in light of the principle of freedom of expression and the importance of protecting the privacy of users.[41] Freedom of expression is defined in the Act as having the meaning intended by Article 10 of the European Convention on Human Rights),[42] which is enshrined in the Human Rights Act 1998.

Other principles for which Ofcom must have regard include those set out in the Online Safety Act, notably that measures imposed by Ofcom on providers must be both proportionate and technically feasible.[43]

Q48.2 Technical feasibility

With regard to technical feasibility, Ofcom has stated in the Consultation that it cannot impose measures on encrypted services that are not technically feasible.[44] We are grateful to Ofcom for its acknowledgement of this position. However, as we've said in our response to Question 1, there is a disparity in Ofcom's position. In Volume 2, Ofcom asks for an extensive risk assessment of harms on encrypted services. The assessment concerns around a dozen harms, detected using a variety of content moderation technologies. In the Introduction to the whole Consultation,[45] it says that encrypted services still have to mitigate risks of CSEA on their services (without saying what that means) and are subject to all the safety duties in the Online Safety Act (which is contradictory to the Act):

> *"Using a technology called 'hash matching' to detect and remove known CSEA [...] this proposal does not apply to private communications or E2EE communications. We are not making any proposals that would involve breaking encryption. However, E2EE services are still subject to all the safety duties set out in the Act and will still need to take steps to mitigate risks of CSEA on their services."*[46]

We recommend that Ofcom clarify its position on risk mitigation on encrypted services.

Q48.3 The Principle of proportionality

The rights at stake are the right to freedom of expression and privacy, which the State should balance against national security and protection of the rights of others. When considering the proportionality of measures, the precise objective should be clear. Any restriction must be lawful, which generally means it must be clearly described in legislation. The lawfulness of the interference must be balanced against the rights of other users whose rights may be arbitrarily interfered with. The measures must be

[41] Schedule 4 (10) (1).

[42] S.236.

[43] Schedule 4 (10) (2).

[44] Volume 4, Section 14.14-14.16; see also p90.

[45] P2.

[46] Introduction, P2.

necessary and specific to achieve a legitimate purpose, ideally with a fact-based assessment of their effectiveness.[47] The least intrusive restriction should be used to meet the policy aim and it should be possible to show that other less intrusive measures have been evaluated.

The quality of the law and existence of adequate safeguards will be factors. The law must be sufficiently clear so that users will be able to foresee when their communications could be interfered with. The scope of discretion for private actors implementing the measures must also be clear. The legal framework must include adequate safeguards against abuse by State or non-State actors, which should be assessed at each stage of the process.

A key factor in the proportionality assessment for an encrypted service, is the possibility of arbitrary surveillance of users who are not the target of the measures, sometimes referred to as "collateral damage". It's important to consider the big picture, rather than individual measures, and look at the regime that is being created and that Ofcom will oversee. The question is whether it creates "collateral damage" by interfering in an arbitrary way with the rights of innocent users. On an encrypted service, the creation of backdoors and systemic vulnerabilities and weaknesses is known to result in that kind of interference, as the ECtHR stated.

We feel the guidance would benefit from a redraft to clarify the proportionality of the measures and set out the fair balance of rights.

Q48.4 Ofcom's approach

Ofcom states correctly that:

> *"In designing our Codes, the Act requires us to have regard to several principles and objectives, and we must also consider our duties under the [...] the Human Rights Act 1998."* [48]

However, in setting out its approach to proportionality, Ofcom says:

> *"We have used proportionality as a key yardstick with which to decide whether to propose certain measures, and the final shape of those measures. To assess whether the proposals are proportionate, we considered the costs and benefits of different options, including how this might vary across services. As well as consideration of financial costs, the potential impacts on users – including both potential harm reduction and their human rights – was a central part of this assessment."* [49]

This appears to combine a human rights assessment with an assessment of economic proportionality (cost-benefit analysis for providers), offset against the anticipated "benefits" of content removal. The weakness of this approach is that it does not provide insights into the potential downsides and in so doing it misses some important outcomes that would influence decision-making.

[47] EDPB-EDPS Joint Opinion 04/2022 on the Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse, Paragraph 42.

[48] Volume 4 (11.5).

[49] Volume 4, (11.26).

Q48.5 Proportionality - encrypted services

The case of *Podchasov v. Russia*[50] sends a clear signal that the measures Ofcom is proposing would in all probability be unlawful on encrypted services. Although there are some obvious differences in the specific measures that were challenged in the court, the principles set out in the judgement would apply to any instance where a service was asked to break, weaken, or compromise end-to-end encryption, or introduce backdoors, weaknesses, or vulnerabilities.

Central to the case was the proportionality test and the court's reasoning as to why breaking end-to-end encryption would be disproportionate. In a nutshell, it is not possible to monitor specific content on an end-to-end encrypted service without creating indiscriminate interference with the privacy of other users who are not the target of the measures.

The ECtHR judgement confirmed that regardless of the technology choice, the screening of uploaded content from every user on the system engages privacy and free expression rights. This is because it is not possible to monitor specific users' content, without arbitrarily affecting others on the network. To read the content of one user, providers have to install software—either through a backdoor on the server or on the end-user devices—that will indiscriminately impact all users.

The reasoning for the ruling went along these lines: end-to-end encryption is a privacy-protecting tool that protects the integrity and confidentiality of communications and in doing so keeps individuals safe from attacks on their messages by hackers and other malicious actors. The content of the message is protected from everyone, even the platform provider, and only the sender and recipient can read it. This puts a barrier in the way of identifying the targeted material. It can only be identified by intercepting the communication, and reading the message in clear text, which involves either reading it before it is encrypted, or decrypting the message in transit. Potentially, if on-device scanning (client-side scanning) is deployed, the scanning software and database would be held on the users' smartphones. All of this is without the users' consent.

Mapping this onto the Online Safety Act, long-standing protections for British citizens against State intrusion into their private lives could be undermined if such measures were required. A UK government-sponsored study of proof-of-concept tools for scanning encrypted services stated:

> *"...from a Human Rights perspective, the confidentiality of the E2EE service users' communications cannot be guaranteed when all content intended to be sent privately within the E2EE service is monitored pre-encryption."* [51]

These measures would engage privacy rights because they require providers to intercept and scan communications. They are very broad powers, without a warrant or suspicion that the individual has

---

[50] European Court of Human Rights, Podchasov v. Russia. Judgement 13 February 2024 https://hudoc.echr.coe.int/eng/#{%22itemid%22:[%22001-230854%22]}. See also Third-Party Intervention by European Information Society Institute (EISi) https://husovec.eu/wp-content/uploads/2021/10/Podchasov-v-Russia-Brief.pdf

[51] REPHRAIN: Towards a Framework for Evaluating CSEA Prevention and Detection Tools in the Context of End-to-end encryption Environments: a Case Study. February 2023

committed a crime. They would engage the right to freedom of expression because they may deter people from speaking, creating a 'chilling effect'.

They also lack safeguards, something which regulators in other jurisdictions, such as Australia, consider important.[52] Similar measures proposed by the EU have been put on hold following a political compromise adopted in the European Parliament that explicitly rules out any scanning of encrypted content.[53]

A significant percentage of the population could be affected: two-thirds of the UK adults say that WhatsApp is their main communications service.[54] A legal opinion from a leading barrister concludes that these measures are unlikely to be in accordance with the law and would be open to challenge on the basis of that they would constitute a disproportionate interference with privacy rights.[55]

Volume 4, Question 20 Automated content moderation user to user

"Automated content moderation user to user i) Do you agree with our proposals? i) Please provide the underlying arguments and evidence that support your views."

We will respond to the following elements of the proposal:

> *"We propose to recommend that certain types of service should use an automated technique known as hash matching to analyse relevant content to assess whether it is CSEA, [...] We propose to recommend that certain types of service should use an automated technique known as URL detection to analyse relevant content to assess whether it consists of or includes a CSEA URL, [...] These proposals only apply in relation to content communicated publicly on U2U services, where it is technically feasible to implement them. Consistent with the restrictions in the Act, they do not apply to private communications or E2EE communications. In Annex 9 to this consultation, we have set out draft guidance which is intended to assist services in deciding whether content has been communicated "publicly" or "privately" for this purpose."* [56]

We welcome the clear statement that Ofcom will not be recommending these proposals for E2EE services. We welcome the acknowledgement that automated content moderation technologies can interfere with freedom of expression. However, we are not convinced that these intentions are fully embedded in the overall approach in the Guidance. As in our response to Question 1, we observe a contradictory position, and we underscore the reasons why these proposals would be dangerous for the security of encrypted services.

---

[52] Australian government Independent National Security Legislation Monitor: Trust but Verify, A report concerning the Telecommunications and Other Legislation Amendment (Assistance and Access) Act 2018 and related matters

[53] "Historic Agreement on Child Sexual Abuse Proposal (CSAR): European Parliament Wants to Remove Chat Control and Safeguard Secure Encryption." Patrick Breyer, 14 Nov. 2023, www.patrick-breyer.de/en/historic-agreement-on-child-sexual-abuse-proposal-csar-european-parliament-wants-to-remove-chat-control-and-safeguard-secure-encryption-2/.

[54] "Whatsappening in the World of Online Communications?" Ofcom, 25 Oct. 2023, www.ofcom.org.uk/news-centre/2023/whatsappening-in-the-world-of-online-communications.

[55] Index on Censorship/ Matthew Ryder KC: Surveilled and Exposed: how the Online Safety Bill creates insecurity, November 2022.

[56] Volume 4, Chapter 14.

Q20.1 Our response

We are grateful for the clear statement that the proposals for automated content moderation, such as perceptual hash matching techniques, do not apply to private or E2EE communications. The Guidance admits that it is not technically feasible to scan the content of E2EE services.[57] We strongly support Ofcom in standing by this statement.

We welcome Ofcom's proposals, to hold back on the deployment of automated content moderation technologies that engage the right to freedom of expression, acknowledging the strong risk of interference.[58]

It is not clear whether these intentions are embedded in the overall Guidance. Providers of encrypted services are asked to make a risk assessment of encrypted services for around a dozen different offences.[59] Content and behaviours are described that could only be addressed by the automated systems. We interpret this as an indication of a future direction, where Codes of Practice could ask providers to monitor for these offences. These two positions present a mixed message.

Please also see our response to Question 21: We do not find that there is a sufficiently clear definition of "content communicated privately".

We would appreciate clarification from Ofcom on these related issues.

Q20.2 Specific issues for encrypted services (client-side scanning)

We underscore the reasons why perceptual hashing and other automated moderation techniques cannot be recommended. There is a lack of evidence regarding the accuracy of hash matching, false positives, and security vulnerabilities, and this is acknowledged in the guidance.

If perceptual hashing were ever to be required for an encrypted service, the scanning of images would either take place on the server or on the users' devices (also known as client-side scanning). Please see our response to Question 2 regarding systemic risks of the technologies to run perceptual hashing on devices or servers for encrypted services. The updates to the device software could be subverted to install malicious code. The database could be corrupted with unauthorised material and the client software could be corrupted. For example, a secondary purpose of facial recognition could be hidden in the system. Such a capability would raise serious concerns that millions of smartphone users could be targeted without their knowledge.[60] It could be done by state or non-state actors.

These kinds of risks are raised by client-side scanning. They go to the heart of our democracy. If the database and scanning software is located on the users' devices, the question of governance must be more strongly addressed in Parliament, not by the regulator. It raises many challenges which go deeper

[57] Volume 4, Section 14.16.

[58] Volume 4.

[59] Schedule 5, 6, and 7 of the Online Safety Act.

[60] Volume 4, footnote 197 Jain, S., Cretu, A., Cully, A., and de Montjoye, Y., 2023. Deep perceptual hashing algorithms with hidden dual purpose: when client-side scanning does facial recognition.

than the Guidance's[61] focus on mitigating the risks outlined in Volume 2, but fail to address risks that could arise out of a system implementation. The challenges are around the transparency and oversight of the database, the scrutiny of algorithmic change, and the audit procedures. There should be policies and controls to prevent re-purposing and we would urge Ofcom to strengthen its governance requirements to include a comprehensive transparency and oversight regime.

[61] Annex 15 or in Volume 3, Chapter 8.