Open Rights Group

Response

Ofcom Online Safety Call for evidence: First phase of online safety regulation

12 September 2022

Preliminary question

Q1. Please provide a description introducing your organisation, service or interest in Online Safety.

Open Rights Group (ORG) is the leading UK-based digital campaigning organisation. We work to protect fundamental rights to privacy and free speech online, including data protection, the impacts of the use of data on vulnerable groups, and online surveillance. With over 20,000 active supporters, we are a grassroots organisation with local groups across the UK. We have worked in this policy field throughout the 'online harms' processes and consultations, and both Digital Economy Acts (2010 and 2017), accurately highlighting which parts of both DEAs would prove extraordinarily difficult to implement practically or fairly.

We have experience of dealing with users whose content has been restricted by content moderation systems and have published research which has informed our responses. In 2019, we published an extensive piece of research investigating filtering by broadband providers and its impact on users who were blocked. Entitled Collateral Damage in the War Against Online Harms our report examines what gets blocked, the scale of errors, complaints made by users, appeal processes and responses from providers. The research data was provided by1881 unblock requests which were made through our Blocked.org.uk tool from 2017-2019. Since the project began in 2014, we indexed over 35,000,000 websites, creating a database of over 760,000 blocked websites, allowing users of the site to search and check domains which could be blocked.

A second piece of research was published by our Policy Manager – Freedom of Expression in a personal capacity before joining ORG. This research examines shadow bans, the demotion or limiting distribution of content through recommender systems. It focusses on Facebook Pages that experienced these restrictions and conducts analysis using Facebook's own data provided by the users, to reveal how shadow bans are actioned by the platform and how they interfere with freedom of expression. This research can be found here Algorithms Patrolling Content: Where's the Harm?

We would be willing to share our more detailed findings with Ofcom.

We are responding to Questions 5, 7,11,14, 18, 27, 28. Our responses in **blue**.

Terms of service and policy statements

Q5. What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements?

ORG response to Q5: We would like to comment on the way that certain large online platforms write their terms of service and the need for precision drafting. We look at it from the users' viewpoint.

A case in point is Facebook's "Dangerous individuals and organisations" in its Community Standards. Facebook does explain that users should not glorify or praise terrorist or other dangerous organisations, or post images linked to a terrorist organisation. However, the platform restricts their content even when the users posting it believe it should be clear that they have posted with the aim of rightfully discussing it or expressing a view. Examples we know of include a photograph from the Syrian conflict showing a war crime and a flag of Isis, a known terrorist organisation. We've also seen restrictions of posts commenting on contemporary politics using quotes from the infamous Nazi propagandist Joseph Goebbels, as well as a Facebook post restricted for discussing an article about the case of Shamima Begum.

The effect on these users, whose content has been restricted under this Community Standard, is that they frequently feel outraged because they are not terrorists.

These cases highlight some of the difficulties in drafting terms of service in ways that users understand, and that also align with the way that the artificial intelligence systems and algorithms are coded. The danger is that under the terms of the Bill, Ofcom would hold tech companies to something that does not work for the users.

There is a wider issue that concerns not only the content to be restricted, but the nature of the restrictions themselves. Based on our own experience of dealing with users who have been restricted, we would say as follows: terms of service need to be very clear and precise about the terms of any restrictions – what exactly the action is, the justification or criteria for it, the effect that the user will notice, its duration, whether the duration could be extended and how many extensions could be imposed, and any other consequences (for example, the account will be suspended after three "strikes"). It is possible to collate a large chart of options for restrictive actions.

The terms should also be clear how the user may appeal any restrictive actions, the process for the appeal, and any further options. Users should also be offered a possibility to contact platform staff and raise any concerns they have about restrictions.

Reporting and complaints

Q7. What can providers of online services do to enhance the transparency, accessibility, ease of use and users' awareness of their reporting and complaints mechanisms?

ORG response to Q7: We would like to comment on complaints mechanisms. However we will comment not on what providers can do, but on what Ofcom has a duty to do. Our comments apply regardless of the type of user, whether registered or not, or whether child or adult. Our focus is on the users' experience.

Terminology: We are particularly concerned with users who are falsely or wrongfully sanctioned (as per Question 14). By this we mean users who have posted or communicated lawfully, but stand accused of posting or communicating unlawfully. In many cases, this is due to error rates in the content moderation systems themselves (see our response to Question 14). In these cases, it is more appropriate to refer to an appeals process than a complaints mechanism. This is not an issue about a purchase, or a faulty product, but a restriction placed on someone's human rights. These appeals should be dealt with diligently, using due process.

Human rights standards: The process of restricting unlawful speech is governed by the European Convention on Human Rights Article 10, paragraph 2. Under this provision, restriction must be prescribed by law, be targetted to meet a specific aim, and be the least restrictive option available. In terms of online content, courts have interpreted this as, for example, specifying a precise url for the content, and only blocking or removing that piece of content, not a whole account or website. The justification would usually include an examination of the context - the intention of the speaker, the social and political environment - as well as the way the technology operates.

The Bill as it stands does not recognise that unjustified interference with freedom of expression can occur as a result of the measures it puts forward. It does not even include freedom of expression as a right that applies to all individuals. This is deeply problematic and ideally it should be rectified on the face of the Bill. Failing that, Ofcom could seek to rectify it by introducing due process into the Codes of Practice, notably with procedural measures that would serve to safeguard the free speech rights of users who are wrongfully restricted.

Procedural safeguards: The Bill as it currently stands has a weak complaint mechanism, with no specification as to how the process should operate. [Clause 18] It could be nothing more than an email address. This is not good enough, whether one is considering this from a consumer standpoint, or human rights standards.

Under human rights standards, all individual users are entitled to an effective complaints process. It should enable users to challenge decisions regarding their content or their accounts, and give them a right to be heard. They should have a right to administrative or judicial redress. This process should be a statutory one, on the face of the Bill. Failing that, Ofcom should be sure to define it in such a way that it will be clear to both online platforms and users, what should be expected.
We would like to see online platforms given a duty to protect, respect and advance users' rights to freedom of expression, supported by the following *ex ante* and *ex post* procedural safeguards:

- Users have a right to an effective appeals process and remedy, when online platforms make decisions to take down content which they generate, upload or share, or in any other way restrict access to their content or account.
- Online platforms should notify the user of the specific content to be removed or restricted such as the URL or a screen shot; or where relevant, the nature of the restriction, the length of time that the restriction will stay in place and when it will be lifted.
- Online platforms should provide a clear and specific statement of reasons for that decision including the rule or the law that was used, with evidence, and information regarding how the decision was taken, whether algorithmically or by human intervention
- Online platforms should state how users may appeal the decision, with the deadline to lodge the appeal, and inform the user of the possibility for judicial redress

The Bill's two-tier complaints process: The Bill creates a fast lane – or VIP lane – for media organisations to complain in the event that one of their social media posts were to be fall foul of the content moderation policies of the large platforms. These organisations would be able to contact the social media platforms, and be assured of a swift response, if a situation arose where their content or accounts were restricted in any way. A government amendment asks the online platforms to contact the user prior to taking it down, and not to restrict until they have had the opportunity to put their case. We note that the beneficiaries of this fast lane are certain large media organisations who want to protect their social media content which is part of their global commercial operation.

We believe this proposal would fall foul of human rights law. All users are entitled to due process where their content is being restricted and should be given the same possibility to challenge appeals as the media organisations are given.

Based on our understanding, we believe that Ofcom has not been asked to implement such a process due to the anticipated volumes of complaints. We would respectfully suggest that this is unfair and unreasonable for the many ordinary citizens who may be affected by the measures in the Bill, and that Ofcom should ensure that there is a fair appeals process available to all.

Online platform procedures: Currently, our experience of the large online platforms and their appeals process, is disappointing to say the least. Users often don't know how to complain, or where they do find an email address they get no reply. From our research, "Algorithms Patrolling Content: Where's the Harm" (see above) we know of one user, who, having endured two months of continued shadow bans, tried to find people on the platform teams who could help him. He did manage to communicate with them via email. However, the responses were less than helpful. For example, they stated that the restriction had been correctly imposed, and would have to run its course. So they admitted that it was a restriction, but would not lift it, or say what he had done or why it was imposed. We would be able to share some of this data with Ofcom.

This is supported by our experience on our Blocked project (see above). There was a minimal take-up of appeals where they were offered (in the case of broadband filtering they were only offered by mobile suppliers). Some of the issues highlighted were the referral of appeals to the third-party supplier of the filtering system, a variable process depending on the provider, and call-centre operatives being uninformed.

Content Moderation

Q11. Could improvements be made to content moderation to deliver greater protection for users, without unduly restricting user activity? If so, what?

ORG response to Q11: Content moderation refers to the seeking out, detection, identification, evaluation or assessment of content, and the enforcement action taken on the basis of that evaluation. We note that the definition in the Online Safety Bill of content moderation systems is insufficient and misleading as it fails to acknowledge this complete suite of functions.

The Bill also suggests that content moderation only applies when the platform is proactively seeking out content, but not when it is reported. This is not correct. The term "content moderation" applies regardless of how the platform finds out about the content.

This question appears to be driving at a specific response without explaining what is the intended functionality of content moderation systems that you would like us to comment on. When you refer to "greater protection for users" we are not sure if you are referring to restrictions placed on the content that is labelled as illegal or harmful and that users should not see. When referring to "not unduly restricting users activity" we are not sure what you mean.

However, we are concerned with how the users' experience content moderation Based on our understanding of some of the evidence provided the Parliamentary committees that examined this Bill, we think you may be referring to what we would call a shadow ban. Or at least, this may be one element of content moderation functionality that could be intended by this question.

Content moderation systems: We have identified several different methods of content moderation, and the decision to use a particular method may depend on the system. Previous systems implemented by broadband providers (see our Blocked research project) relied on a database of static urls that have been classified as falling into a category to be blocked.

With online platforms, the game has changed significantly. It's a dynamically shifting process, that continually seeks out posts uploaded across the entire platform by users every minute of the day, and even segments of those posts. These AI-driven systems use perceptual hashing techniques, or various forms of machine-learning that is not necessarily in a form recognisable to humans.

All of these systems operate to a system of rules that has been coded in, and tell the system what to do when it finds prohibited content.

Shadow banning: This is a method of restricting users applied by online platforms. It leaves the content live on the platform, but gives it no distribution. It may be deemed not to unduly restrict user activity, because they can still post, but their posts are seen by far fewer people than normal, and hence the user experiences a ghosting effect, and is put in the online shadows. A shadow ban does not relate to content but instead looks at the behaviour of an account as exhibited by the analytics data, as outlined in this research report.

Shadow banning is often presented as being a soft option, a mitigating measure when in fact it can be a draconian sanction that has a similar effect to removing the content entirely. We believe that shadow banning restricts content to almost the same level as taking it down, but without any reference to specific content that is unlawful. This makes it deeply problematic from a human rights perspective. We believe that shadow banning restricts content to almost the same level as taking it down, but without any reference to specific content that is unlawful. This makes it deeply problematic from a human rights perspective. It is a positive interference with freedom of expression. We would urge Ofcom to exercise caution in mandating or allowing shadow bans in Codes of Practice.

For more detail on shadow bans, please see our response to Question 14.

Upload filter: We are also concerned that compliance with the Illegal Content Safety Duties would require an upload filter. By this we mean an automated system that intercepts content as it being uploaded by the user, and follows the process of identification and evaluation, and enforcement action as outlined above. We see this as a possible interpretation of the requirement to "prevent" users from engaging with illegal content, and in that regard it may correlate to your question about delivering greater protection to users. However, the upload filter raises serious human rights issues.

This upload filter implies a de facto general monitoring obligation that was illegal under UK law until January 2020; and it is characteristic of prior restraint, a particularly draconian form of censorship where the content is banned before it has been published. In the online scenario, the content would be removed without the user being able to know. This is contrary to human rights standards. We would urge Ofcom to exercise caution in drafting codes of practice to avoid mandating an upload filter, and ideally not to mandate it at all.

Actioning content and sanctioning users

Q14. How are sanctions or restrictions around access (including to both the service and to particular content) applied by providers of online services?

ORG response to Q14: Once again, we are concerned with the users' experience of restrictions imposed by online platforms. Our remarks relate to the interference with users' free speech rights by the incorrect or unlawful restriction of their content. Before mandating any restrictions or drawing up a Code of Conduct for the application of restrictions, Ofcom should conduct detailed research into the nature of restrictions applied by online platforms. Ofcom may be surprised to find that there is a multiplicity of ways that restrictions can be applied, and that the method will vary from one platform to another.

We are pleased to see that Ofcom acknowledges that sanctions and restrictions will be imposed on users, and the possibility for inaccuracy of content moderation systems. The Bill does not do so. At most, it refers to "actions" taken by providers. For the most part, the Bill does not specify what actions would be taken by providers. However, we note that some actions are detailed in Section 13, and these are: taking down content, restricting users' access to the content, limiting distribution or merely "allowing" the content (giving it no distribution or promotion, as per Amendment 71 at Report Stage).

We would like to comment on some actions that we are aware are applied by social media platforms and also on error rates. Please see our response to Question 11 regarding safeguards.

Taking down content: this refers to content being removed from the user's account. It is not always clear whether it is deleted from the platform, or merely blocked, or made inaccessible to the user. In the context of messaging service, the term is something of a nonsense. They cannot take down the content, but they can remove it from the account.

Taking down or removing content is often accompanied by other measures on the users' account. Sometimes the account is also suspended, or the user is given what's called a 'strike', rather like a black mark against their account. If they accumulate 'strikes' they could be sanctioned in other ways. For example, 3 strikes means their account will be suspended for a week. It is not clear on what basis these actions are taken. Users are typically not informed, and just wait out the penalty until their account is back to normal.

Suspending or terminating users' accounts: Suspending a user's account means that they cannot access any of their content, they cannot post, and depending on the platform, they may also lose access to messaging services. Suspensions are usually for limited periods of time, which may vary by platform. Facebook, for example, sometimes uses a pattern of 3-10-30 day suspensions. Suspending an account is an interference with freedom of expression.

Terminating an account is a very serious interference with freedom of expression. The user loses all access to all services and functions associated with the account, including the analytics, and has no means of retrieving anything.

In our experience, social media platforms sometimes simultaneously take down content and suspending accounts. Sometimes an account is suspended along with a restriction on the content itself, or the platform uses a system of "strikes" as a heavier sanction against users whose content has repeatedly been restricted. We have seen is "three strikes and you're out", but mostly it seems to be that the number of strikes prior to suspension is arbitrary.

Millions of people in the UK use online platforms as an essential means of communication. They have replaced the old-fashioned telephone for day-to-day contacts. In this scenario, we have known people who, for example, could not make contact regarding a friend's funeral because their account was suspended. We also know it affects non-profit organisations who rely on these services. There is a case in Poland of an NGO supporting drugs rehabilitation, whose Facebook Pages were terminated without warning or explanation (SIN v Facebook). The Polish courts have determined that it is an interference with freedom of expression, and said that all existing Pages must be kept online, but the case has stalled.

Limiting the recommendation of the content : this could have two meanings depending on the nature of the service. We address it separately for social media platforms and for search engines.

Limiting recommendation on social media platforms: In the context of user to user services, it means a shadow ban. Social media platforms rely on recommender systems to distribute content that is posted by one user and show it to other users. A shadow ban operates through recommender systems. This is where the user's account is demoted, or is deprived of distribution, to the effect that traffic drops suddenly, without the user being informed. Depriving content of distribution in context of social media platforms, means that it is not shown to users in their newsfeeds and timelines, and its visibility is reduced. It is the recommender systems that determine what content is shown in those newsfeeds and timelines. In applying a shadow ban, the online platforms apply algorithms to those systems in order to drive the demotion of the content.

Shadow bans are applied in an arbitrary way. On Facebook they typically last 7 days, and may be extended for further periods of 7 days, or multiples of 7 days. Users are often not informed beforehand, they are not told the reason, and not informed when the ban is to be lifted. They know they are under a ban because their reach is vastly reduced, and they know the ban has been lifted because their reach starts to come back. A graph of a shadow ban shows a "U" shape, indicating how traffic to the Page plunges when the ban is instated, and goes sharply upwards when it is lifted. Facebook also operates partial shadow bans, or selective banns, where individual posts are limited, and others are given the usual distribution. Users only notice this when they see unusual patterns in their post reach analytics. The working paper "Algorithms Patrolling Content: Where's the Harm" draws on analytics data from 20 Facebook Pages, and outlines the key issues on shadow bans. We would be happy to share some of this data with Ofcom and present the research findings.

Shadow bans don't operate on the basis of the content, but on the basis of behaviour. This is not behaviour that a human would understand, but it is about patterns in the data that reveal the behaviour of the account. It is a technique rooted in computer security. It punishes the content, on the basis that the account is displaying behaviour that appears suspicious at the level of the data, but in fact may be perfectly genuine. The above-mentioned paper explains this in detail. The effect is almost the same as suspending the account. The user has no means of appealing, because they don't even know what they should be appealing. We believe that shadow bans are a draconian measure and a serious interference with freedom of expression. As a minimum, we advocate for the safeguards we recommend above in our response to Question 7.

Limiting the recommendation on search engines: In the context of search engine services, "limiting recommendation " means demotion or de-ranking, or de-indexing. Search engines are resistant to de-indexing, the permanent removal of a website or page from the search engine index, without a court order to protect them from liability.

The position of search engines is different from that of social media platforms. They have no relationship with the websites or pages that are in their index. The information is obtained by automated bots crawling the web. Therefore, if a website or page is demoted or de-indexed in a search listing, the search engine has no means of informing them and they would not know. The only way they could know is by a fall in traffic. This Bill therefore represents a serious interference with their freedom of expression.

Until now, UK services have been protected from liability if they have 'actual knowledge' that a piece of content is illegal or unlawful. Actual knowledge means, in simple terms, that they have received a notification from a lawyer or from law enforcement authorities, that the content is unlawful. This Bill puts the full liability on the provider to make the assessment and determine the action, subject to potentially large fines if they are deemed not to comply. There is an outstanding question as to the position if a user - who could be a website in the case of search engines – decided to take legal action for loss of traffic and revenue.

As an example of the harm that can be caused by this type of blocking, we draw on our Blocked project which has a similar effect to what we expect from search engine de-motion. It shows how blocking urls can damage businesses who rely on websites to bring in customers. For example, a wine merchant whose website was blocked by alcohol filters. There was even a small watchmaker and a Porsche consultancy who lost business through their website being wrongly filtered. Most of the businesses who contacted the project found out by accident that their websites had been included in filtering systems.

We urge Ofcom to proceed cautiously, and not to limit accessibility of content on the British Internet by mandating draconian measures like de-indexing unless a court has ordered it.

Allowing the content: This means merely allowing content to be on the platform without any form of distribution. We believe the effect will be more or less the same in its effect as a shadow ban. Other users would only find the content if they have the direct url, but it would not appear in newsfeeds or timelines.

Error rates and accuracy: When we consider the actions taken to restrict content, our concern is about error rates which can be quite surprising. Our Blocked project tested around 9,000 Scottish charity websites and discovered that around 50 of them were blocked by one or more providers. The project found 59 sites dedicated to supporting domestic violence or sexual abuse victims incorrectly blocked by filtering systems, as well as 40 LGBQT+ sites, 55 charity websites and 98 counselling sites. Erroneous blocking by filtering systems operated by broadband and mobile providers continues. A recent example from August 2022 was the blocking of the campaigning website wesayenough.co.uk, , by the mobile network Three and reported by ORG's Blocked project. The network provider said it was incorrectly classified, which raises questions around how classifications are made.

Error rates by AI systems are also reported as a problem, both for perceptual hashing systems and text-based classifiers. It's notable that Apple cancelled it's perceptual hashing system for monitoring CSAM material due to the considerable criticism of its error rates by cyber experts. Perceptual hashing is way to compare two different images via a digital fingerprint (known as as hash file). It is possible for two different images to generate the same hash or for a slightly different version of the same image to be missed. Error rates can vary even for the same type of image, and a quote accuracy rate of say 85%, does not necessarily mean a consistent finding.

We would urge caution for Ofcom in relying on error rates quoted by service providers or system vendors. A constructive approach would be to build in the procedural safeguards for users as we outline in Question 7.

Design of service …including functionalities and algorithms

Q18. Are there any functionalities or design features which evidence suggests can effectively prevent harm, and could or should be deployed more widely by industry?

ORG response to Q18: Once again, this question appears to be driving at a specific response, but we are unclear of the exact intention. We will therefore comment on some deeply troubling measures mandated by the Bill that we assume may be what Ofcom is thinking of.

The filter button: here we refer to the Bill's so-called user empowerment clauses which are claimed to protect vulnerable people from abusive or harmful content. In particular, there is a provision that mandates users being offered the possibility to see, or be seen by, only verified users. This is being positioned as a positive measure to protect users from potentially abusive comments. However, we believe these measures need to be more clearly thought through when Ofcom drafts its Code of Practice. We are not convinced that the measures suggested will empower users or protect them. For example, verification is no guarantee of integrity. A basic check to see who is verified on social media will reveal some insidious characters, who are unlikely to be helpful to people in a vulnerable situation. A bad actor, or a stalker, or ex-partner, could get their account verified and remain in a position where they could taunt their victim.

The people who are most likely to want to press a button and filter out unverified users are those in a position of power. This filter button would enable the rich and powerful, and influential people to escape scrutiny and to hide from the public discourse.

We do not believe this is a good idea, and call on Ofcom not to enforce it as mandatory.

It is not clear from the way the Bill is drafted whether the intention is to create an environment on these platforms where users will no longer be anonymous, on the basis that abusers hide behind anonymity. We believe that overall, anonymity also protects users and would urge Ofcom to protect the ability for users to be anonymous on online platforms if they so choose.

The Internet is premised on the possibility to be anonymous. For certain groups of people in society, anonymity is an essential protection from harm as they may be vulnerable. We would also highlight that it is used by many people who are not necessarily in a vulnerable category, but do not want their employer or their family to know their views on particular topics, such as politics. To force name-only accounts would lessen the vibrancy of our political discourse. For these reasons, we would call on Ofcom to hold back from asking providers to go down the route of removing anonymity.

Age- assurance or age verification: The issue that most concerns us here is the use of biometric data to guess or estimate the age of a person, whether child or adult. Such data may include hand or head measurements, or voice. This is a very new area of technology. Its effectiveness or the ability to implement it at scale has not been proven. The governance structures are not in place and the likelihood of abusive or unlawful practices is high. We urge Ofcom not to mandate the implementation of these systems, even though they are on the face of the Bill, at least until the governance and regulatory issues have stabilised.

Transparency

Q27. For purposes of transparency, what type of information is useful/not useful? Why?

ORG response to Q 27: We will comment on the large platform transparency reports.

In general, these are driven by statics on content taken down. We believe this is not an especially helpful metric as it tends to incentivise content removals, and does not provide a holistic picture of the platform actions. We also note that some platforms have reported relatively low numbers of appeals. This may look like the platform is doing a good job. In fact, it is more likely to indicate the opposite. On some platforms, it is not made clear to users how to appeal, or they cannot find the relevant page with the details on how to do so. This was our experience with Facebook in research done in 2019-2020 as discussed in the paper Algorithms Patrolling Content: Where's the Harm? . Users who don't know how to make an appeal, or who get no response, are less likely to file appeals.

With regard to shadow bans (limiting recommendation- see our response to Question 14), a metric based on taking down content would be useless because no content is taken down, although its visibility is suppressed. Ofcom should develop an alternative metric to measure the volume and efficacy of shadow bans.

Ofcom should also develop metrics to assess the quality of the reports about harmful content. The experience from copyright enforcement is that these reports are not always accurate. It would be harmful to users, if these reports were incorrect and platforms were being asked to act on the basis of them. For example, Ofcom could develop a metric to measure the volume of content that is falsely flagged as illegal or harmful.

Other

Q28. Other than those in this document, are you aware of other measures available for mitigating risk and harm from illegal content?

ORG response to Q28: We are aware of other measures that are being suggested for mitigating risk and harm from illegal content. We have concerns about them, as set out here.

Upload filter: We would draw your attention to our concerns expressed in response to Question 11. These systems impact directly on the right to freedom of expression by implementing a form of prior censorship. The harm to users of their implementation would need to be fully assessed by Ofcom before there could be any justification for mandating them. We would urge Ofcom not to mandate upload filters when drafting their Code of Practice.

Age assurance technology: We also draw your attention to our concerns expressed around the use of biometric surveillance and the gathering of biometric data, in response to Question 18. We make the opposite recommendation and suggest that Ofcom do not mandate the deployment of these systems. Currently, there are no standards for these systems, and there impact on the user experience has not been clearly established. There are no governance bodies in place to ensure that the vendors and service providers operate to human rights standards.

Monitoring and scanning of private messaging services: This is a very worrying issue as we are looking at a disproportionate surveillance of the entire population and yet the Bill is silent on the exact nature of the technical implementation and the functionality, and lacks a precise definition of what it is to do. The lack of definition is contrary to human rights standards, which is deeply concerning for a mandate that will interfere with free speech and privacy rights on a mass scale. The inherent danger is that the so-called "accredited technology" that could be mandated by Ofcom under Section 104 of the Bill, would compromise encrypted services, creating back doors and other opportunities for bad actors to interfere with the integrity of the system and the rights of the users. Far from suggesting that it be deployed more widely, we call on Ofcom to avoid deploying it at all.

As we understand it, the efficacy of these systems is highly questionable. To identify images, it uses perceptual hashing, a technique that research has shown to be error prone. The quality of the database is difficult to audit. The alternative is to employ AI systems using text-based classifiers, which have high error rates and are dependent on having sound data to train them on.

The likely technology to be implemented would be client-side scanning. This would require scanning software on every device, with the database of hashes. Aside from the repugnance that many people would feel towards this intrusion into their personal space, this technology is a form of mass surveillance and vastly disproportionate to the policy aims.

---ENDS - -