Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression

Ref.: OL GBR 5/2022
(Please use this reference in your reply)

14 March 2022

Excellency,

I have the honour to address you in my capacity as Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, pursuant to Human Rights Council resolution 43/4.

The purpose of this letter is to share with you my concerns regarding the **Online Safety Bill** (the "Bill), which was published on 12 May 2021 and is scheduled to be put to a vote in Parliament in 2022. In February 2022, your Excellency's Government announced that three new offences will be introduced through the Bill, based on the UK Law Commission's 'Modernising Communications Offences' report published in July 2021,[1] and that additional 'priority offences' will be included in the Bill.[2]

I wish to acknowledge at the outset the very real concerns regarding the societal impact of social media platforms, including the online safety of children and other vulnerable groups, that have motivated this legislative proposal. I appreciate the commitment of your Excellency's Government to promote transparency and accountability with respect to social media intermediaries. Given the potential impact on a wide range of human rights, I am also pleased to note that the consultations involved a broad diversity of relevant stakeholders.

Having said that, I believe the proposed Bill, as currently drafted, contains some key provisions that could undermine its overall objective as well as international human rights principles. I am concerned that the Bill promotes and incentivises the removal of so-called 'legal but harmful' content on overly broad grounds. I also note that the responsibility for the removal of such content as well as of illegal content is assigned to social media companies without judicial oversight, thereby preventing access to remedy in case of human rights violations as a result of the enforcement of such provisions. I believe that certain provisions analysed below, as well as the enhanced powers of the regulatory body to impose hefty fines, could create a chilling effect on freedom of expression on social media platforms. I therefore encourage review and reconsideration of the Bill.

Before turning to my concerns in further detail below, I would like to underline that article 19(2) of the International Covenant on Civil and Political Rights (the 'Covenant') provides the right "to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice." As you are aware, restrictions to this right are permissible under international law only if they are provided by law and are necessary and proportionate to achieve one of the legitimate aims expressly enumerated in article 19(3) of the Covenant and article 10(2) of the European

[1] DCMS, Update on the Law Commission's Review of Modernising Communications Offences, Written Ministerial Statement (HCWS 590), 4 February 2022.
[2] DCMS, Online Safety Update, Written Ministerial Statement (HCWS 593), 7 February 2022.

Convention on Human Rights ('European Convention').

Duty of care to remove lawful but 'harmful' content

Under the Bill, regulated providers likely to be accessed by children and certain "Category 1 services" to be identified in future by secondary legislation will be required to carry out risk assessments to identify the presence of harmful content on their service, and

- for services likely to be accessed by children
    - take proportionate steps to mitigate and manage the impact of harmful content on children
    - ensure that the terms of service specify how children will be protected from harmful content which they may encounter
- for Category 1 services, ensure that the terms of service specify how the service will address harmful content.

Regulated providers will need to comply with the duties set out above for content that they reasonably identify as having a 'material risk' of a 'significant adverse physical or psychological impact' on a child or adult of 'ordinary sensibilities' (**Clauses 45(3) and 46(3)**). The regulator must take into account, in particular, (i) the number of users that may be assumed to have been exposed to the content, (ii) how easily, quickly and widely the content may be disseminated by the service provider (**Clauses 45(5) and 46(5)**).

The Bill provides very little detail regarding the interpretation or application of these obligations. These will require further elaboration by the regulator (UK Office of Communications - Ofcom). According to the explanatory notes to the Bill, harmful content could range from online bullying and abuse to advocacy of self-harm, to spreading disinformation and misinformation.

I am concerned that the Bill's creation of a duty of care on regulated providers based on vague and broad formulation of the provisions create risks to freedom of expression. It is contrary to the notion of legality in article 19(3) of the Covenant which has been interpreted as requiring a restriction to be "formulated with sufficient precision to enable an individual to regulate his or her conduct accordingly" (Human Rights Committee, General Comment 34). Furthermore, any restriction to freedom of expression must "not put in jeopardy the right itself" (ibid), and so should be narrowly and clearly defined.

The duty of care placed upon online providers to protect users of their services against legal but harmful content uses vague terms that are open to broad interpretation, such as "reasonably identify", "material risk", "significant adverse physical or psychological impact", "ordinary sensibilities", and so risks undue removal of content. In effect, platforms will be required to undertake a subjective assessment of the potential impact of content on an unknown individual. Without appropriate clarification of what is and is not 'legal but harmful' content, it may prove difficult for regulated service providers to properly carry out the balancing act required under the legislation.

Even with additional guidance, regulated providers will be required to make difficult judgement calls regarding content removal that they may be ill-equipped to undertake. That creates an acute risk of overbroad removal of content, particularly when an evaluation of context or nuance may be required, as in the case of satire, commentary on matters of public interest where there is heightened debate or controversy, or content seeking to raise awareness of important societal issues like suicide, self-harm or eating disorders. A sufficiently precise definition is essential to prevent the possibility that legitimate expression is taken down in violation of international human rights law standards.

I note that the Bill places a duty on all regulated providers when deciding on and implementing safety policies and procedures, to have regard to the importance of (a) protecting the right to freedom of expression within the law, and (b) protecting users from unwarranted infringements of privacy. However, that might result simply in a box-ticking compliance exercise by providers as the Bill does not adequately guide service providers on how to resolve the inherent tension between removal of legal content while also upholding freedom of expression.

I welcome the provision in the Bill to carve out exceptions for content of 'journalistic' or 'democratic importance'. However, again, without clear definitions it may be difficult for regulated providers to apply those provisions.

Duty to remove illegal content

Regulated providers will be required to remove content generated and shared by other users that is unlawful under UK law, or that the provider has "reasonable grounds to believe" is illegal. In the Bill, illegal content is divided between four different categories, namely 'terrorist' type of content, child sex abuse material, priority illegal content as specified by the Secretary of State in regulations and other illegal content where the actual or intended victim is an individual (**Clause 41**). I note that your Excellency's Government has indicated that 'priority illegal content' will include *inter alia* encouraging or assisting suicide; revenge and extreme pornography; incitement to and threats of violence; and hate crime.[3]

Regulated providers will be left to police the limits of acceptable speech based on their own interpretation of whether content is likely to meet the threshold for criminal prosecution. I am concerned that this obligation delegates to private companies a responsibility that should be exercised by law enforcement, particularly for offences where the boundary between offensive but legal and illegal conduct may be difficult to discern, such as hate crime.

Removal of disinformation or misinformation

I am concerned that the expression of some ideas categorised as false information may be improperly restricted in violation of international standards and national law. The right to freedom of expression applies to all kinds of information and ideas, including those that may shock, offend or disturb,[4] and irrespective of the truth or falsehood of the content.[5] I would like to draw attention to the Joint

[3] DCMS, Online Safety Update, Written Ministerial Statement (HCWS 593), 7 February 2022.

[4] Human Rights Committee, General Comment No. 34 (2011), para. 11. See also European Court of Human Rights, Handyside v. the United Kingdom, application No. 5493/72, judgement, 7 December 1976, para. 49.

[5] General Comment No. 34 and European Court of Human Rights, Salov v. Ukraine, application No. 65518/01, judgement, 6 September 2005, para. 113.

Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda[6] published by my mandate and regional experts on freedom of expression, which delineates the applicable human rights standards in this context. The Joint Declaration notes that general prohibitions on the dissemination of information based on vague and ambiguous ideas, including "false news," are incompatible with international standards for restrictions on freedom of expression, and should be abolished".[7] Disinformation may be restricted in certain circumstances, such as where it causes harm to individual reputation and privacy, or incites violence, discrimination of hostility against identifiable groups in society. However, any measures to prevent the dissemination of disinformation must comply with the criteria set out in Article 19(3) of the Covenant, as well as under Article 10 of the European Convention.

Lack of judicial oversight

I note that the Bill foresees limited appeal of decisions made by Ofcom with respect to regulated providers (**Clause 104**), appeal of Ofcom notices (**Clause 105**) and the ability for designated entities to make super-complaints (**Clause 106**). However, the Bill does not foresee any external appeal mechanism for individual content moderation decisions. This means that for both illegal and "legal but harmful" content, regulated providers will make judgement calls regarding content removal without adequate judicial or independent oversight.

Requiring private companies to remove broad categories of content without a court order would run contrary to international standards on freedom of expression. Content removals must be undertaken "pursuant to an order by an independent and impartial judicial authority, and in accordance with due process and standards of legality, necessity and legitimacy".[8] States should not require the private sector to take steps that unnecessarily or disproportionately interfere with freedom of expression, whether through laws, policies or extra-legal means, nor should they delegate censorship measures to private entities.[9]

In addition, I am concerned that **Clause 9(3)** appears to establish a general monitoring obligation that could lead to the monitoring and filtering of user-generated content at the point of upload. This form of restriction would enable the blocking of content without any form of due process before it is even published, with the risk that providers may over-regulate expression to avoid sanctions.

I note that the Bill would require regulated providers to operate an internal user complaint mechanism that "provides for appropriate action to be taken" and which is "easy to access", "easy to use (including by children)" and "transparent" (**Clause 15**). However, as pointed out by the Joint Committee on the Draft Online Safety Bill, the Bill does not currently set any minimum quality standards, nor does it identify how Ofcom should assess requirements such as ease of access and use, or transparency.[10] In addition, if a regulated service provider implements a system that leads to systematic under-flagging or over-flagging of content, it is unclear how this will be treated by Ofcom from a compliance perspective when it comes to

6 Joint Declaration
7 See also A/HRC/47/25.
8 A/HRC/38/35, para. 66
9 A/HRC/32/38
10 Joint Committee on the Draft Online Safety Bill, Draft Online Safety Bill, Report of Session 2021-22, 14 December 2021 – HL Paper 129 – HC 609, paras. 440-446.

enforcement of the Bill.

Furthermore, the substantial cost of compliance with the monitoring and risk assessment obligations could pose a substantial burden on smaller service providers, thereby contributing to strengthening the role of dominant online platforms. The overall result could be to reduce online information diversity and pluralism, and/or prevent its emergence in the future.

Limits on the right to privacy

The rights to privacy and to freedom of expression are closely linked in the digital age -- and by extension, encryption and online anonymity are protected under international standards because of the critical role they can play in securing those rights.[11] By enabling users to protect their data and communications from unauthorised access, encryption and anonymity give individuals, including journalists, those persecuted because of their sexual orientation or gender identity, human rights defenders and others, a zone of privacy in which to hold, share and develop opinions and exercise their freedom of expression without arbitrary and unlawful interference or attacks or social pressure.[12] Accordingly, my mandate has consistently urged States to refrain from blanket restrictions on encryption and anonymity.[13]

I am concerned that the inclusion of direct private messaging within the scope of the Bill could impact negatively on encryption, security and privacy. I have similar concerns regarding Ofcom's ability to compel a service to use technology to detect child sexual exploitation and abuse (CSEA) and terrorism content on private and public channels and CSEA content on private communication channels and to "swiftly take down that content" (**Clause 64(4)(b)**), as well as the related issuance of technology warning notices (**Clause 63**). Any measure that has the effect of weakening, bypassing or removing encryption puts everyone at risk by creating points of access that could be exploited by third parties.

Furthermore, I note that the Bill extends protection to users from "unwarranted infringements of privacy" (**Clause 12**), without providing any definition of 'unwarranted', thus leaving an open question about the circumstances under which a privacy infringement will be permitted.

Oversight and enforcement

With respect to the oversight and enforcement mechanisms laid out in the Bill, I note that Ofcom, as the foreseen regulator under the Bill, will have the mandate to register all entities that fall within the scope of the Bill. In addition, Ofcom will be empowered to exempt news publishers from its statutory oversight if they meet the Bill's criteria of a 'recognised news publisher', such that they benefit from journalistic protections under the Bill. This could have the effect of indirectly introducing elements of a statutory licensing scheme for the media, notwithstanding the UK's strong tradition of press self-regulation as enshrined in the Royal Charter on self-regulation of the press.

---

11 A/HRC/23/40.

12 A/HRC/29/32

13 See as well the Special Rapporteur on right to privacy's support of judicial decisions that reject breaking encryption as "disproportionate, privacy-intrusive measures" (A/HRC/31/64, para. 58).

I also note with concern the severity of the envisaged penalties which could have a chilling effect on freedom of expression. The Bill:

a. authorises Ofcom to issue a fine of £18 million or 10% of annual worldwide turnover (whichever is greater); and

b. contains a deferred mechanism to impose criminal sanctions on senior management of regulated providers where they have failed to take reasonable steps to prevent offences from being committed.

At the same time, the Secretary of State is granted extensive powers, including to designate priority illegal content **(Clauses 41 and 47**), and to direct Ofcom to modify its guidance (i.e. Codes of Practice) to reflect government policy (**Clause 33**), which may undermine Ofcom's independence.[14]

Under international standards, when deciding whether to apply sanctions authorities must take care to meet the standards of necessity and proportionality provided under article 19(3) of the Covenant.[15]. As my mandate has previously noted, States should "refrain from imposing disproportionate sanctions, whether heavy fines or imprisonment, on Internet intermediaries, given their significant chilling effect on freedom of expression", and refrain from adopting models of regulation where government agencies, rather than judicial authorities become the arbiters of lawful expression."[16] In this case, the proposed hefty fines, criminal liability, and other measures introduced in the Bill raise concerns whether they would meet that test. They also heighten the risk that providers are likely to take down content without fully investigating whether it is harmful so as to avoid incurring potential liability.

Concluding observations

The State has a legitimate interest and responsibility to promote online safety and to protect individuals against terrorism, child pornography, and speech that constitutes incitement to discrimination, hostility or violence. The questions that arise relate to the way in which the Bill seeks to achieve these legitimate objectives, in particular the responsibilities it places upon private companies to regulate the exercise of freedom of expression, and the inclusion of measures aimed at so-called legal but harmful content in overly broad terms is lawful under international human rights law.

As it is my responsibility, under the mandate provided to me by the Human Rights Council, to seek to clarify all cases brought to my attention, I would be grateful for your observations on the following matters:

1. Please provide any additional information and/or comment(s) you may have on the above-mentioned matters.

2. Please provide your observations on how the Online Safety Bill is consistent with your Excellency's Government's obligations under international human rights law, especially the requirements of Article 19 of the Covenant.

[14] Joint Committee on the Draft Online Safety Bill, Corrected oral evidence: Consideration of government's draft Online Safety Bill (23 September 2021) 10 am.

[15] See UN Human Rights Committee, Korneenko et al. v. Belarus (Communication no. 1274/2004, 31 October 2006), paras. 7.6-7.7.

[16] A/HRC/38/35.

This communication, as a comment on pending or recently adopted legislation, regulations or policies, and any response received from your Excellency's Government will be made public via the communications reporting website after 48 hours. They will also subsequently be made available in the usual report to be presented to the Human Rights Council.

I stand ready to provide Your Excellency's Government with any technical advice it may require in ensuring that the Bill is fully compliant with international human rights obligations,

Please accept, Excellency, the assurances of my highest consideration.

Irene Khan
Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression