Online Safety (recommitted clauses and schedules) Bill

Written evidence from Carnegie UK

December 2022

Introduction

1. This note sets out our view on the amendments that are under consideration by the Committee in this recommittal stage. This evidence has five sections:

   a) the new duties on terms or service (Section 1),
   b) the possible impact of those duties on service provider behaviour (Section 2)
   c) the User Empowerment Duty
   d) the need to reinstate some aspects of risk assessment (Section 4)
   e) changes to categorisation of companies (Section 5)

2. In assessing the Government's amendments, we have referred to the WMS published by Minister Paul Scully on 30 November[1] and the list of amendments, including those from the Government and opposition parties, as published on 9th December. [2]

3. As many of the contributors to the Report day 2 debate observed[3], the development of this Bill has been a long-running saga. We have published thousands of words on the subject, which can be found on our website[4], and spent many hours giving evidence in Committee hearings at various points of the Bill's passage.[5] We do not intend to rehearse all that again here but would be more than happy to provide the Bill Committee with more detail on any of our thinking during this short recommittal stage.

4. For new members of the Committee, we attach, at annex A, a short explainer of how the Bill works; while the focus of this recommittal stage is limited to a handful of new clauses, they cannot be scrutinised effectively without a basic understanding of the framework into which they are intended to fit.

5. We also attach at annex B our recent blog post on why the "harmful but legal" provisions in the Bill should, in our opinion, have remained in the Bill. Given that the Government has made clear that they do not intend to reverse this decision, the rest of the evidence focuses on the new amendments.

1 https://questions-statements.parliament.uk/written-statements/detail/2022-11-30/hcws403
2 https://publications.parliament.uk/pa/bills/cbill/58-03/0209/amend/onlinesafety_rm_pbc_1209.pdf
3 https://hansard.parliament.uk/Commons/2022-12-05/debates/E155684B-DEB0-43B4-BC76-BF53FEE8086A/OnlineSafetyBill
4 https://www.carnegieuktrust.org.uk/programmes/tackling-online-harm/
5 We would refer the Committee to the most recent appearance of Prof Lorna Woods at the Lords Communications and Digital Committee discussion on the Online Safety Bill on 6 December: https://parliamentlive.tv/event/index/79b3c3c5-e592-48e3-b75e-ff1ed982853b

The Government's amendments

6. The tabled amendments do the following:

- Remove the requirement for service operators to assess the risk of harm to adults occurring through their platform originally found in cl 12;
- Remove the harmful to adults' transparency obligations originally found in cl 13;
- Introduce duties around enforcement of Terms of Service (ToS);
- Remove the category of "content harmful to adults"6 and, consequently, "priority content that is harmful to adults" but introduce a list of content areas to which the user empowerment duties (cl 14) apply;
- Amend user empowerment duties (cl 14);
- Amend rules in relation to determination of Category 1 services;
- Leave illegal content duties as before, although the press release and the first Government WMS refer to these duties as one element of the triple shield for adults (note additional criminal offences will be introduced but these are not part of the amendment package); and
- Tighten drafting around age verification in relation to children's safety duties.

Section 1: Duties on terms of service

7. There are two amendments introducing new duties. The first is to:

Ensure that the largest or most risky online service providers design systems and process so that the service provider can't take down or restrict content (so that a person cannot see it without further action by the user) or ban users unless it is in its own ToS or constitutes breaking the law or the OSB regime. (NC3)

This duty is referred to as the '*Duty not to act against users except in accordance with terms of service*'.

The second is to improve the effectiveness of terms of service so that:

if the provider's ToS say that certain types of content are to be taken down, or restricted or that certain behaviours will lead to banning then providers must run systems and process to make sure that these things happen;
people are able to report breaches easily and that a complaints service delivers 'appropriate action' in response to complaints; and
the service allows certain types of complaints – including about the service provider itself. (Mainly NC4)

This 'effectiveness' duty is referred to as '*Further duties about terms of service*'.

There are dozens of consequential amendments flowing from these duties.

8. Both the duties are focused narrowly on banning, take down and restriction, rather than upstream design changes or softer tools that might be more effective and allow different types of users to coexist on the same service more easily. Nonetheless, the duties bear on the quality of systems

6 This is defined in clause 55 of the current version of the Bill as content "of a kind which presents a material risk of significant harm to an appreciable number of adults in the United Kingdom".

and processes that companies operate to achieve this effect in general, rather than a hard liability for one wrong decision.

9. Notably the systems and processes required by the new duties should be 'proportionate' to the 'size and capacity' of the service provider: the government says that these factors are "in particular, relevant" (NC6(7)). Other safety duties also have to be 'proportionate' a word that also encompasses riskiness – it is odd to draw out economic factors here despite the fact that the obligations will apply only to Category 1 services which will tend to be the largest.

10. In relation to large and risky user-to-user services (Category 1), the Government - having dropped some of the 'harmful but legal' provisions - seems to expect that if those providers claim to be tackling such material they must deliver on that promise to the customer/user (NC4(3)). Service providers have terms of service (defined in cl 203 as "all documents (whatever they are called) comprising the contract for use of the service (or of part of it) by United Kingdom users") that may go beyond the criminal law and encompass material that they believe is harmful to their users, doesn't 'fit' the service or is unpopular with advertisers.

11. The Government list of harmful content provided for the "Third Shield" (user empowerment tools, below) is an example (though there is no obligation on any provider to include provisions dealing with these issues in its ToS). The intention to make companies stick to their ToS reflects a widespread view that companies pick and choose their application of ToS or implement them with little vigour. These failings lead to harm when people encounter things which they had thought would not be there when they signed up.

12. The amendments also include provisions for making ToS more legible (NC4(7)) which reflects obligations found in the original transparency obligations (cl 13(6)). These legibility obligations carry echoes of measures in financial services which had similar problems and where the ToS complexity issue hasn't been solved.

13. Service providers which do not fall within Category 1 need not enforce their terms of service, or may do so erratically or discriminatorily (subject to any constraints in other areas of law) – and that includes search engines no matter how large. We discuss below that the Government may be changing the definition of Category 1 to broaden it somewhat.

Section 2: Impact on service provider behaviour

14. The new duties will make the largest and riskiest companies expend more effort on enforcing their ToS for UK users. The Government has not yet presented modelling nor, say, game-theory-type work on what this effect this will have on company ToS.

15. There are risks arising from the fact that there is no minimum content specified for the terms of service for adults – though of course all providers will have to comply with the illegal content duties (and these include specific lists of priority illegal content in Schedules 5-7). One former social media company senior executive felt the amendments would make ToS much longer and 'lawyered'. This might be especially so in circumstances in which "restrictions" (NC2) are in play – and this might militate against a company using non-take-down methods to control content.

16. Another view is that, faced with stringent application companies might make their ToS shorter, cutting out harmful material that is hard to deal with because they now might be liable if they don't deal with it. Or, if a service provider does deal with it, they may suffer competitively and reputationally if they run into issues with OFCOM and end up having to publish breach notices (following amendments being introduced in the Lords). By comparison, companies that chose to do nothing have an easier life. They might suffer reputational hits from content-harm when that becomes public - for example, because of whistleblower action or media reporting - but not from the regu-

lator under the new duties. We note that big service providers do have a track record of changing the nature of enforcement (Meta on modern slavery – see Haugen) or the terms of service themselves (Twitter recently on COVID misinformation) in response to management decisions, likely driven by financial considerations.

17. The Government presents no evidence either way. It isn't clear what companies might say now about what they will do in future to seek to influence policy thinking. The fact that there is no minimum requirement for ToS in the regime means that companies have complete freedom to set ToS for adults – and the ToS may not reflect the risks to adults on that service. Service providers are not obliged by the government amendments to include ToS in relation to the list of harmful content proposed by the Government for the user empowerment duties (below). Moreover, the removal of both the risk assessment for harms to adults (cl 12) and the previous obligation to summarise and publish the results (cl 13(2)) means that users will lack vital information to make an informed choice as to whether they want to engage with the service or not.

18. The decision to drop the broadly based adult risk assessment required in old clause 12 is regrettable. This would have provided an exceptional insight into trends and a forward look at harm reduction problems on the horizon. It would have provided in particular a useful insight into radicalisation and (non-terror) extremism. It would be useful to get onto the record why the Government has removed it; there was no discussion in the media material or WMS nor evidence justifying such a change. The benefit of such an 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 could be informed by OFCOM's use of information gathering powers.

19. We would therefore like to see the Bill amended (at clause 112, subsection 1) to require OFCOM to include in its annual report an assessment of the threats of unmitigated harm arising from the operation of online services in the United Kingdom.

20. Malign actors should not be able to buy their way around well enforced ToS. If advertising ToS are not well-applied they become a weak link to exploit and cause harm. One factor in the commercial success of online services is the very low barrier to buying ads, even with small budgets. It is possible that the way "Terms of Service" is employed here means that platforms' advertising content policies will be outside the scope of this clause. The definition of the term in cl 203 refers only to the documents constituting the relationship between service provider and user. Concerns about targeted advertising will likewise not be assuaged by the regime.

21. OFCOM is required to produce guidance on the above.

Section 3: User Empowerment - amendments 8-17

22. The final part of the so-called "Third Shield", these amendments allow a user to manage what harmful material they see by requiring the largest or most risky service providers to provide tools to allow a person to "effectively reduce the likelihood of the user encountering OR effectively alert the user to" [our emphasis] certain types of material. The Government proposes a list of such material in amendment 15 to go on the face of the Bill as follows:

'if it encourages, promotes or provides instructions for—
(a) suicide or an act of deliberate self-injury, or
(b) an eating disorder or behaviours associated with an eating disorder.
(8C) Content is within this subsection if it is abusive and the abuse targets any of the following characteristics—
(a) race,
(b) religion,
(c) sex,

(d) sexual orientation,
(e) disability, or
(f) gender reassignment.

Content is within this subsection if it incites hatred against people—
(a) of a particular race, religion, sex or sexual orientation,
(b) who have a disability, or
(c) who have the characteristic of gender reassignment.'

23. We don't think that this list nor the similar one published by former Secretary of State Nadine Dorries in a WMS in July has been debated before.

24. There is no linkage of these terms to existing definitions in statute or Crown Prosecution Service guidance. Some terms ('eating disorders', 'abuse') might have no current definition in statute or common law but understanding what each of these mean is crucial to the scope of protection provided. For example, is the term "eating disorders" intended to refer just to those disorders of a level triggering clinical diagnosis or is the intention to provide tools in relation to content related to a wider range of disordered eating? (The phrase 'associated behaviours' does not answer this question.)

25. We do not understand why the Government has omitted harmful health content when the July WMS, tabled by Dorries, set out a list of harmful but legal material including:

'Harmful health content that is demonstrably false, such as urging people to drink bleach to cure cancer. It also includes some health and vaccine misinformation and disinformation, but is not intended to capture genuine debate.'[7]

26. If this is the first time this list is debated, there is an opportunity in committee for a first principles discussion of what should be in it. For instance, climate change disinformation as we discussed in an article last year[8]. There may be other significant content areas that should also be considered.

27. Hateful extremism is a long-standing omission from the Bill which hasn't been properly debated (Rehman Chishti MP raised it at Report[9]). 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.[10]

We therefore support Opposition amendments 15 a) and 16 a) to address these points above.

28. The new amendments do not alter the provisions for user empowerment tools in relation to unverified users which already exist in the Bill.

29. Government amendment 9 requires the user empowerment tools to be 'effective' (they had not been required so previously) and amendment 12 requires the tools now to be 'easy to access'. OFCOM has to produce guidance on this and the tools in general above. We assume that OFCOM

7 https://questions-statements.parliament.uk/written-statements/detail/2022-07-07/hcws194
8 https://www.carnegieuktrust.org.uk/blog-posts/an-inconvenient-truth-radical-change-needed-to-online-safety-bill-to-tackle-climate-disinformation/
9 "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
10 https://www.gov.uk/government/publications/operating-with-impunity-legal-review

will need to include the above in its risk assessment processes and, possibly, risk profiles.

Should the user empowerment tools be turned on by default?

30. The tools now proposed in clause 14 reflect an *a priori* view of the Secretary of State that the content referred to is harmful and that people need to be provided with tools to protect themselves. The amendments provide that an adult should be able easily to find such tools and turn them on. We note that in a number of cases people at a point of crisis (suicidal thoughts, eating disorders, etc) might not be able to turn the tools on due to their affected mental state; for others default on saves them from having to engage with content to utilise tools in the first instance. Given that a rational adult should be able to find the tools they should be able to turn them off just as easily. The existence of harms arising from mental states in our view tip the balance in favour of turning the tool on by default. In the work we did with Anna Turley MP in 2016, we proposed an abuse blocker that was on by default in PMB Malicious Communications (Social Media) 2016:

'Operators of social media platforms ..must have in place reasonable means to prevent threatening content from being received by users of their service in the United Kingdom during normal use of the service when the users—
(a) access the platforms, and
(b) have not requested the operator to allow the user to use the service without filtering of threatening content.'[11]

We therefore support Opposition amendments 102 and 103.

31. One question is how the user empowerment tools relate to the Terms of Service duties. If a tool gives rise to false positives would that constitute a restriction for the purposes of the new duty? NC2 (3)(a) seems to exclude effects arising from the user empowerment duty. It is unclear whether user tools that are useable beyond the context of the areas listed in the user empowerment duty would benefit from NC2 (3)(a) or whether– when used outside the areas envisaged by cl 14 –they would constitute 'restrictions'? This is particularly important if tools are optimised to specific types of content, rather than having a one size fits all approach to tools (which would be likely less effective). Of course, this could be covered in the ToS but that might give rise to the risk of 'lawyering' and the disincentives noted above. The precise relationship between the two duties should be clarified.

32. One final point to consider is how effective these tools (as well as the enforcement of terms of service) will be when the service itself is contributing to the problem either through the delivery of adverts or the personalisation – and therefore repeat delivery – of content. As regards adverts, it is not hard to envisage that problems might arise through the content of adverts. Imagine, for example, the targetting of nutritional supplements or wellness products to a young person who had an eating disorder. Will the effectiveness of tools be assessed taking into account the content of adverts too? The Minister should be asked about the relationship between the new tools and paid-for advertising.

Section 4: reinstating some form of risk assessment

33. As noted, the companies are not obliged to carry out risk assessments for any of these duties, making them very different from others in Part 3 of the Bill. 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, the Government says that measures have to be 'effective'. This sug-

11 https://bills.parliament.uk/bills/1877

gests 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.

34. The removal in this block of amendments of the adult risk assessment obligation (in original Clause 12 companies had to assess the risk of harm to adults and (original Clause 13) report it), will mean it is much harder for users and civil society to assess what problems arise on the platforms – and the role of product design in those problems. The consequential removal of the transparency obligation (service providers had been required to inform customers of the harms their risk assessment had detected) means that users (as consumers) will not have the information to assess the nature/risk on the platform.

35. 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.

Section 5: 'Categorisation' of companies NC7, amendments 48, 49, 76-93

36. The Government appears to be preparing to broaden the criteria for selecting which companies are likely to be in Category One. They add the 'characteristics' of a company's service to the rather crude size and functionality metrics employed before. The Government also allows for a list to be drawn up of companies that are close to the margins of categories or 'emerging Category One'. This will give more regulatory certainty. Whether this deals with all concerns arising with regard to the drawing of boundaries (eg Secretary of State's powers in this regard) is another question.

37. Sir Jeremy Wright MP, who had tabled similar amendments at Report stage, indicated in the debate on 5th December that the Government probably would not go far enough in either of the amendments that have now been tabled at Committee. On characteristics, Wright said:

"I welcome the Government's adding other characteristics to the consideration, not just of threshold criteria, but to the research Ofcom will carry out on how threshold conditions will be set in the first place, but I am afraid that they do not propose to change schedule 11, paragraph 1(4), which requires regulations made on threshold conditions to include, 'at least one specified condition about number of users and at least one specified condition about functionality.' That means that to be category 1, a service must still be big. I ask the Minister to consider again very carefully a way in which we can meet the genuine concern about high harm on small platforms. The amendment that he is likely to bring forward in Committee will not yet do so comprehensively. I also observe in passing that the reference the Government make in those amendments to any other characteristics are those that the Secretary of State considers relevant, not that Ofcom considers relevant—but that is perhaps a conversation for another day."

38. On where the boundary lies in drawing up the list, Wright went on to say:

"Again, the Minister may say that there is an answer to that in a proposed Committee stage amendment to come, but I think the proposal that is being made is for a list of emerging category 1 services—those on a watchlist, as it were, as being borderline category 1—but that in itself will not speed up the re-categorisation process. It is the time that that process might take that gives rise to the potential problem that new clause 1 seeks to address."[12]

39. Minister Scully, summing up, replied "I look forward to working with him to get the changes to the Bill to work exactly as he describes."

12 https://hansard.parliament.uk/commons/2022-12-05/debates/E155684B-DEB0-43B4-BC76-BF53FEE8086A/OnlineSafetyBill#contribution-BA305D5F-0EB2-4EB3-8F22-394566A08F88

Child safety

40. We note the amendments on child safety (1-5) and wait for a considered view on these from civil society organisations expert in child protection.

Carnegie UK
December 2022
Contact: maeve.walsh@carnegieuk.org