# THE ONLINE SAFETY BILL: OUR INITIAL ANALYSIS

**March 2022**

This is our initial response to the Online Safety Bill. We have written extensively about our concerns on the draft Bill and are pleased to see that many of those have been addressed in the Bill. As always, we acknowledge the immense effort that has gone into getting the proposed legislation to this stage - from countless DCMS and other Government officials, multiple Ministers, the members of the Joint Committee who scrutinised the draft so effectively and other Parliamentary Committees whose inquiries have shaped the final Bill, and an array of civil society campaigners whose insight, analysis and tenacity has led to some significant Government concessions.

The Bill is by no means perfect, however; and we still have many concerns that we will need to work through in the coming weeks and during the passage of the Bill through the House. For now, this initial response is structured around a number of tests that we had set for the Bill before its publication which reflect many of the areas and topics we have written about in recent years. More will follow from us in due course. In the meantime, for those coming to this fresh, we have also published a short explainer on what the Online Safety Bill does, which can be read alongside this initial analysis.

## Summary: areas of improvement and areas of concern

We note the following improvements:

- Compared to the draft, the approach has become more clearly systems-based, which is welcome. But it is still some way short of a truly systems-based approach: the structure of the Bill, the addition of content-specific new sections, and its drafting still pushes services to addressing harmful content, rather than systems - including the business model and algorithms. **Further amendments to consolidate its focus on systems would make the regime more effective** and provide reassurance to those worried about its impact on free speech.
- The protections for children have been significantly strengthened, including with the introduction of measures in relation to access to pornography, although children's rights groups still have concerns about the Child Access Assessment.
- The enforcement measures seem to be more effective and tied back more explicitly to companies' risk assessments and the obligations that flow from them.

There are still areas of concern including:

- The Bill remains too complex. Some simplifications have been brought in. But **the overall structure is still too difficult to navigate**, the principle-based elements (such as the Online Safety Objectives, which appear in Schedule 4) and definitions are buried at the back, rather than upfront, and new concessions have been, in the main, bolted on. Complexity, which is the Government's choice here, increases the regulatory burden.

- The decision to stick with different categories of services, which we and the Joint Committee had argued against in the draft Bill, does not fit with a proportionate or risk-based regime and will not catch harms on fast-growing platforms. This is a significant hole in a systems-based approach: very large size itself can be an absolute indicator of risk and using very large size as a proxy brings administrative simplicity, but it is wrong to suppose that smaller size means lower risk. **We continue to recommend that categories of providers are removed and risk assessment duties apply across the board**.
- Too many powers remain with the Secretary of State, with too much of the regime's detail left until secondary legislation. **We suggest that the Secretary of State's powers to direct OFCOM on the detail of its work (such as codes) are removed**. For National Security, government should have carefully constrained powers. OFCOM's Board needs to be bolstered to oversee National Security issues.[1]
- The OSB remains weak in relation to addressing harms to adults and wider societal harms. Given that the Government is listing harms in the Bill, **it should list harms both to children and to adults in a new Schedule 7(a) and 7(b) respectively** to enable debate and give victims and companies more certainty. Three years on from the White Paper, the Government must know what it intends to include.
- Measures to tackle mis/disinformation are still largely absent from the Bill. The Government should reform the unaccountable system of nudging service providers to deal with issues using civil service teams, on a possibly piecemeal basis, and **bring disinformation into scope**. It should create a formal mechanism for state actor disinformation, such as that from Russia, and for public health issues such as COVID.
- The Bill envisages cross-organisation working between OFCOM and other regulators but does not create the powers that may be needed for domestic inter-regulator co-operation. **The Government should give clear powers to OFCOM (similar to those in Clause 97) to ensure that case files can flow through and between regulatory systems in accordance with the law.**
- Fraudulent advertising has the same impact on victims, whatever platform it occurs on. **The new fraudulent advertising powers should apply with similar strength to all companies, not just Category 1 and 2A**.
- **Human trafficking offences are a serious omission from Schedule 7 that should be rectified**. Advertising is an important route for modern indentured servitude and should clearly be in scope for that offence.
- The regime will not start working properly until 2024. **The Government should seek to bring some aspects forward;** one area worth more study could be making companies' terms and conditions enforceable under the regime from the moment of Royal Assent.

We work through these areas in more detail below.

1 This paragraph was amended 08/04/22 to clarify the recommendations on the Secretary of State's powers.

## Is the Bill workable?

When we published our proposals for extensive amendments to the Online Safety Bill in November 2021, we argued that the Bill was too complex and needed to be simplified and strengthened:

> *A simpler Bill will lead to better outcomes for victims. Others benefit too: legislators, who will need to scrutinise and amend it further; companies (and their lawyers) who will need to comply; the regulator, Ofcom, who will take enforcement decisions based on it; and civil society organisations advocating for victims.*

The Bill's structure remains complex and opaque, a consequence of policy decisions made by the Government as well as drafting approaches chosen by Parliamentary Counsel. The re-ordering proposed by the Joint Committee to make the objectives of the regime clearer, which were similar to proposals we made in our amended Bill, have not in the main been actioned and the Bill still lacks any general explanation at all. The structure of the Bill is cumbersome, with much nesting of qualifiers on terms scattered through the Bill: for instance, risk assessments link back to illegal or harmful content; and harmful content depends on an amended definition of harm (at clause 187) and of content (clause 189). Definitions are listed at the end, with other definitions scattered throughout the Bill, and an index brings them all together at clause 190. While this is not unusual, the implications are hard to assess because of this nesting. (We also have concerns about the drafting of the journalistic content exemptions (clause 16) and particularly the definition of what that content is (explanatory notes, para 132), which we will return to in due course.)

The Explanatory Notes have improved a little from the draft. But overall comprehensibility would be improved by some more simple use of diagrams and flow charts that show 'user journeys' from a number of perspectives[2]. It is in the Government's interest to make the regime easier to understand - it will help passage of the legislation as well as reducing the regulatory burden. If they insist on keeping the complex structure, then the Government and/or OFCOM (in due course) should provide better supporting materials.

The Online Safety Objectives (which we had argued should have been at the front of the Bill in a new clause 1A) are now in schedule 4, which is unhelpful. These objectives, in addition, then have an overlap with some of the requirements in each of the duties in the main body of the Bill re illegal content etc.

We do, however, note some welcome simplifications to the structure, which include: the consolidation of the duties in clause 6(2); the incorporation of the child sexual exploitation and abuse (CSEA) and terrorism offences into priority illegal content; and the reordering of the risk assessment duties so that they each sit next to the relevant safety duty, which improves flow and readability.

But, where new concessions have recently been made, they are introduced as new sections (for example on pornography, user verification and fraudulent advertising); and the impact of these additions is unclear on the overall schema, as yet. Certainly, it has led to some convoluted drafting around pornographic content;[3] and the provisions around user verification (in clause 57) have been split off from those relating to the right to filter out non-verified users'

2 User journey mapping is a technique for understanding complex system problems espoused in particular by the UK Cabinet Office https://www.gov.uk/service-manual/design/map-a-users-whole-problem

3 See e.g. Clause 7(1)

content and other user empowerment tools found in clause 14 as part of the Category 1 safety duties.

Moreover, the consequences of the boundary between advertising and other content being removed is not clear. If adverts fall within the definition of user-generated content (and it seems that they would, per clause 181 and 49(3)), then adverts are regulated content and the machinery behind advert delivery comes within scope where the content is either criminal or harmful to children or to adults (noting that economic/financial loss is not a relevant harm). This inclusion is likely to be a step forward though there will be awkward boundaries to navigate, especially given the special regime for fraudulent ads is applicable only to some service providers. Fraudulent advertising has the same impact on the victim regardless of where they are harmed by it: the new powers should apply with similar strength to all companies not just Category 1.

The Bill still leaves much to do in secondary legislation - with resultant delays to clarity about the regulatory scope. For example, companies will not know which category of service they fall into until after the Secretary of State has published definitions on thresholds and laid these as secondary legislation and OFCOM then "as soon as possible after" those regulations are made publishes the register of categories (cl 81); and as a result will not know all the duties that could apply to them. The Government could easily give broad indications of its thinking in a carefully caveated speech by the Secretary of State, reducing uncertainty and making debate during passage more practical than theoretical. In the three years since the White Paper, the Government must have formed its outline views on the basics of which companies will be subject to which regimes.

The scope of protection offered by the Bill also won't be clear until after secondary legislation. While there is a preliminary list of priority criminal content in schedule 7, priority content harmful to adults, and both primary priority content and priority content for children, will be introduced by statutory instrument (for which there is no timing given). These areas are critical for victims seeking to understand if the Bill will protect them in future, as well as for companies that might have to manage these risks. As the Government has embarked upon listing harms in the Bill, it should add the categories of priority content for both content harmful to children and to adults as a new Schedule 7(a) and 7(b) and make its position clear by Second Reading.

More detail on what companies will be expected to do to comply with the regime will follow in OFCOM guidance and codes of practice[4], which cannot be produced or consulted upon until after Royal Assent. For example, the Children's Access Assessment (CAA) is dependent on action from OFCOM. This sequencing is problematic in that the obligations in the regime on companies will not become clear until these products from OFCOM are in existence. We note that OFCOM has said recently it will be ready to proceed with these as soon as practicable and is preparing for "phased implementation" of the regime as its powers come onstream. Note also the existence of materials in relation to

- the Video-Sharing Platforms provisions,
- the draft CSEA and Terrorism Codes produced by the Home Office; and
- the voluntary safety by design guidance from DCMS

though OFCOM's Online Safety codes may differ from these.

4 There are three codes of practice that are named (terrorism, CSEA and fraudulent advertising) and a requirement on OFCOM to produce code(s) on "relevant duties "OFCOM also has at least 10 requirements to produce guidance (risk assessment, child access, child risk assessment, adult risk assessment, record-keeping, user ID verification, transparency reporting, porn duties, guidance re proactive tech, guidance re enforcement action). More details on what should be in some of the codes can be found in Schedule 4.

The draft OSB defined illegal content on the basis of when the service provider had "reasonable grounds to believe" something amounted to a "relevant offence". In exercising their reasonable belief, service providers were to have regard only to content on their respective services. There has been a shift: now the definition in clause 41(3) is simply that material "amounts to a criminal offence". This seems to lose the language that recognised that services would be making a judgement call (either when they thought something was criminal which was not, or when they thought something was acceptable which was not). The Bill does, however, include language recognising this point in clause 9(6) ("the provider *reasonably considers* is illegal content") and 24(6) (same as 9(6)), with regard to consistent enforcement of terms of service/ community standards. However, this space for uncertainty is not reflected in the obligations to have systems in place. We question whether this is quite right: surely the obligation there should reflect the fact that a system will be operating by reference to classes of content - but "amounts to" seems to suggest a more individual assessment as to whether content is criminal – and if so, this would make the regime less workable.

It seems that the regime will not start working properly until 2024. The Government should seek to bring some aspects forwards - one area worth more study could be making companies' terms and conditions enforceable under the regime from Royal Assent, which is expected at the end of 2022.

## Is it systemic and risk-based?

Our work to describe and advocate for a duty of care regime for online harm reduction has always been rooted in our belief that systemic, risk-based regulation - such as that which is established in countless other sectors - is the most appropriate approach for the online environment, requiring companies to account for, and mitigate, the harm that arises from the design and operation of their systems rather than focusing on individual items of content.[5]

Our judgement on an initial, but close, reading of the Online Safety Bill is that, compared to the draft, the approach has become more clearly systems-based. This is partly a result of DCMS tidying up and addressing many errors and loose ends and making positive drafting improvements in relation to both the risk assessment and safety duties, as well as the description of harm and functionalities. It is also due to a clearer risk assessment process, which we describe below. However, by retaining the categorisation of services - which we have argued against previously, as did the Joint Committee - the risk-based regime does not apply equally and will lead to gaps in enforcement and the likelihood of harms arising and proliferating unchecked on smaller, but potentially fast-growing platforms, before the process for recategorising them can kick in.  Very large size itself can be an absolute indicator of risk and using very large size as a proxy brings administrative simplicity, but it is wrong to suppose that smaller size means lower risk. We continue to argue for the categories of services to be removed.

### System-focused

Overall, we welcome the fact that the language used around risk assessments reflects that used in other regulatory systems and sectors where there is great experience of using this technique to reduce harm. This is a positive development. For instance, risk assessments must now be "suitable and sufficient" – language we proposed some time ago[6]. The Explanatory Notes describe this meaning (in relation to illegal content, for example) that: "service providers

5 See our full reference paper from 2019 here: https://d1ssu070pg2v9i.cloudfront.net/pex/pex_carnegie2021/2019/04/06084627/Online-harm-reduction-a-statutory-duty-of-care-and-regulator.pdf

6 "Suitable and sufficient" now aligns with the Management of Health and Safety at Work Regulations 1999 https://www.hse.gov.uk/managing/delivering/do/profiling/the-law.htm

will therefore need to assess how likely content is to be illegal, and therefore how likely it is that illegal content is present on their service."[7]

This description however does not focus so much on the role of the hazards in constituting that problem, although an emphasis on features can be seen; for example, in clause 8(5), especially sub-clause (d), which identifies things a service should cover in its risk assessment. The question is whether the language chosen, particularly in relation to the safety duties, implicitly pushes platforms mainly to addressing the problem - that is the content - rather than the underlying causes and exacerbating factors of platform design. For example, the specific duties in clause 9(3) (takedown of content and minimisation) can be seen this way. Conversely, clause 9(2), which obliges service providers to "take or use proportionate measures to effectively mitigate and manage the risks of harm to individuals, as identified in the most recent illegal content risk assessment of the service", ties back more clearly to harm and the risk assessments. Given this possible ambivalence about systems or content-based, the guidance and codes of practice that will come from OFCOM take on a higher significance.

We note that the language in the safety duties remains that of proportionality (as opposed to reasonableness - the threshold found in other duties of care), defined by reference to size/resources as well as level of risk/severity of harm.[8] This definition is new.

It is good to see the emphasis that the duties "apply across all areas of a service, including the way it is operated and used as well as content present on the service"[9] - though we note a significant omission in that this language is not replicated in relation to content that is harmful to adults. Having said that, the expanded definition of "harm" (clause 187) provides that references to harm include "the manner of [content's] dissemination (for example, content repeatedly sent to an individual by one person or by different people)"[10]. Presumably this could include where the platform personalises content to similar effect (assuming any other relevant thresholds are satisfied) as well as pile-ons and mobbing. Whether this is sufficient, given the omission in the safety duty for content harmful to adults itself, remains to be seen. Another issue in this context is that the fact of dissemination - in the context of the children's safety duty - is specifically excluded in relation to some of the duties: 11(3) (prevention from encountering content) and 11(5) (terms of service in relation to clause 11(3)) There is some degree of tension between this and clause 11(4).

The definition of harm itself is modified to be equivalent to "risk of harm and potential harm" which is important given the centrality of risk assessments (which take place before any harm has occurred) to the regime[11].

More generally, the role of functionalities (and options other than take down) are more visible than in the draft Bill. While the risk assessments always took into account functionalities (the definition though is tucked away as the back of the Bill), new examples include language in the safety duties - clauses 9(4), 11(4) and 13(4) (though the list in 13(4) is very limited); and analogous provisions for search (clauses 24(4) and 26(4)). Whether these lists are sufficient requires further thought: on user-to-user platforms, how are disposable accounts dealt with, for example, and how do default settings get taken into account? The user empowerment obligations (and not just the rules around engagement with anonymous accounts but more generally in clause 14)

7 Explanatory note: para 81
8 See e.g. clause 9(9), 11(9)
9 Clause 9(4); cl 11(4); 24(4); 26(4)
10 Clause 187(3)(c)
11 Clause 187 (5)

refer to features. Even the new "proactive technology"[12] provisions act to surface the role of the underlying system in the content environment created.

## Risk assessment process

OFCOM can carry out a broadly based risk assessment of all harms. This assessment (which was found in the draft Bill at clause 61 and is largely unchanged, aside from the welcome requirement to publish a risk register and the risk profiles) underpins the entire regime. Here, there remains a welcome focus upon the "characteristics" of the service (that is functionalities, user base, business model, governance and other systems and processes[13]) rather than just the content[14] - though the risk of harm to be taken into account is only that deriving from the three types of content (addictive design per se would not seem to be covered, for example). Nonetheless, a broadly based risk assessment will inform better policymaking downstream.

The Bill does not seem to explicitly require that OFCOM assess all risks at once. This is pragmatic as OFCOM must be able to prioritise the more serious and significant issues when the regime starts up; for instance, to begin assessing risks for children and from terrorism first. Recent comments by OFCOM at an event on the Online Safety Bill suggest that this will be their approach, with four codes of practice ready for consultation at Royal Assent covering terrorism, CSEA, illegal priority content and the risk assessment.

By contrast to the draft Bill, OFCOM cannot develop risk profiles by reference to identified harms to adults that are not designated as priority.[15] Does this mean that non-designated content that is harmful to adults nonetheless poses no relevant risk? Publishing OFCOM's risk assessment will potentially reveal those unaddressed harms. Moreover, OFCOM has an obligation to carry out reviews on content that is harmful to children and content harmful to adults and publish a report not more than three years apart.[16] The report must include advice as to whether the regulations specifying the priority content need to be changed, though the Secretary of State is not obliged to follow that advice. Note that there is no equivalent update process as regards illegal content.

A significant and positive change to risk assessment in the regime is the inclusion of specific provisions allowing OFCOM to take action against deficient risk assessments.[17]

Perhaps as a consequence of the obligations relating to fraudulent advertising, the exclusion of "paid-for advertisements" found in the draft Bill has been removed. This suggests that the ad delivery systems in general could be relevant to the risk assessment and risk mitigation duties, in addition to the specific provisions on fraudulent ads (which seem to relate to specific content-based rules). This we feel to be potentially an important positive step given the role of the advertising funding and the overall business model in supporting certain types of problematic harms. The characteristics of advertising delivery systems should explicitly be brought into scope for risk assessment.

Overall, these small changes to risk assessments and an increase in the role of the system in managing risk are welcome and should strengthen the regime by making it more systemic.

12 Defined clause 184
13 Clause 83(12); 'functionalities' is defined at cl 186
14 Clause 83 (2)
15 Clause 83(6)
16 Clause 56
17 Clause 114

## Does the Bill protect fundamental rights?

There is a duty on platforms to consider freedom of expression and privacy when deciding on and implementing safety measures and policies. This is a change from the language in the draft Bill (which referred to "safety policies and procedures"). Precisely what is meant by this is not clear - is it any step taken to mitigate as a result of a risk assessment (including potentially a decision not to include a feature)? Or is it narrower? The former could be too broad - potentially including purely commercial decisions - but there is still the question of when the obligation kicks in.

Another change from the draft Bill is in the language relating to privacy. As regards the obligation on all service providers, rather than referring to the right itself (as has been done in the case of freedom of expression), the text now refers to "any statutory provision or rule of law concerning privacy that is relevant to the use or operation of a user-to-user service."[18] This seems a narrower formulation, arguably excluding some considerations that are protected by Article 8 ECHR. Of course, one might argue that providers, which are not public bodies, are not directly bound by human rights obligations. But in that instance, why is there a direct reference to freedom of expression (rather than statutory provisions and rules of law as for privacy)? It is still not clear where the balance is to be struck between the rights of vulnerable users oppressed (and silenced) by sub-criminal speech and bad systems versus the speech of other individuals and companies. The extra obligations on Category 1 providers in relation to freedom of expression and privacy (without the difference in language found in the duties on all providers) remain, though they are - as is the case for the general duty - only obliged to consider those rights in relation to their users and not, for example, the victims of image-based sexual abuse or racial abuse (assuming victims not to be users of the platform).

The Bill contains an obligation on OFCOM to include in its annual report a statement on steps and processes to ensure that OFCOM's functions have been exercised compatibly with Articles 8 and 10 ECHR. This was not in the draft Bill. It is good to see this obligation introduced but, while both Articles 8 and 10 are mentioned, the other civil and political rights are not. Other rights could have relevance online. It would be good for OFCOM to report on these broader considerations where appropriate; for example, when revealed by transparency reports or super complaints.

## Does the Bill protect the most vulnerable?

The Bill has been significantly strengthened with regard to children, with the new duties in relation to pornography (part 5), and the new requirement on providers to report the existence of CSEA to the NCA (part 4, chapter 2). This is on top of the existing measures with regard to CSEA and the requirement for a children's access assessment. We will leave others more expert in this area to comment on these measures for now and focus our analysis here on vulnerable groups of adults.

The definition of harm (clause 187) now includes a welcome reference[19] to the way in which an individual can do something which affects another individual due to their characteristics or membership of a particular group. The reference to a group means that consideration is not limited to the categories identified in the Equalities Act, or for the purposes of hate crime, and presumably could be understood to include intersectional groups also. Note, however, that in assessing whether content is harmful there is a quantitative threshold of "an appreciable num-

18 Clause 19(3)

19 Clause 187 (4) (b)

ber" of adults[20] or children. Does this remove smaller groups from protection? In sophisticated user-to-user networks, or search services, with tens of millions of users, advertisers can address numerically very small audiences - and commercial claims are made of personalisation of user experience - so "appreciable" could be a small number. We note that the Explanatory Notes say that "kinds of content that affect only one person or very few people" are not in scope";[21] but more guidance would be helpful. Moreover, to require action, the harm must be significant and the risk must be material.

In general, however, the Bill remains weak on harms to adults - all the more so because the types of harms to be prioritised are as yet unknown and will not be set out until the Bill has received Royal Assent; and the duty applies only to Category 1 providers (which we refer to above). Non-designated content that is harmful does not require action on the part of service providers, even though by definition it is still harmful.

We know that many campaigners are concerned that protections for women and girls are not included in the draft Bill, a concern supported by the Petitions Committee in its report on Online Abuse.[22] While Schedule 7 does include a list of sexual offences and aggravated offences, to which the Government has referred in its response to the Petitions Committee,[23] the Bill makes no concessions here and the wider context of Violence Against Women and Girls (VAWG) is not addressed. We will be looking further at the types of harms that might need to be included to ensure the protections required are delivered.

Human trafficking offences (noted by Frances Haugen's whistleblowing revelations) are a serious omission from Schedule 7 that should be rectified. Advertising is an important route for modern indentured servitude and should clearly be in scope for that offence.

We also have some concern that the overarching risk assessment for adults in clause 12 (which only covers category 1 companies), expects services to consider the fact that some groups are more likely to encounter harmful content and behaviour and are more likely to be harmed by it (cl 12(5)(d), is constrained by the scope of harmful content. As noted above, there is a quantitative threshold for determining which content is harmful (and therefore to be taken into account in the risk assessment) which might not well serve smaller groups.

## The role of the state

Compared to the broadcasting regime, there is considerably more intrusion by the executive and Parliament into the work of the independent regulator. We have written extensively on why this was problematic in the draft Bill - and it remains so here. Clause 40 permits the Secretary of State to direct OFCOM to change a code of practice. Whilst the draft Bill permitted this 'to

20 Clause 54(3)(b)

21 EN para 329; p54

22 We recommend that the Online Safety Bill should include abuse based on the characteristics protected under the Equality Act and hate crime legislation as priority harmful content in the primary legislation. It should also list hate crime and Violence Against Women and Girls offences as specific relevant offences within the scope of the Bill's illegal content safety duties and specify the particular offences covered under these headings, as the draft Bill already does for terrorism and Child Sexual Exploitation and Abuse offences. (https://committees.parliament.uk/work/307/tackling-online-abuse/publications/)

23 "We agree with the Committee's suggestion that hate crime offences and offences associated with Violence against Women and Girls should be listed as priority illegal offences on the face of the Bill. Government announced this change on 5 February. This includes offences relating to sexual images (i.e. revenge and extreme pornography), and harassment and stalking offences, as well as acts intended to stir up racial hatred, religious hatred or hatred on the grounds of sexual orientation and racially or religiously aggravated harassment and public order offences. This means all services will need to take steps to remove and prevent users from being exposed to this content. This will result in women and girls being better protected online and proactive measures to tackle illegal abuse on the grounds of the listed characteristics. Beyond the priority offences, all services will need to ensure that they have proportionate systems and processes in place to quickly take down other illegal content directed at women and girls once it has been reported or they become aware of its presence". (https://publications.parliament.uk/pa/cm5802/cmselect/cmpetitions/1224/report.html)

ensure that the code of practice reflects Government policy', clause 40 specifies that any code may be required to be modified 'for reasons of public policy'. While this is more normal language, it is not clear in practice what the difference between the two sets of wording is. Implicitly it seems that this excludes 'national security or public safety', which are specifically dealt with in relation to the CSEA and terrorism codes in clause 40(1)(b). This provision would be unnecessary, given clause 40(1)(a) applies to any draft code include CSEA/terrorism, if public policy were to cover national security or public safety. Different rules apply in relation to CSEA and terrorism codes in that they are reviewed[24].

There appears to be no Parliamentary control over the Secretary of State's actions in respect of public safety or national security nor oversight by a competent body such as the National Police Chiefs Council or Directors of Public Health for public safety. We suggest that Secretary of State intervention in the substance of the codes is, as a matter of principle, problematic. We do, however, recognise the importance of the Government to identify when there are national security issues that need to be addressed; this flagging of a problem is different from saying how it should be resolved, which should be a matter for OFCOM. We also think that any special intervention by the State in public speech must be limited to national security. When public safety issues become sufficiently serious to allow the Government to issue directions to a regulator (such as COVID) we suggest that they are national security issues.

The new relationship between intelligence services and OFCOM (see clause 99) is now codified but with no oversight (mundane issues like OFCOM having board members and senior staff who have undergone say Developed Vetting[25] etc would help). And also - while protecting security advice is necessary - the caveat that allows this as a reason for redacting information from reviews of codes (clause 43 (6)) does not have any oversight. Who will ensure that it is only the national security/public safety relevant content that has been legitimately removed or obscured before publication?

We suggest that the Secretary of State should not have powers to direct OFCOM on the detail of its work (such as codes). The public policy and public safety sections of Clause 40 should be deleted. The Secretary of State should be able to give high level guidance on national security issues to the regulator, but not interfere in its detailed work such as codes, guidance and enforcement strategy.[26] OFCOM should have sufficient capability to assess and discuss national security issues with the Government - such as board members with appropriate backgrounds and sufficient security clearance for key staff as well as the Chair and CEO.

More generally, the process used to make regulations now mainly requires the affirmative process. While this is an improvement on the draft Bill, it is important to note that this still does not allow Parliament the same freedom to consider the substance of the text as when matters are dealt with by primary legislation. (We have noted above, in relation to content harmful to children and content harmful to adults, that the priority content will only be identified by statutory instrument rather than listed on the face of the Bill (as for criminal matters)).

The Secretary of State also has powers to amend the Online Safety Objectives (Schedule 4, paragraph 7) - and indeed the Bill (clause 173) - by regulations too.

Overall we are pleased that there is greater use of the affirmative procedure in making of regulations but the extent of delegated law-making by the Government in the OSB remains worry-

24 Clause 43 and 40(2)-(4)

25 DV clearance can allow the sharing of intelligence material with an individual https://www.gov.uk/government/publications/united-kingdom-security-vetting-clearance-levels/national-security-vetting-clearance-levels

26 This paragraph was amended 08/04/22 to clarify the recommendations on the Secretary of State's powers.

ing for independent regulation.

## Transparency

Statutory transparency is important to give enough information to identify where problems arise and the need for an investigation and as a corollary the use of the information powers. We suggested a number of amendments to the draft Bill in this regard last year.

There are three aspects to transparency: what information should be given; by whom; and disclosed where. As regards the first, Schedule 8 listing the headings for public transparency reports now gives OFCOM the possibility of specifying "any other measures … which relate to online safety matters", rather than a closed list as before. The obligations apply, however, only to Category 1, 2A and 2B - and we still do not know which providers will fall within these categories – and as a corollary which services fall outside. As a consequence, the extent of the limitation on transparency is unclear.

There is no requirement to publish risk assessments or actions relating to compliance with safety duties (with the exception of the impact assessments Category 1 providers must carry out in relation to freedom of expression and privacy of users) – even to OFCOM (though it can presumably ask for that information if it is investigating). So, while it might now have clearer powers of enforcement around deficient risk assessments, it is unclear how it will know that there is a problem. Presumably the transparency reports are envisaged as filling this gap.

While OFCOM must publish its own transparency reports, these will be based on the providers' reports and seemingly therefore also not cover providers which are not obliged to report.

Transparency will be vital to civil society groups seeking to make informed and effective use of the super complaints mechanism. To what extent does this limitation affect their ability to do so?

The Government's aspirations for the regime to be world leading would better be met by publication of risk assessments for high-risk companies. These would prove invaluable to regulators around the world seeking to contain online harms.

## Societal harm

The Government's concessions on a number of criminal offences now listed in Schedule 7 (such as fraud; sale of a realistic imitation of firearms) have used the priority illegal harms route to open up areas of harm which are not specifically to individuals. The new definition of harm also possibly opens the door for societal harm to be included as it recognises that members of a group being affected by comments directed at another member of a group. For clarity, we will have to wait and see on statutory instruments as regards content harmful to adults and children and as we have suggested, we feel it would be preferable for the initial list of priority harms to be included in the bill. Presumably, societal harms that could be harmful to adults could equally (if not more so) be harmful to children.

We know that many civil society organisations remain concerned about the lack of provisions for mis/disinformation; we will return to this issue in more detail in due course but it is notable that the Government's response to the Joint Committee's report, which rejected its recommendations in this area, went no further than all its previous statements on this issue during the course of the development of this legislation. Given the level of evidence-based concerns

about the scale and impact of this - which are comparable to those expressed by campaigners in relation to e.g. fraud and scams, or anonymity - it is difficult to understand objectively why the Government was not prepared to consider similar concessions here.

There are two significant disinformation issues that the Bill does not address: disinformation supported by state actors and COVID disinformation. Several unaccountable civil service groups[27] exist to nudge service providers on these issues but we have no record of their effectiveness. This very direct state interference in the media gives rise to concerns. The Government should reform this system and bring disinformation firmly into the scope of the regime, putting the disinformation cells under OFCOM's independent supervision.

## Effective enforcement

We had raised concerns previously about this, in particular the requirement to "comply or explain". This seems to work better now – OFCOM can assess whether compliance is in breach of duty or not.[28] The record-keeping duties also have been expanded both as regards providers which are following a code[29] and those which seek to use "alternative measures"[30]. This should facilitate enforcement. While there are obligations on service providers to review compliance,[31] it is unclear what their internal consequence of finding defective compliance is. While OFCOM may ask for this documentation under its enforcement powers, service providers do not need to provide this to OFCOM as a matter of course; again that presumably is the role of transparency reporting.

The useful table in Clause 111 now sets out clearly the enforceable requirements. Apart from the change in formatting, we see the new user empowerment provisions and the linked user identity verification provision, the fraudulent advertising obligations in clauses 34 and 35, the obligation to report CSEA content to the NCA and the obligations on providers of pornographic content being included as enforceable requirements. These changes reflect new substantive obligations. Even under the draft Bill, risk assessments were enforceable requirements. The Bill, however, clarifies that this means insufficient risk assessments can constitute a failure in duty as well as a failure to do a risk assessment at all.

The OSB extends the 'skilled persons' provision. As was the case in the draft OSB, OFCOM may appoint a skilled person or require providers so do. This latter provision[32] will allow OFCOM to bring in skilled talent when required (and paid for by the company[33]) which is sensible resource management in a tight skills market. Additionally, however, service providers are required to give the skilled person assistance[34], which is an 'enforceable requirement' for the purposes of clause 111.

The fulcrum of OFCOM's enforcement is the "confirmation decision".[35] These are issued after notices ("provisional notices of contravention") allowing a service to correct behaviour in relation to obligations listed in clause 111.[36] In confirmation decisions, OFCOM can direct service providers to take specified steps to comply with a notified requirement or to remedy a failure

27 Dinenage-Puttnam – letter re counter disinformation cell https://committees.parliament.uk/publications/1280/documents/11300/default/
28 Clause 45(6), and implicit in wording 'seeks to comply' in cl 45(5)
29 Clause 20(3); 30(3)
30 Clause 20(4)-(5); 30(4)-(5)
31 Clause 20(6); 30(6)
32 Clause 88(4)(a)
33 Clause 88(7)
34 Clause 88(6)
35 Clause 112 et seq
36 Clause 110

to comply with one such.[37] A requirement to take steps can be immediate.[38] OFCOM can back requirements (with seemingly the exception of immediate requirements) up in court, including through an injunction.

There are specific provisions relating to "proactive technology"[39], a new concept introduced by the Bill.[40] The significance of these provisions is not that OFCOM can now do something that it could not under the draft Bill - its ability to specify steps there was relatively unconstrained[41] - but rather that there have been constraints introduced, presumably in the service of the right to privacy or confidentiality, as well as differentiating between content harmful to adults and the other two categories of harmful content.

As mentioned above, there is now an explicit confirmation decision for risk assessments, the central part of the regime.[42] This allows OFCOM to take action if it thinks a service provider has not satisfied its duty on the basis that risks that OFCOM has identified are not effectively mitigated or managed. So it does not seem to target directly the process of risk assessment but seems concerned with risk management. It still seems to be a step forward from the draft Bill. There is also a separate provision dealing with compliance with the children's access assessments, which is also new; having clarity around these foundational elements of the system is an improvement.

From confirmation decisions flow measures to enforce. These include fining (with more detail now than in the draft) and business disruption measures. OFCOM has to apply to a court for business disruption measures. The starting point seems to be service restriction orders[43] – these apply to people who provide services ("ancillary service") (such as banking, search visibility, advertising[44]) in relation to a failure on the part of a regulated service provider to comply with an enforceable requirement; time-limited Interim Service Restriction Orders[45] – for urgent, high-harm cases; Access Restriction Orders, where Service Restriction Orders are not likely to work and are mainly directed at ISPs and App Stores; and Interim, time-limited Access Restriction Orders (clause 126).

## Is the Bill future proofed?

The regime seeks to be dynamic and there are specific measures in place at key points that seek to ensure it remains up to date. For instance, risk assessments are required to be kept up to date[46], including when there is a significant change in their service. Companies are (for illegal content and content harmful to children) required to take note of issues raised with them that are harmful. As we have written at Carnegie UK, it seems likely that a user-to-user regime would encompass the metaverse - the Government should clarify that this is the case. There are two sets of questions arising: do the lists of measures in the safety duties seem appropriate for the dynamic, real-time environment that the metaverse is likely to be? Insofar as the harms in relation to which service providers must take action are described by reference to current criminal offences, is there a danger that those offences will themselves become outdated (consider a sexual assault on an avatar)? This more generally illustrates the difficulties of relying

37 Clause 113
38 Clause 113(5)
39 Defined at clause 184 as "content moderation technology", "user profiling technology" and "behaviour identification technology" each with their own sub-definition.
40 Clause 116
41 See Clause 83 draft OSB
42 Clause 114
43 Clause 123
44 Clause 123(12)
45 Clause 124
46 For instance, Clause 8 (3) and Clause 83 (9) Ofcom's market risk assessment

on lists of types of content in terms of keeping the regime relevant.

There is a process for reviewing whether new categories of content harmful to children and/or adults are needed.[47] However, this in the end requires a statutory instrument, which is cumbersome. In TV, radio advertising and cinema regulation, the regulator - based on research - can move independently of the executive to combat novel harms which fall within the more general category of harmful content (rather than a more detailed list of specific issues as here) it was tasked with tackling.

For harmful topics that are well understood, and new permutations of those, the system will stay up to date. There are questions though about its ability to cope with novel topic areas due to the constraints the government has chosen to put on OFCOM.

## Working with other regulators

The requirement to work with other regulators was dropped between the response to the White Paper and the draft Bill. Despite a significant suite of recommendations on this matter from the Joint Committee, there is no mention at all of any requirement to cooperate or co-ordinate with other regulatory bodies - though the ICO has been added in a few places as a statutory consultee. While OFCOM may not need a new "co-designation" power[48], it will need to work with other regulators - for example, the FCA in relation to OFCOM's powers in enforcing the new duty on fraudulent advertising. This has consequences not just for - e.g. sharing of information between the regulator or ensuring clear lines of responsibility and cooperation in relation to evidence-gathering, horizon-scanning or enforcement - but also for upstream policy oversight. For example, which department holds the ring on the policy oversight and related Ministerial advice on the implementation of the duty on fraudulent ads? DCMS, as the sponsor of OFCOM, or HMT, as the sponsor of the FCA, or Home Office, as the department with the policy responsibility for combatting fraud? From the other end of the process, it is instructive to contemplate what would happen to a case file on an issue as it winds its way between regulators. Are all the powers in place to allow information to cross boundaries and for bureaucracies to resource their teams for cross boundary working?

Curiously, clause 97 makes express provision for OFCOM to work with overseas regulators, but not UK ones.

It would do no harm to set out in the Bill a requirement on OFCOM to define the terms of its relationships with other regulators and the power, if needed, to get them to work effectively together.

## Conclusion

We will return to many of these issues - and more - in blogs and more detailed publications in the coming months and will continue to work in support of civil society partners, policymakers and Parliamentarians as the Online Safety Bill makes its way through Parliament. In the meantime, as always, we welcome feedback on our analysis - where we have got things right and, as importantly, where we may also have got things wrong. Contact us on info@carnegieuk.org.

47 Clause 56

48 See para 223 of the Government response to the Joint Committee: "Ofcom already has the power to co-designate other bodies to carry out separate functions by virtue of Section 69 of the Deregulation and Contracting Out Act 1994, and Section 1(7) of the Communications Act 2003. We are satisfied that these powers are sufficient, should other bodies be required to deliver aspects of the regime. As such, we do not see a need to amend the Online Safety Bill" (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1061446/E02721600_Gov_Resp_to_Online_Safety_Bill_Accessible_v1.0.pdf)