To: Baroness Tina Stowell, Chair, Lords Communications and Digital Committee From: Prof Lorna Woods OBE, Professor of Internet Law, University of Essex

13th December 2022

BY EMAIL

Dear Baroness Stowell

Thank you for the opportunity to give evidence to your Committee last week on the Online Safety Bill. The Committee's clerk has asked me for further information required by the Committee in writing and I set this out, along with some further thoughts on one or two other topics, below. Do please get in touch with me again if there is anything here that is unclear.

Removal of clause 12 – harms to adults risk assessment

The decision to drop the broadly based adult risk assessment required in old clause 12 is regrettable. It means that companies would no longer be required to consider what impact their features and functionalities have on the creation, promotion and prevalence of content harmful to adults and we have seen (eg in the Molly Russell case, information from whistleblower about Instagram's impact on body image) that personalisation and design for user engagement can have significant impacts. The corollary of this is that the range of possible solutions becomes restricted. Rather than having a choice of interventions right the way across the distribution chain (from account creation, content creation and upload, discovery and recommendation – as well as user empowerment tools and company moderation), the focus is on *ex post* interventions.

Yet, as the UN Special Rapporteur for Freedom of Expression noted, other interventions are more proportionate from a freedom of expression perspective. In a public comment on a Facebook Oversight Board decision, Irene Khan, taking a similar stance to her predecessor – David Kaye, wrote: "*International law requires use of the least restrictive measure available to*

confront the problemsof 'hate speech'... It may be appropriate to use measures such as downranking, demonetizing, friction, warnings, geoblocking and countermessaging, and the criteria for application of these measures should also be transparently disclosed to allow users to govern their behaviour accordingly and understand when these measures have been applied."[1]

Moreover, a lack of a risk assessment means that there will no assessment of the types of content or features likely to be harmful. This would have provided an exceptional insight into trends and a forward look at harm reduction problems on the horizon, especially those that could indicate a path towards criminal activity. It would have provided in particular a useful insight into radicalisation and (non-terror) extremism neither of which are criminal offences.

It would be useful, when the Committee receives evidence from the Government, to find out from them why the risk assessment was removed. There was no discussion in the media material or WMS, nor evidence justifying such a change. To some extent, retaining a risk assessment in relation to content harmful to adults would be practical from the service provider's perspective, as service providers will be carrying out risk assessments for illegal content and, most likely, children's duties in any event. Alternatively, we are of the view that the benefit of such a risk assessment exercise could be gained by changing the emphasis and requiring instead OFCOM to publish an annual Threat Assessment rather than companies themselves carrying it out. This would need to be informed by OFCOM's use of information- gathering powers, which may need to be extended.

Of course, this still leaves the point that there may be a difference between what the companies choose to take action about and the range of risks on the platform. We suggest that some form of transparency obligation should be in place around the risks – whether these are assessed by the service provider or OFCOM - so as to allow users to make an informed choice as to whether they want to use the service or not. Under the current proposals, users will not have access to that vital information. The obligation to publish a summary of the risk assessment from original clause 13 could be re-instated were the service providers to be subject to a risk assessment obligation, or the obligation to publish OFCOM's breach notifications could be extended to include a statement as to the outcome of the Threat Assessment if introduced.

Note that the obligation to publish breach notices will not suffice in this regard as it could be easily gamed.

With regard to the duties that form the third part of the "Triple Shield", in lieu of the old "harmful to adult" duties – the duty not to remove content and the user empowerment duties – the lack of any risk assessment provision means they are outliers in relation to the other two duties in the Bill, notably the duties with regard to illegal content and to protecting children.

In amendment 13, the Government proposes probability of occurrence of a type of harm as a factor in assessing the proportionality of user protection. In another (amendment 9), the

---

[1] https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/Legislation/Case_2021_009-FB-UA.pdf

Government says that measures have to be 'effective', providing this detail in the EN that it is strengthening "the duty in this clause by requiring that the systems or processes used to deal with the kinds of content described in subsections (8B) to (8D) (see Amendment 15) should be designed to effectively increase users' control over such content."

This suggests that a company deciding whether or not to offer a tool would have had to carry out a risk assessment, especially as, in assessing whether the user empowerment duties had been met, OFCOM would be likely to investigate what grounds the provider had for determining that particular tools were thought to be effective. We think that there perhaps should be consequential amendments (not yet made) to OFCOM's risk profiles to ensure that this aspect can be included.

If so, it seems that the requirement for risk assessment has moved from being explicit and public to implicit and private between the companies and the regulator, insofar as it exists at all.

Secretary of State's powers

Our concerns about the Secretary of State's powers remain unchanged and we set out some of our analysis regarding the amendments we would wish to see in the Lords, below. While the amendments to the "harms to adults" clauses remove one set of circumstances in which the Secretary of State may exercise powers, the changes do not have any bearing on the suite of powers that remain in the Bill beyond this. A key concern for us has been the potential threat to OFCOM's independence - when independence of the regulator is seen as crucial to maintaining a diverse and pluralistic communications environment. In this we draw a distinction between the setting of the framework within which OFCOM must operate and identifying the objectives of the regime – which are rightly set by Parliament (and not the Executive) – and the implementation of that regime, which should be for the regulator, making decisions based on evidence. For these reasons, while we maintain our concerns about the Secretary of State's powers in relation to the codes of practice, our concerns are not limited to Clause 41.

Clause 41 (previously clause 40): powers to direct Ofcom for reasons of public policy

Clause 41(1)(a) gives the Secretary of State power to direct OFCOM to make specific changes to some of OFCOM's draft codes of practice if s/he '*believes that…modifications are required…for*

reasons of public policy'. The codes are the fulcrum of the regulatory regime and this is a significant interference in OFCOM's independence. Jeremy Wright MP and the DCMS Select Committee put forward amendments at Commons Report that both address this issue.

OFCOM itself has noted that the 'reasons of public policy' power to direct might weaken the regime. If OFCOM has undertaken a logical process, rooted in evidence to arrive at a draft code then it is hard to see how a 'reasons of public policy' based direction is not irrational for the purposes of judicial review. This then creates a vulnerability to legal challenge.
Our position is that this clause should be removed.

We also have concerns that the same clause gives the Secretary of State powers to direct OFCOM on national security or public safety grounds the Terrorism and CSEA codes of practice. The government has not demonstrated why it needs a power to direct. In the broadcasting regime, there are no equivalent powers and the Secretary of State was able to resolve the case of Russia Today on national security grounds with public correspondence between the Secretary of State and OFCOM.

The Secretary of State can use the Cl 41 powers to direct OFCOM continuously in a form of ping pong before laying a code – this signals a willingness of Government to wear OFCOM down and impose the Government's view on the regulator. It seems to be preparing the ground for an irrational or highly disputed request.

We noted that, in the WMS of 7/7/22, the then Secretary of State signalled a potential Government concession in the Lords to address the concerns raised about this clause. The WMS states:

> *We recognise the concerns raised that the Bill allows too great a degree of executive control. These have focused in particular on the power for the Secretary of State to require Ofcom to modify a draft of a code of practice for reasons of public policy. We remain committed to ensuring that Ofcom maintains its regulatory independence, which is vital to the success of the framework. With this in mind, we have built a number of safeguards into the use of the Secretary of State's powers, to ensure they are consistent with our intention of having an independent regulator, and are only used in limited circumstances with appropriate scrutiny.*
>
> *We will make two substantive changes to this power: firstly, we will make it clear that this power would only be used 'in exceptional circumstances'; and secondly, we will replace the 'public policy' wording with a more clearly defined list of reasons for which the Secretary of State could issue a direction. This list will comprise national security, public safety, public health, the UK's international relations and obligations, economic policy and burden to business.*

It is our view that this does not go far enough to address the problems with this clause and the exercise of the Secretary of State's powers and should be resisted.

We would prefer to see Clause 41 replaced with the ability for the Secretary of State to write to OFCOM in public with observations on codes on limited grounds and for OFCOM to have regard to but not be bound by such letters. This is the conventional approach to UK government/regulator relationships. The Government's ability to write letters should not be infinite. Codes issued after such correspondence should be approved by the House using the affirmative procedure.

Clause 153 (previously clause 148) Secretary of State's Guidance

The Secretary of State takes powers in Cl 153 to issue detailed tactical guidance to OFCOM on the 'exercise of their functions', to which OFCOM should have regard. This includes enforcement of the regime. This is in addition to Cl 149 (previously 144) which allows the Secretary of State to make a statement of strategic priorities relating to online safety (and which is subject to some Parliamentary safeguards). The tactical guidance is incredibly broad and detailed in scope with no constraints. The guidance has no Parliamentary input into its drafting. Even qualified by 'have regard', Cl 153 is interference in how OFCOM carries out regulation. We believe this clause should also be deleted.

Tackling violence against women and girls

I mentioned to the Committee that I felt that, notwithstanding the new offences being added to the Bill that have a particular relevance to women and girls, there needed to be a more systemic requirement on companies to recognise the particular vulnerability of women and girls to online violence. VAWG as a series of discrete offences fails to recognise the environmental nature of the risk, and does not take into account how criminal and non-criminal abuse together make up this particularly risky environment. The heightened exposure of women and girls to harm needs special recognition in the risk management process. To this end, Carnegie UK has worked with a number of civil society organisations and academics to develop a code of practice that we believe should be included in the Bill and which was launched by Maria Miller MP and Baroness Nicky Morgan.[2]

Extremism: content that falls short of criminal threshold

I also mentioned to the Committee that I had concerns about the fact that the Bill does not cover extremism, most notably the type of content that can lead to radicalisation but that falls short of the criminal threshold. I have provided below the relevant extract from our written evidence to the Bill Committee, which we have submitted today, which sets this out:

[2] The code is available here: https://d1ssu070pg2v9i.cloudfront.net/pex/pex_carnegie2021/2022/05/24163713/VAWG-Code-of-Practice- 16.05.22-Final-1.pdf

> Hateful extremism is a long-standing omission from the Bill which hasn't been properly debated (Rehman Chishti MP raised it at Report[3]). There is increasing concern about extremism leading to violence and death which does not meet the definition for terrorism and might not fall within "abuse" as specified in the list for user empowerment tools. The internet and it seems user-to-user services play a central role in the radicalisation process. The OSB does not cover extremism. Sara Khan, the former Lead Commissioner for Countering Extremism, provided a definition of extremism for the Government in February 2021 but we are not aware of a formal response from the Government.[4]

We therefore support the amendments tabled by the Opposition to Government amendments 15 and 16, which relate to the list of content that is to be covered by the user empowerment duties:

> 15 (a): Line 21, at end insert— "(8E) Content is within this subsection if it— (a) incites hateful extremism, (b) provides false information about climate change, or (c) is harmful to health."
>
> And,
>
> 16 (a): Line 2, at end insert— "hateful extremism" means activity or materials directed at an out-group who are perceived as a threat to an in-group motivated by or intended to advance a political, religious or racial supremacist ideology— (a) to create a climate conducive to hate crime, terrorism or other violence, or (b) to attempt to erode or destroy the rights and freedoms protected by article 17 (Prohibition of abuse of rights) of Schedule 1 of the Human Rights Act 1998."

Risks of over-removal of content

Finally, if I may, I would like to return to one of the main topics of discussion at the Committee last week with regard to the Online Safety Bill and threats to free speech. It is the contention of many of those who oppose the Bill that in seeking to comply with the duties, regulated companies will err on the side of over-removal of content and as a result, take down much that is below the threshold of illegality, with an impact on the rights of individuals to free speech.

---

[3] "Terrorism is often linked to non-violent extremism, which feeds into violent extremism and terrorism. How does the Bill define extremism? Previous Governments failed to define it, although it is often linked to terrorism." https://hansard.parliament.uk/commons/2022-12-05/debates/E155684B-DEB0-43B4-BC76- BF53FEE8086A/OnlineSafetyBill#contribution-962913D8-A7F7-41C4-8A72-08AEC01DFC7E

[4] https://www.gov.uk/government/publications/operating-with-impunity-legal-review

It is worth emphasising here that the Bill is a civil regulatory regime, not a criminal one. I set out below some of the points I made in correspondence last year with the Chair of the Joint Human Rights Committee on a similar point, which have relevance here:

> "*the assessment of a regime and its compatibility with fundamental rights is different from an assessment of an administrative decision within a regime for compliance. While individual decisions can review all the facts in relation to that decision, the review of legislation setting up a regulatory regime operates at an abstract level. The European CourtofHumanRightshas notedthat:"aStatecan,consistentlywiththeConvention, adopt general measures which apply to pre-defined situations regardless of the individual facts of each case even if this might result in individual hard cases (Ždanoka v. Latvia [GC], Application no. 58278/00, §§ 112-115, ECHR 2006-IV). Contrary to the applicant's submission, a general measure is to be distinguished from a prior restraint imposed on an individual act of expression (Observer and Guardian v. the United Kingdom,26November 1991,§60,SeriesAno.216)".[ADIvUK,(Applicationno. 48876/08) (Grand Chamber), para 106]*
>
> "*Assessment in this context would seem to require looking at the underpinning legislative choices, the quality of review and the risk of abuse if a general measure were to be relaxed. It is clear in this that the State is not limited to tackling content that is bad enough to be criminalised. Moreover, a law may grant discretion, provided that – bearing in mind the objective of the legislation –there is sufficient clarity to safeguard the individual against arbitrary interference (eg Magyar Kétfarkú Kutya Párt v. Hungary [GC], Application no 201/17, para 94). Unless general prohibitions are in issue (as was the case in ADI v UK), the question would seem effectively to be that of whether the regime allows actors enough space to make properly balanced and proportionate decisions.*
>
> *"This is particularly the case in a regime, as in the Online Safety Bill, where the regulation does not necessarily impose direct content regulation, but rather requires the subjects of regulation (companies) to have in place systems to deal with content proportionately. The systems will be designed, not looking at a particular item of content, but rather at categories of content. Against such a background, it would seem that relevant questions to ask operators are not 'what did you take into account in this case?', but ratherto look at the service's processes, and the extent to which they have been designed to take into account relevant human rights (and not just freedom of expression). For example, do the terms and conditions deal with different types of speech in an appropriately granular way, or does the company rely on broad, vague categories? Is its decision-making in this regard transparent and predictable? How much investment is there in moderation and response to complaints, so how fast are decisions made and appeals processed? If a service employs automated systems (highly likely), what datasets have been used to train those systems? What are the tolerances for false positives and*

false negatives – and why are they set at this level (and should the level be the same for all types of speech)?

I hope that additional context is useful for the Committee's deliberations and that I have also provided all the extra information requested in response to the follow-up questions. Do please let me know if there is anything that is unclear or where you would like more detail.

Yours sincerely

Prof Lorna Woods OBE