Your response

Question (Volume 2)	Your response
Question 6.1: Do you have any comments on Ofcom's assessment of the causes and impacts of online harms? Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.	*Is this answer confidential? No* The risk analysis carried out by Ofcom is comprehensive in addressing the various issues that can impact the safe use of the Internet. Considering the myriad of topics, we would like to emphasise how encryption is treated in Ofcom's evaluation. As highlighted by Ofcom, encryption plays an essential role in information security as it ensures the integrity, confidentiality, and availability of data. For this reason, its use permeates numerous fields, such as protecting private communications, securing web traffic (HTTPS / SSL), storing information, among many others. Considering these applications, services with encryption are employed by various sectors, such as businesses and governments. Ofcom's analysis of encryption is based on the 'Going Dark' perspective. According to this narrative, security forces are supposedly being hindered by the use of encrypted devices. As digital rights researchers, we would like to offer another perspective on encryption as a guarantor of Human Rights. Various international organisations, civil society entities, and scholars point to the importance of encryption for the realisation and strengthening of various rights. The report produced by UNESCO, entitled "Human Rights and Encryption," highlights how the use of encryption is being applied by various actors in society. Among the rights reinforced by encryption are freedom of expression, anonymity, access to information, private communication, and privacy. In the 2022 report "The Right to Privacy in the Digital Age," by the Office of the United Nations High Commissioner for Human Rights, a section is dedicated to assessing restrictions on the use of encryption by governments. The report evaluates that there is a significant impact on the right to privacy and other human rights. It also offers the case of armed conflicts, such as in Ukraine, where the use of end-to-end encryption apps was essential for the communication of civilians. In this same direction, we point out the documentary produced by WhatsApp "We Are Ayendas," which demonstrates how, after the Taliban's rise to power in Afghanistan, the encrypted platform was essential for women players to communicate and escape the country. Other materials are significant to demonstrate the importance of encryption for Human Rights. The English organisations Child Rights International Network (CRIN) and Defend Digital Me published in 2023 the report "Privacy and protection: A children's rights approach to encryption." In it, CRIN and defenddigitalme consider how the use of encryption negatively and positively impacts the lives of children. The report points out how its use brings new challenges, but also opportunities. Thus, the report points out several benefits for children in the use of encryption to guarantee their rights and argues that policymakers should keep in mind the rights of children when legislating on any law that may impact the use of encryption. We emphasise that policies affecting the Internet can have extraterritorial effects, as it is a global network that connects local and regional networks. Thus, weakening encryption can cause harm not only in the UK

Question (Volume 2)	Your response
	but in several regions of the planet. This is the case with WhatsApp blockages in Brazil between 2015 and 2016. In addition to the enormous social and economic harm caused by the blockage of the most used digital communication tool in the country, several neighbouring nations also had the service blocked. In this sense, weakening or blocking encryption and other techniques that ensure security and privacy on the Internet can lead to the fragmentation of the Internet, causing even more social and economic harm. Considering that encryption is especially important for vulnerable groups, such as women, the LGBTIQA+ population, and children, these groups are even more vulnerable in the Global South. Therefore, Internet fragmentation and weakening encryption can amplify the social and economic harm to these groups in peripheral nations.
Question 6.2: 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.	*Is this answer confidential? No* Ofcom's risk assessment places encryption as a major risk for users. In a quick analysis of frequency, in over 90% of encryption mentions in the material it was associated with a negative description. In three other cases, Ofcom associated the positive use of encryption with a caveat about its dangers. We identified only one case in which Ofcom was able to recognize the importance of encryption on the Internet, in the passage from page 03: "End-to-end encryption plays an important role in safeguarding privacy online. Pseudonymity and anonymity can allow people to express themselves and engage freely online." It is important to note that the use of encryption can bring challenges to regulatory environments. However, we cannot treat its use only from negative perspectives that seek to criminalise its use. Ofcom's report seems to rely on the "Going Dark" narrative, in which security forces are becoming obscured by the use of end-to-end encryption. This argument ignores that security forces also have powerful technological tools for investigation, such as those that bypass encryption systems, like Cellebrite's UFED. The practice of governmental hacking was the subject of investigation by IP.rec, in the report "Merchants of Insecurity: conjuncture and risks of governmental hacking in Brazil", which identified that the use of hacking devices is widespread throughout the country. In the work, we indicated how the use of these tools puts at risk an entire ecosystem of electronic device security, strengthening a parallel market for vulnerabilities in systems, which are sold to companies like NSO Group and Cellebrite. It is important to reinforce that one of Cellebrite's offices is precisely in the UK. First, communication services with end-to-end encryption have content moderation mechanisms for unencrypted information. Thus, such services are able to ban users, groups, who are sharing inappropriate images and content, such as CSAM, without even accessing the transmitted information. The report produced by the Center for Democracy and Technology (CDT), "Outside Looking In: Approaches to Content Moderation in End-to-End Encrypted Systems", reviews methods that platforms use to moderate content in end-to-end encrypted systems and highlights the importance of user agency in reporting inappropriate content and also in using metadata for moderation. Furthermore, it assumes that many times when services with encryption are used by criminals, this is due to technical privacy knowledge - which is not always true. Several factors need to be considered, which demand a more detailed analysis of this use, such as the application's market share, the users' perception of private communications, digital literacy, among others. In Brazil, several scams are carried out through the WhatsApp application, not because of its

Question (Volume 2)	Your response
	encryption function - which many users are not aware of - but because it is one of the main forms of communication in the country. In fact, without strong end-to-end encryption, there would be a higher number of digital fraud and theft. It is necessary to draw attention to the fact that Ofcom's risk assessment indicates that encryption "stands out as posing a particular risk." However, this assessment expands the scope of risks far beyond the dissemination of CSAM, including 12 other categories of crime, respectively, hate crimes, terrorism, drugs, immigration, sexual violence, extreme pornography, abuse of intimate images, proceeds of crime, fraud, foreign interference, and false communications. By employing the inaccurate argument that encryption reduces the chance of detecting crimes and is therefore adopted by criminals, Ofcom is asking service providers to assess the likelihood of various types of illegal harm occurring on their services, expanding the obligation to not only investigate the presence of illegal content but also of crimes being committed or facilitated in these services. Adopting this approach will mean the worrying expansion of the scope of what services with encryption will have to monitor. In this way, the function creep foreseen and denounced by various organisations, experts, and service providers would be realised, bringing worrying violations of the right to privacy, as well as an increase in the risks of errors and false positives that are already existing and previously denounced.

Question (Volume 4)	Your response
Question 14.1: Do you agree with our proposals? Do you have any views on our three proposals, i.e. CSAM hash matching, CSAM URL detection and fraud keyword detection? Please provide the underlying arguments and evidence that	*Is this answer confidential? No* According to experts such as Professor Susan Landau, Professor in The Fletcher School and Tufts School of Engineering, Department of Computer Science, the current technique for recognizing photos of known CSAM uses a technology called “perceptual hashing,” which is a method that allows an image to be recognised even after a small change. PhotoDNA, YouTube’s CSAI Match, Apple’s NeuralHash, Meta’s PDQ and TMK-PDQ are all examples of perceptual hashes. However, as Professor Landau, we believe that even though this technique might work with static images, video files might be difficult to analyse as they have far more bits than a photo. Therefore, due to the amount of data, using perceptual hashing on each frame of a video to match it with videos of known CSAM wouldn’t work as this would take a considerable amount of time. In order to address this issue techniques that involve sampling frames or splitting the video in small segments were created as alternatives, those techniques used, for example, for the immediate removal of real-time filming of terrorist attacks. Facebook’s TMK+PDQF uses a video-hashing algorithm to recognize matches with known offensive content such as terrorist attacks or CSAM.

Question (Volume 4)	Your response
support your views.	Moreover, it has been observed by experts that perceptual hashes can be fooled. False negatives are the main problem related to the use of this method, i.e., when the system does not identify a photo or video when, in fact, there is CSAM in it. In fact, experts suggest that this tends to happen quite often as it is not hard to manipulate an image so that its perceptual hash appears to be quite different from the image containing CSAM - even though the image still contains the illegal content. Moreover, a high percentage of CSAM images and videos cannot be recognized by perceptual hashing techniques as there is no a priori known image. In the article *Bugs in our Pockets: The Risks of Client-Side Scanning*, a group of experts observed, perceptual hashing is vulnerable to "adversarial attacks," i.e., deliberate malicious efforts to fool the algorithm. In two weeks researchers reverse engineered Apple's NeuralHash algorithm leading to a breach, producing a hash "collision", i.e., two images that look nothing alike but have the same perceptual hashes. It is important to stress that such capabilities are available for anyone. As computer scientists Carmela Troncoso and Bart Preneel observed, "In the arms race to develop such detection technologies, the bad guys will win: scientists have repeatedly shown that it is easy to evade detection and frame innocent citizens." Once again, Professor Landau reminds us that the impact of false positives can be quite serious on those accused. While for some types of criminal investigations, once the person is cleared, the taint may go away, that is often not the case for accusations of CSAE. Such accusations come with real—and often permanent—costs for the innocent people who have been accused.

Question (Volume 4)	Your response
Question 14.3: Do you have any relevant evidence on: An effective application of hash matching and/or URL detection for terrorism content, including how such measures could address concerns around 'context' and freedom of expression, and any information you have on the costs and efficacy of applying hash matching and URL detection for terrorism content to a range of services.	*Is this answer confidential? No* The use of perceptual hash matching technology for detecting terrorist content presents more risks to users' freedom of expression than other hash detection modalities. In its own text, Ofcom's public consultation argumentation highlights the possible inaccuracies and vulnerabilities of this technology, with a special focus on its use in detecting CSAM. The text emphasises the high importance of the quality of the databases that will feed the tool and the risks that inaccurate data could pose to the effectiveness of security tactics and to users' rights. In the case of detecting terrorist content, which is often difficult to correctly classify due to sociopolitical nuances, the chances of inaccuracies in the databases increase. The integration of perceptual hash matching technology with machine learning and artificial intelligence tools exponentially multiplies the consequences of errors in the databases, as it increases the likelihood of the system reproducing racist and xenophobic biases, resulting in a high rate of false positives and false negatives. Issues of political and religious belief, sociocultural and geographical origin, and geopolitical conflicts that produce public animosity could penalise already vulnerable users (racialized people, non-Christians, from the Global South, etc.) who have no connections to terrorist activities and ignore effectively terrorist publications (made by white, Christian, Northern Global users, etc.). Known or biassed data and processes cannot serve as a basis for user discrimination. We do not recommend the use of ML and AI technologies as they bring known racist biases that could threaten users' freedom of expression and negatively impact the accuracy of hash detection. Perceptual detection alone brings vulnerabilities that could threaten users' rights, which is why we do not recommend its adoption in its current state of development. Therefore, we recommend: 1. Detecting, blocking, and reporting URLs already identified as related to terrorist content is a more effective tactic for combating terrorist organisations that operate through multiple platforms while presenting lower risks of errors that could threaten users' freedom of expression; 2. Providers should favour a design and service culture that encourages reports made by users themselves; 3. Human moderation activities should receive greater attention and investment so that: a) no measure that could impact users' freedom of expression and privacy is taken without human supervision; b) measures are taken quickly, especially in the case of live broadcasts of terrorist attacks.

Please complete this form in full and return to IHconsultation@ofcom.org.uk.