ONLINE SAFETY ACT NETWORK

OFCOM'S ILLEGAL HARMS CONSULTATION: EMERGING CONCERNS

The Online Safety Act Network brings together over 60 civil society organisations, campaigners and advocates with an interest in the implementation of the OSA. You can read more about us here.

Many of these organisations will be submitting individual responses to Ofcom's consultation on the illegal harms duties, reflecting their particular areas of interest or expertise across the broad online harms landscape. But the 23 organisations and experts who have signed this statement share a number of fundamental concerns about Ofcom's approach which, we believe, will greatly reduce the effectiveness of the regulatory regime in the early years of its implementation and limit the protections that online users, particularly the most vulnerable, can expect in the coming years.

While we appreciate the scale of the task that Ofcom has been handed by the complexity of the Act and the work that has gone into this first consultation, we urge a rethink before the first codes of practice are published later this year. The evidence-based measures that are included are welcome but we fear that the strategic choices that Ofcom have made limit the protections for users, baking them into the regime in a way that subsequent iterations of the codes are unlikely to rectify. We set out our concerns in brief below, with more detail in the attached annex. The OSA Network's full response to the consultation will be published shortly.

Strategic choices

Ofcom has made a number of strategic decisions about its approach to the illegal harms duties which are problematic and run counter to what we believe to be the legislative intent of the Act, when passed by Parliament. These decisions are not up for consultation but, in following them through, they

fundamentally affect the choices that Ofcom has made about the measures it recommends in its codes of practice, which in turn weaken the protections for users.

These decisions include:

- **The approach to "illegal content" set out in its Illegal Contents Judgement Guidance:** this does not take account of the Act's overarching "safety-by-design" approach based on obligations related to *systems* (not individual items of content); focuses primarily on identification of criminal conduct, rather than the content associated with a criminal offence; and uses criminal standards of proof, rather than civil standards which would be the norm in a regulatory regime.

- **Burden of proof/weight of evidence**: there are aspects of this issue that are problematic in relation to the illegal contents judgements guidance and there is a separate issue which is the apparently preferential weighting that Ofcom gives to evidence already collected from industry, eg "best practice" from companies, and the undefined threshold it sets for other evidence to meet for inclusion in the codes, which seems very high. The effect of this is to set the bar too low in terms of the measures with which regulated services must comply via the codes and to reinforce the status quo which the legislation was intended to improve. Conversely, measures that Ofcom acknowledge might mitigate harm but which do not meet their (undefined) threshold for evidence are discounted. This does not align with Parliament's expectation of a systemic, risk-based regime, focused on outcomes rather than prescriptive rules.

- **What proportionality means**: Ofcom's approach to proportionality is primarily economic: to avoid imposing costs on companies. While the OSA requires regulated services take a "proportionate" approach to fulfilling their duties, and indeed requires Ofcom to look at resources, Ofcom is also required – among other issues – to look at the severity of harm.

- **The prioritisation of users' freedom of expression above adverse impacts on fundamental rights of others**: amongst other things, this has significant implications for protection of women and those from minoritised groups, for whom targeted online abuse is a means of silencing them. This is especially concerning in light of increased media attention regarding – and the government's recognition of – digital threats to democracy; increases in misogynoir and other forms of online abuse limit democratic participation among those most adversely impacted by online abuse. In the past few years, high-profile racist and misogynistic online attacks on footballers and TV commentators have led to many of those targeted being hounded off platforms.

Implementation specifics

There are a number of other specific decisions on the scope of the codes of practice and the measures recommended within them which also, in our view, significantly limit the likely impact of the measures proposed in this initial consultation. Some of these include:

- **Weak "safety by design" foundations:** a disconnect between the evidence of harm in the risk profiles and the mitigation measures in the codes of practice.

- **Lack of focus on outcomes**: the regime is not outcome-orientated (eg to deliver improved safety) but focused on a prescriptive, tick-box, process-driven approach via the codes.
- **Compliance expectations**: the proposals take at face value evidence from the platforms that they are "doing much of this already" and Ofcom continuously emphasises the proposed measures will not incur any additional costs.
- **Small vs large companies**: there is a significant differentiation in Ofcom's approach to the risk assessment duties and the codes between large companies (7m+ monthly users) and small companies (everything else).
- **Limited improvement in the online safety of children, women, Black women especially and other minoritised groups:** while there are specific measures relating to child sexual abuse material (CSAM), overall, the impact of all the decisions taken by Ofcom above will do little to shift the dial in terms of improving safety for children, women, especially Black women and other minoritised groups.

Consultation process

Finally, there are two aspects of the consultation approach that are a concern.

- **Speed vs comprehensiveness:** the codes are a "first iteration" and will be revised. However, a "lowest common denominator" regime is very likely still to get watered down further.
- **Civil society response:** the size and complexity of the consultation has caused accessibility challenges for under-resourced third-sector organisations, and there is no mechanism for victims to respond.

February 2024

LIST OF ORGANISATIONS & EXPERTS SUPPORTING THIS STATEMENT

Alliance for Countering Crime Online (ACCO)
Barnardo's
CEASE
Clean Up the Internet
5 Rights Foundation
Institute for Strategic Dialogue
Kick it Out
Reset
UCL Digital Speech Lab
Antisemitism Policy Trust
CARE
Center for Countering Digital Hate (CCDH)
End Violence Against Women Coalition
Glitch
Just Algorithms Action Group
Molly Rose Foundation
Samaritans
Suzy Lamplugh Trust

Prof Clare McGlynn KC (Hon)
William Perrin, Carnegie UK Trust
Prof Lorna Woods OBE, University of Essex
Julian Coles

ANNEX

Strategic choices

- **The approach to "illegal content" set out in its Illegal Contents Judgement Guidance:** the guidance focuses primarily on individual items of content and assessing whether they should be taken down – it even refers to the obligation being "to take content down", rather than, as the Act says, to operate a proportionate system designed to have that effect. There are parts of the consultation which interpret this correctly - for example, in the "Overview" document where Ofcom says "A new legal requirement of the Act is for all services to swiftly take down specific illegal content when they become aware of it," the Act's systemic language is ignored in the draft guidance itself. Choices about design should happen before you get the content flowing across them. There is also no consideration of scale - the sheer volume of information that is potentially involved. This then defines the scope of Ofcom's overall illegal harms approach, with an ex-post focus on measures, such as content moderation and take down, which we discuss in more detail below.

  Furthermore, by requiring that a criminal offence has taken place each time content is posted (rather than acknowledging that content which has been deemed illegal remains illegal when shared as it is still connected with the original offence), an unnecessarily limited view of relevant content is baked into the proposals compounded by an approach that sets the standard of proof at the criminal level – at odds with what is a civil regulatory regime. It is also unfortunate that Ofcom has not considered any of the existing non-priority offences, specifically s 127(1) of the Communications Act, which unlike 127(2) has not been repealed. We set out more in this blog post. Whilst organisations such as Glitch support justice for victims and survivors of online abuse, the criminal justice system has proven time and time again that it is not fit to deliver that justice, particularly for Black women. By focusing on tackling online abuse through the criminal justice system, we are therefore creating avenues for justice that are exclusionary of the groups most impacted by it. Instead, tech companies should be held accountable to the harms their design choices allow.

- **Burden of proof/weight of evidence**: there are aspects of this issue that are problematic in relation to the illegal contents judgements guidance and there is a separate issue which is the apparently preferential weighting that Ofcom gives to evidence already collected from industry, eg "best practice" from companies, and the undefined threshold it sets for other evidence to meet for inclusion in the codes, which seems very high. Much store is set by the amount of evidence already collected to support eg the risk management approach, and on the "best practice" already provided by platforms to justify the approach. But this sets the bar too low. Conversely, where there is weak or limited evidence relating to the potential for a particular measure to address a particular outcome, this is given as a reason not to include it within the codes until more evidence comes available. This approach reinforces the status quo, setting a "lowest common denominator" approach to a process-driven regime, rather than one that is

focused on the outcomes described in the Act.

- **What proportionality means**: Ofcom's approach to proportionality is primarily economic: to avoid imposing costs on companies. While the OSA requires regulated services take a "proportionate" approach to fulfilling their duties, and recognises that the size and capacity of the provider is relevant, the Act also specifies that levels of risk and nature and severity of harm are relevant. This focus on costs and resources to tech companies is not balanced by a parallel consideration of the cost and resource associated with the prevalence of harms to users (for example, on the criminal justice system or on delivering support services for victims) and the wider impacts on society (particularly, for example, in relation to women and girls and minority groups, or on elections and the democratic process). The assumption in the proportionality analysis that "small" means "less harm" due to less reach is also an issue, particularly given that it downplays the severe harm that can occur to minoritised groups on targeted, small sites - which we discuss further below.

- **The prioritisation of users' freedom of expression above adverse impacts on fundamental rights of others**: Ofcom's approach to human rights considers only the rights of users (as speakers) and has principally focused on their freedom of expression. In doing so, it has not really considered the nature of the speech (which the Convention court does take into account), nor provided evidence that speech in some instances would be chilled – it has rather hypothesised a rather theoretical concern. It has not considered the rights of other users and non-users that require steps to be taken against rights infringing harms – and where the infringement of a right has been recognised in the judgements of the European Court, or the opinion of UN Special Rapporteurs. This means that any balancing exercise is skewed towards not taking action for fear of inconveniencing users (who could well be infringing the rights of others) and companies. This is especially concerning in light of increased media attention regarding – and the government's recognition of – digital threats to democracy: increases in misogynoir and other forms of online abuse limit democratic participation among those most adversely impacted by online abuse. . In the past few years, high-profile racist and misogynistic online attacks on footballers and TV commentators have led to many of those targeted being hounded off platforms. We set out more in this blog post.

Implementation specifics

- **Weak "safety by design" foundations:** There is a disconnect between the evidence of harm presented in the risk profiles and the mitigation for those harms proposed in the codes of practice: the former identifies the significant role of systemic issues, design and functionalities (many of them being a factor in multiple different risks/offences) but the latter does not adequately address these aspects in terms of mitigation, focusing mainly on post-hoc actions (content moderation, takedown, complaints etc). A specific example here is livestreaming, which is listed in volume 2 of the consultation documents as one of the four functionalities that "stand out" as posing particular risks (along with end-to-end encryption, pseudonymity and

anonymity and recommender systems) but there are no measures required relating to livestream in volume 4 (codes of practice). Ofcom does not consider the possibility that features which are demonstrably harmful but for which no ex post measures exist should be redesigned, or withdrawn, which is unlikely to be the case in other industrial sectors. An absence of any requirement for safety testing of new products or services is also a factor here as is the fact that there is too much emphasis on detection rather than upstream prevention measures. There is no joining up with other existing initiatives which may lead to improved safety and better outcomes for users - such as the recent proposals from Ofcom on media literacy by design principles or the 2021 "Principles of Safer Platform Design" published by DCMS.

- **Lack of focus on outcomes**: the overall message that emerges from the consultation documents, compounded by the weak "safety by design" foundations, is that the regime is not outcome-orientated (eg to deliver improved safety) but focused on processes that companies need to follow in a tick-box way to comply. The obligation to measure the result of mitigation measures and improve them in risk assessments is undermined by the decision to take a process-driven approach, listing lowest common denominator measures to follow, in the codes.

- **Compliance expectations**: The governance and risk assessment proposals take at face value evidence from the platforms that they are "doing much of this already" and therefore the suggested measures will not incur any costs. This does not take account of the costs to society ("polluter pays"), the impact of business models nor the principle that the regulatory approach should focus on the overall objective (making services safer), rather than a tick-box process for compliance.

- **Small vs large companies**: Parliamentary assurances were given that all companies would have to comply with the illegal duties, but there is a significant differentiation in Ofcom's approach to the risk assessment duties and the codes between large companies (7m+ monthly users) and small companies (everything else). Many of the measures – including board or governance oversight of risk management – only apply to "large" companies, and the threshold for this is too high; for example, it will not catch Roblox or Fortnite. While Ofcom have defined "risky" in relation to particular harms by reference to the existence of specific risk factors - and some of the measures apply to small "risky" companies - the weak risk assessment obligations on small companies potentially mean that those with low or multi-risks will not identify themselves as such. Volume 4 (on the codes) also very specifically says that small companies are exempt from following many of the measures to avoid incurring costs or stifling innovation. This lets many potentially harmful and/or risky small companies off the hook, such as - for example, collector sites for image-based sexual abuse or suicide fora, where the risk is high and the targeted nature makes them very problematic. There is also an issue that is unaddressed regarding the relationship between small and large companies: bad actors develop their practice on smaller platforms so their actions are sophisticated enough to bypass the T&Cs of the big tech platforms.

Weighing public safety from illegal harms (grooming, terrorism, intimate image abuse etc) against costs to private companies - some of which may be worth billions of pounds - does not align with Parliamentary or public expectations on what the regulatory framework will achieve. It also wrong to suggest that safety stifles innovation or competition or indeed that, if a small service is unsafe, it is wrong of the regulator to take measures that might hinder its ability to compete. It would be helpful to understand the legal basis upon which Ofcom has determined it is acceptable to use size as a factor in determining safety standards online, particularly in light of the changes late in the Parliamentary passage of the Bill to allow category 1 designation to apply to companies on the basis of either size or risk which brought small, high-harm services into scope of additional duties).

- **Children are overlooked:** in addition to ensuring that particular risks of harms to children that are identified in volume 2 are effectively covered by mitigations in volume 4, Ofcom should ensure there is a greater focus on children throughout. For example, whilst Terms of Service must be child friendly, this requirement is missing from reporting mechanisms even though volume 2 highlights the fact that children experience increased barriers to reporting which could be overcome through design.

- **Limited improvement in the online safety of children and Black women:** there are a number of new criminal offences proposed that address online VAWG, which are welcome. But the impact of all the strategic and policy decisions taken by Ofcom above will do little to shift the dial in terms of their overall safety online. Until the Government conceded on Baroness Morgan's amendment in the latter stages of the Bill's Parliamentary passage, the Government promised that the new offences would go a long way to improving protections for women and girls and that a separate code of practice was unnecessary. The opposite is true – and Ofcom's guidance on VAWG, which was the Government's concession, will not be consulted on for at least another year.

  Further, as evidenced in Glitch's Digital Misogynoir Report, Black women continue to be disproportionately impacted by online abuse, and the online abuse directed towards Black women is interconnected with other forms of hate online, like antisemitism, Islamophobia and transphobia. While the OSA accounts for intersectionality, it remains to be see how those vulnerable to harm because of their intersectional identities will be protected; nor is it clear how Ofcom plans to develop and implement frameworks for ensuring Black women – and many other multiply-marginalised communities – do not fall through regulatory and legal gaps. More detail on these concerns will be provided in due course in a detailed joint submission from 15 experts and organisations working to address online Violence Against Women and Girls (VAWG).

Some points on the process

- **Speed vs comprehensiveness:** this has been used by Ofcom officials in a number of discussions as a reason for not covering particular aspects (eg things added late in the passage of the Bill) with a promise that the illegal content codes will be a "first iteration" and will be revised. What are the timescales for this? What evidence will be needed for these next iterations? The risk here is that the regime gets embedded in this "lowest common denominator" form and watered down, via company lobbying, Judicial Reviews etc, from there, rather than being built on stronger foundations and continuously improved.

- **Approach to the consultation:** the size of the consultation and the resources required to engage with it in a meaningful way is a barrier to the engagement of a huge swathe of civil society, where resources and capacity is limited, despite the role many of these organisations played in the development of the Online Safety Act - for example, in campaigning to ensure that there were protections for women and girls included. While Ofcom, as they frequently point out, have provided a very succinct and clear summary of their proposals, the detail is in the 1700-odd pages and – as we have set out above – is often woven into the fabric of the multiple volumes and annex, rather than set out clearly as a proposal for consideration. Moreover, there does not seem to have been a mechanism set up with appropriate safeguards and protections for individual survivors, victims or those with lived experience of the harms set out in this consultation to contribute in a way that could enable Ofcom to judge the effectiveness of their proposals against the reality of harm and its impact on users.