Ofcom Call for Evidence: First Phase of Online Safety Regulation

Global Encryption Coalition Steering Committee Submission
September 2022

The Global Encryption Coalition welcomes the opportunity to provide comments on Ofcom's call for evidence to strengthen its understanding of the range of approaches and techniques platforms can employ to help them meet their proposed duties under the Online Safety Bill.

Your response

Please refer to the sub-questions or prompts in the annex of our call for evidence.

Question	Your response
Question 1: Please provide a description introducing your organisation, service or interest in Online Safety.	*Is this response confidential? – Y / N (delete as appropriate)* The Global Encryption Coalition (GEC) was launched in 2020 to promote and defend encryption in key countries and multilateral fora where it is under threat. It also supports efforts by companies to offer encrypted services to their users. With more than 300 members in 95 countries, the Coalition is led by a steering committee made up of three global organisations: the Internet Society (ISOC), Global Partners Digital (GPD) and the Center for Democracy and Technology (CDT). GEC Members and Friends of the Coalition support the GEC's founding statement: *Encryption is a critical technology that helps keep people, their information, and communications private and secure. However, some governments and organisations are pushing to weaken encryption, which would create a dangerous precedent that compromises the security of billions of people around the world. Actions in one country that undermine encryption threaten us all. As a global coalition, we call on governments and the private sector to reject efforts to undermine encryption and pursue policies that enhance, strengthen and promote use of strong encryption to protect people everywhere. We also support and encourage the efforts of companies to protect their customers by deploying strong encryption on their services and on their platforms.*

Question 2: Can you provide any evidence relating to the presence or quantity of illegal content on user-to-user and search services?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 3: How do you currently assess the risk of harm to individuals in the UK from illegal content presented by your service?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 4: What are your governance, accountability and decision-making structures for user and platform safety?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 5: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 6: How do your terms of service or public policy statements treat illegal content? How are these terms of service maintained and how much resource is dedicated to this?	*Is this response confidential? – Y / N (delete as appropriate)*

Question 7: What can providers of online services do to enhance the transparency, accessibility, ease of use and users' awareness of their reporting and complaints mechanisms?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 8: If your service has *reporting or flagging*mechanisms in place for illegal content, or users who post illegal content, how are these processes designed and maintained?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 9: If your service has a *complaints*mechanism in place, how are these processes designed and maintained?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 10: What action does your service take in response to *reports* or *complaints*?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 11: Could improvements be made to content moderation to deliver greater protection for users, without unduly restricting user activity? If so, what?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 12: What automated moderation systems do you have in place around illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*

Question 13: How do you use human moderators to identify and assess illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 14: How are sanctions or restrictions around access (including to both the service and to particular content) applied by providers of online services?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 15: In what instances is illegal content removed from your service?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 16: Do you use other tools to reduce the visibility and impact of illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 17: What other sanctions or disincentives do you employ against users who post illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*

Question 18: Are there any functionalities or design features which evidence suggests can effectively prevent harm, and could or should be deployed more widely by industry?	*Is this response confidential? – N (delete as appropriate)* As described in a DRCF report from a roundtable held by Ofcom, the ICO and the FCA (January 2022), there are a variety of functionalities that can help prevent harm. These include a 'safety by design' approach that focuses on ensuring users can control the data they receive and share. For example the report suggests "a safety by design approach focusing on preventing online services being used for illegal activity; User controls for blocking or verifiable reporting within E2EE environments; Flagging and removing accounts that violate platform standards (in a transparent manner); The use of non content signals such as metadata to identify and address suspicious behaviour (*where the interception, retention and searching of related communications data should be analysed by reference to the same safeguards as those applicable to content- * *our addition*); and accessing the end-user device". Client-side scanning – which is a method that scans message contents on the user's phone, tablet or mobile device, either on the user device or on a remote server – should not be employed to prevent harm. This is because, as noted in the paper "Breaking Encryption Myths," client side scanning "increases the "attack surface" for encrypted communications by creating additional ways to interfere with communications - including by manipulating the database of prohibited content. The method is disproportionate, and is too easily misused to scan for content beyond the original purpose it was created for. This threatens mission-creep from authorities, creates opportunities for criminals to hack communication channels and could allow hostile state actors to surveil the communications of persons of interest. These systems are also prone to false positives and negatives, and create the conditions for censorship, and undue interference with user rights to freedom of expression and privacy. By breaking the expectation of privacy between sender and receiver, client-side scanning breaks the end-to-end encryption trust model, directly putting users' confidentiality at risk, and indirectly undermining trust in online services. Tech Against Terrorism's latest report on encryption includes an overview of the security risks, privacy violations, jurisdictional challenges and longer-term normative risks of breaking

end-to-end encryption through client-side scanning (see pgs 82 - 84)..

In 2021, Apple abandoned plans to apply client-side scanning on its devices, with the aim of addressing the use of its services for CSEA, as it was deemed disproportionate, insecure and unworkable.

Question 19: To what extent does your service encompass functionalities or features designed to mitigate the risk or impact of harm from illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 20: How do you support the safety and wellbeing of your users as regards illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 21: How do you mitigate any risks posed by the design of algorithms that support the function of your service (e.g. search engines, or social and content recommender systems), with reference to illegal content specifically?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 22: What age assurance and age verification technologies are available to platforms, and what is the impact and cost of using them?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 23: Can you identify factors which might indicate that a service is likely to attract child users?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 24: Does your service use any age assurance or age verification tools or related technologies to verify or estimate the age of users?	*Is this response confidential? – Y / N (delete as appropriate)*

Question 25: If it is not possible for children to access your service, or a part of it, how do you ensure this?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 26: What information do you have about the age of your users?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 27: For purposes of transparency, what type of information is useful/not useful? Why?	*Is this response confidential? – N (delete as appropriate)* As the 2022 DCRF report previously cited states, technological remedies themselves cannot provide a comprehensive solution to ensuring safety in an E2EE environment. Businesses and regulators should not overlook other approaches, such as incorporating user safety into the design and development of E2EE services. There are limits to technological interventions in tackling illegal activity". The report also highlights the need for regulators to provide clarity about their expectations re: privacy, safety and security. This includes the need to be more specific about threats and harms faced by users, desired outcomes and technological measures. Transparency measures that companies should provide include: ● Transparency reports that provide aggregated data and qualitative information about moderation actions, disclosures, and other practices concerning user generated content and government surveillance; ● User notifications about government demands for their data and moderation of their content; ● Access to data held by intermediaries for independent researchers, public policy advocates, and journalists; and

	• Public-facing analysis, assessments, and audits of technology company practices with respect to user speech and privacy from government surveillance Note: these transparency measures are also provided in GPD's response to this Call for Evidence (https://www.gp-digital.org/news/gpd-input-to-ofcoms-call-for-evidence-on-online-safety-bill-roadmap/).
Question 28: Other than those in this document, are you aware of other measures available for mitigating risk and harm from illegal content?	*Is this response confidential? – N (delete as appropriate)* Note: these recommendations are also provided in GPD's response to this Call for Evidence (https://www.gp-digital.org/news/gpd-input-to-ofcoms-call-for-evidence-on-online-safety-bill-roadmap/ Yes, some of these measures include: • Deploying counter speech against harmful speech, whether through funding or supporting counter speech projects and initiatives, or through developing automated tools which can generate effective counter speech; • Redirecting users who are searching for or consuming illegal or damaging content, such as terrorist content or CSAM, towards alternative content such as helplines or resources; • Ensuring that private or encrypted services have clear and accessible user complaints mechanisms allowing users to report content shared on the private or encrypted channel that they think is violative of the terms of service. This ensures that online service providers can continue to provide end-to-end encryption, which provides security to online activities and communications and protects data from potential malicious actors – which is particularly important for the protection of vulnerable groups, including LGBTQ+ persons, survivors of domestic violence and human rights defenders – while also

ensuring that illegal or harmful content is not left unchecked on those channels;
- Allowing users to customise their own moderation rules beyond what is prohibited in the terms of service, such as Twitter's Bodyguard tool, which allows users to set their own moderation rules;
- Allowing users to block content from particular people or groups, or on particular topics, or content from unverified or anonymous accounts, such as Twitter's Block Party tool;
- Allowing users to limit their own discoverability, or to have invisible or anonymous accounts;
- Developing software that helps users to review, document and export repeated instances of illegal or harmful content online, such as Google Jigsaw's Harassment Manager tool;
- Allowing users to flag what they believe are underage accounts;
- Implementing additional privacy-by-default settings for children's accounts, such as only allowing their content or profile to be visible to or engaged with by their friends or contacts;
- For younger children, developing parental controls to allow adults to have control over what types of content is encountered, particularly for vulnerable children.

As a note on methodology: a critical factor in mitigating risk and harm from illegal content is clarity about the problem to be solved. The Online Safety framework runs the risk of failing to define the problem with sufficient clarity, with the result that proposed solutions don't work, and/or have unintended and harmful consequences.

First, it is overwhelmingly unlikely that any single technical "fix" will successfully neutralise all societal ills that materialise on the internet.; For many of the Government's intended aims in this policy area , not all of which include criminality, the appropriate intervention is not technical at all, but a matter of user education, awareness-raising, and digital and ethical literacy.

	A disproportionate focus on technical “fixes” is not the most effective solution to what are essentially societal problems.

Please complete this form in full and return to OS-CFE@ofcom.org.uk