Open Rights Group's Response to Ofcom's Illegal Harms Consultation

I. Introduction

Open Rights Group (ORG) is the UK's largest digital rights campaigning organisation, working to protect people's right to privacy and free speech online. We have over 40,000 supporters across the UK and active member chapters in ten cities. Our work includes policy research and analysis, legal challenges, and public campaigning, all in the defence and promotion of digital rights.

Open Rights Group has campaigned on online free expression issues in the UK since its inception in 2005. We have publicly responded to both the Online Harms White paper and the Online Safety Act, while it made its way through the Houses of Parliament as a Bill.

We have submitted numerous policy briefings to parliamentarians about our concerns around the Act's impacts on online privacy, security, and free speech. In June of last year, we coordinated a letter that was signed by over 80 civil society organisations, academics and cyber experts from 23 countries urging the UK government to protect encrypted messaging.

Now, as the Bill has become the Online Safety Act, we welcome the opportunity to respond to Ofcom's consultations for how the proposals in the Act should be implemented. While we strongly welcome efforts to combat disinformation, hate speech, and illegal online content, and recognize the serious impacts that the spread of this content has had, there are still numerous pressing challenges and concerns in both the Act itself and Ofcom's proposed guidance.

To start, many civil society organisations who hope to be meaningfully involved in shaping this guidance face serious time and resource constraints that other lobbying groups and corporate organisations do not. The recent guidance put out by Ofcom for the Illegal Harms consultation was extensive, around 1,700 pages. While we appreciate the thought and effort that has gone into the guidance, and it demonstrates how seriously Ofcom is taking this task, the capacity and financial constraints on civil society are quite large. We hope to see Ofcom do more to meaningfully engage with civil society in the future and to recognize the unequal playing field for smaller, nonprofit organisations. Not only will this give Ofcom a better understanding of the range of viewpoints on their guidance, nonprofits are typically more likely to represent the views of communities who will be most impacted by changing regulations.

Our response, set out below, will cover our main concerns around free expression and privacy. We've chosen to respond in a format other than the template document provided by Ofcom, but we signpost the applicable sections of Ofcom's guidance that our response relates to.

II. Free Expression and Due Process [Volumes 1, 2, 3 & 4]

A. *Freedom of expression concerns and recommendations for content policies*

The Online Safety Act casts a wide net around content that must be removed and is likely to result in increased amounts of lawful content being taken down from the Internet. In July 2023, a legal opinion found that there were, "real and significant issues" regarding the lawfulness of a clause in the then Online Safety Bill, that appeared to require social media platforms to proactively screen their users' content and prevent them from seeing anything deemed illegal. The opinion found that there is "likely to be significant interference with freedom of expression that is unforeseeable and which is thus not prescribed by law"

It is important to ORG that any guidance from Ofcom meaningfully ensures that companies will consider their human rights responsibilities around freedom of expression and non-discrimination. When Ofcom is assessing "risk of harm" towards users, it should be assessing both positive and negative risk of harm (i.e. risk from harmful content *and* risk of a chilling effect or free speech being stifled).

While freedom of expression is acknowledged throughout the guidance, there is a lack of clear policies or processes in place to ensure that freedom of expression is prioritised. Companies are asked to balance the accuracy of content removals with the swiftness of content takedowns, without meaningful guidance on how this balance should be achieved and with incentives and penalties heavily leaning towards speed. We welcome guidance recommending proper resources and training for content moderation teams. However, even with proper training, most moderators are not lawyers, and moderation decisions are complex, difficult calls. We would like further information about how Ofcom plans to encourage services to protect free expression.

Open Rights Group recommends that the following ideas are incorporated into the guidance as *best practice recommendations*:

Already vulnerable or marginalised groups, like activists, racialised or queer communities, and people posting in non-Western languages experience the highest rates of wrongful content takedowns and are likely to be impacted by the increased amounts of content removed under this act. Companies should ensure that content moderation policies, and moderators themselves, have a clear and extensive understanding of the language, culture, and political and social content of the posts they are moderating. Furthermore, companies must ensure that users have access to rules, policies, and complaints processes in their chosen language.

ORG urges Ofcom to make it clear throughout its guidance that companies must ensure human rights and due process considerations are accounted for through all stages of the

moderation process. Companies must also be transparent about how they are incorporating free expression and non-discrimination concerns into these considerations.

Additionally, ORG recommends that many of the protections applicable to Category 1 services be extended *as best practice recommendations* more widely by the Guidance. These protections should be applied, at minimum, to all highly protected speech, no matter the service or service provider.

Highly protected speech could be defined as the following, as set out by the UN Human Rights Council in resolution 12/16:

- Discussion of Government policies
- Political debate
- Reporting on human rights
- Government activities and corruption in Government
- Engaging in election campaigns
- Peaceful demonstrations or political activities, including for peace or democracy
- Expression of opinion and dissent
- Religion or belief, including by persons belonging to minorities or vulnerable groups.

Currently, only News Publisher Content get meaningful protection, which is:

- Advance notice of intended action
- Reasons for intended action (and how FOE is served)
- Time to make representations on intended action.

Or—at least *ex-post action* (s18 (6-7)):

- Notice of action taken
- Justification for the lack of advance notice
- Time to request reversal

And (s19):

- Dedicated expedited complaints
- Swift decisions
- Swift reversal actions

Additionally, the Content of Democratic Importance protection for diversity of political opinions should be recommended for all and importantly for all speakers.

B. Recommendations for complaints processes

We welcome the guidance around clear and easily understandable content policies and complaints for all user to user and search services. In particular, the guidance that all U2U and search services must “Have an easy to find, easy to access and easy to use complaints system. . . and information and processes to be accessible and comprehensible, including having regard to users with particular accessibility needs such as children (if children use the

service) and those with disabilities." Additionally, we welcome that services will be required to "acknowledge receipt of each relevant complaint with indicative timeframes for deciding the complaint." We echo that all appeals processes should be transparent, clear, easy to access, timely, and involve human review. Users should be notified when their content is removed or account is suspended and given clear reasons why and instructions on how to appeal.

ORG requests more clarity and specificity around the requirements for appeals processes for people who believe their content has been wrongly removed. Specifically, ORG would like to encourage and understand whether the guidance will require human review of appeals by people who were not involved in the initial decision.

C. Incentives or penalties for accuracy

Appeals are a necessary safeguard, but they put the burden on users to take action and are not utilised in the majority of cases; ORG would like to see more provisions around incentivizing companies to prioritise accuracy of their decisions. In the current guidance, companies are asked to balance the accuracy of content removals with the swiftness of content takedowns, without meaningful guidance on how this balance should be achieved and with incentives and penalties heavily leaning towards speed. Without a counter balance here and more clarity on minimum thresholds for accuracy, companies will overcensor to avoid penalties and significant amounts of lawful speech and expression will be removed from the Internet.

Open Rights Group recommends that Ofcom should implement enforcement provisions to encourage a prioritisation of accuracy. For example, Ofcom could implement a mechanism to penalise companies who make repeated and significant mistakes that impede freedom of expression, or create a way for users to seek financial redress if their content is wrongly removed or their account mistakenly closed.

D. Government transparency and accountability

As states get more involved in regulating harmful speech, the possibility and opportunity for governments to exploit or manipulate companies' content moderation systems to censor unwanted speech (for example, political opponents or social movements) increases. Ofcom's guidance should assure the public that both itself and other government bodies commit to being transparent about their role in content removal and restriction and allow companies to publish data detailing how the Act and other government requests have affected content and user removals.

III. Privacy [Volumes 2, 3, & 4; Annex 9]

A. *Encryption*

In Volume 2, Ofcom clearly sets out encryption as a high risk for online. While Ofcom also caveats that "the functionalities . . . are not inherently bad and have important benefits" and play "an important role in safeguarding privacy online," it appears that, in Ofcom's view, these benefits are not enough to prevent encrypted services from coming under the scope of

irreconcilable moderation obligations. Open Rights Group urges Ofcom to consider the widespread impacts that weakening encryption could have if end-to-end encrypted services are required to comply with content moderation provisions that are impossible to reconcile with the functionality of their services. As in our June 2023 letter that was signed by over 80 civil society organisations, academics and cyber experts from 23 countries, we urge Ofcom to protect encrypted messaging.

In February 2024, in the case of Podchasov v. Russia, the European Court of Human Rights (ECHR) clarified that governments should not simply require that encryption be removed or limited in order to target criminals and thereby compromise everyone's privacy. The Court ruled that doing so is not proportionate.

Several key messaging services, including WhatsApp, Signal, and Element have said they would remove their services from the UK if encryption is impacted by the Online Safety Act. If encryption is weakened or these services are lost, online communication will be made insecure for everyone in the UK, as evidenced by cyber-security experts worldwide. In the guidance, Ofcom critically highlights the important role that encryption plays for members of the LGBTQ+ community who wish to safely discuss or explore their sexuality or gender. In addition to the LGBTQ+ community, many people in the UK and around the world rely on safe and secure messaging every day, including young people, activists, doctors, lawyers, journalists, victims of domestic abuse, and women seeking abortions in countries with restricted healthcare rights.

Ofcom claims that "the role of the new online safety regulations is not to restrict or prohibit the use of such functionalities, but rather to get services to put in place safeguards. . . managing the risks appropriately." If that is the case, **we believe Ofcom should publish regulations that make clear that there is no available technology that can allow for scanning of user data to co-exist with strong encryption and privacy.** ORG encourages Ofcom to guide encrypted messaging services towards other methods of improving user safety, such as sign posting users towards help services or device-level safety options.

Additionally, we would like clarity or further guidance on the following areas as they relate to encryption:

- Ofcom's proposals include a requirement to track evidence of new kinds of illegal content on the service and unusual increases in particular kinds of illegal content. We would like clarification on whether 'evidence' relates to encrypted content and if there are expectations of monitoring private conversations. [Volume 3]
- Services are required to have systems or processes in place to swiftly take down illegal content of which it is aware and non-priority illegal content where there is evidence of it. We would like clarity around these proposals and recommend that the guidance explicitly takes into account the limitations that encrypted services have and that content moderation cannot take place in private spaces. [Volume 4]
- In Annex 9, Ofcom sets out guidance on whether communication is public or private. We recommend that encrypted messaging be considered private communication.

B. Age verification

Open Rights Group is concerned about Ofcom's plans to implement age verification requirements for services. Age verification poses significant privacy risks for users, and there is no privacy-protective age estimation or verification process currently in existence that functions accurately for all users. This opinion has been backed by several governments in recent years. In September 2022, France's National Commission on Informatics and Liberty (CNIL) published a detailed analysis of current age verification and assurance methods. It found that no method has the following three important elements: "sufficiently reliable verification, complete coverage of the population, and respect for the protection of individuals' data and privacy and their security." Australia's government decided similarly in August 2023, stating "It is clear. . . at present, each type of age verification or age assurance technology comes with its own privacy, security, effectiveness or implementation issues." In short, every age verification method has significant flaws.

Age verification systems will collect data, particularly biometric data. This carries significant privacy risks, and there is little clarity in the Act or guidance about how websites will be expected to mitigate these risks. Platforms like Facebook and TikTok, and even community-based sites like Wikipedia, will have to choose between conducting age checks on all users – a potentially expensive, and privacy-invasive process – or sanitising their entire sites. This will result in an enormous shift in the availability of information online, and pose a serious threat to the privacy of UK internet users. It will make it much more difficult for all users to access content privately and anonymously, and it will make many of the most popular websites and platforms liable if they do not block, or heavily filter, content for anyone who does not verify their age.

Whilst those advocating for age-verification are well-intentioned, the result will be a disproportionate interference with children's and adult's right to access information and their right to privacy.

We recommend that Ofcom's guidance should include provisions specifying that any age assurance or age verification systems should be effective at correctly identifying the age or age-range of users and strongly safeguard individuals' data and privacy and their security.

We encourage Ofcom to work with the Information Commissioner's Office to set out strong, clear guidelines for data protection requirements in these systems. ORG is concerned that with proposed changes to the UK's data protection regime through the Data Protection and Digital Information Bill, people's biometric data will be particularly at risk in coming years.

C. Competition and Interoperability

Ofcom should also consider the importance of interoperability between platforms in its guidance. Safety will improve if users can leave platforms for others that align best with the type of moderation styles, privacy approaches and user features that work for them.

Additionally, if people are required to participate in age-verification systems to access online information or services, there should be some requirement for competition among the systems a customer can choose to verify their age. For example, if an individual trusts Apple or Yoti's age verification system more (for being more privacy protecting or accurate, etc) then sites should be encouraged to accept those methods of verification. People should not be forced into badly implemented age-verification systems to access services. Introducing consumer choice would enable privacy minded consumers to opt for the platform that has the best track record on data protection, security and privacy.

IV. **Conclusion**

As Ofcom continues to develop its guidance on the Online Safety Act, it is essential to consider the broader implications for freedom of expression, privacy, and democratic principles. Failure to do so could not only undermine fundamental rights and freedoms within the UK, but also set a dangerous precedent for online censorship globally if repressive regimes take the Act and Ofcom's guidance as a licence to further censor and penalise legitimate speech.

Open Rights Group welcomes further engagement with Ofcom on the issues discussed above or any related free expression and privacy topics within the Online Safety Act.