Online Safety Act Network

[redacted]

Consultation Title	Consultation: New priority offences - serious self-harm and cyberflashing
Organisation	Online Safety Act Network

[redacted]

[redacted]

Responder Type

organisation

Do you agree with the proposal to categorise suicide and self-harm as a single kind of illegal harm? Please provide the underlying arguments and evidence that supports your views.

We broadly agree with the statement that “the evidence suggests that suicide and self-harm manifest online in a similar way and many pieces of research do not distinguish between the two harms. The evidence also suggests that the risk factors most strongly associated with the two offences are the same.” However, substantial nuance will be required in the way these offences are communicated to social media companies. Both have challenges with definition and therefore identification, but this particularly applies to self-harm content.

Ofcom’s documentation will need to refer to a specific definition of self-harm. The legal definition will not be sufficient. It includes encouragement to cumulative acts that result in self-harm, which will create challenges in identifying individual pieces of content. It also does not specify the type of act causing harm, only that it reaches the threshold of grievous bodily harm (or serious injury in Scotland). While some acts - such as cutting and burning one’s own body - are clearly defined, others acts, like swallowing foreign objects may not be obvious to platforms and it is unclear whether encouragement of eating disorders is included. Consideration will also need to be given to deliberate acts of self-injury that may have a cultural, religious or sexual context.

There are a number of legitimate reasons for posting material online that could be seen to – or have the effect of – encouraging serious self-harm. Again, the legal definition will not be sufficient on its own. It is important that content related to self-harm minimisation or reduction isn’t caught up in the combination of the offences: for example, harm minimisation guides, whether they are medically correct material from responsible organisations or ineffective methods from individuals could be considered as normalising self-harm and, by stating it can be carried out in a safe or nearly safe way, encouraging it; personal stories that describe how someone self-harms and the relief they feel from it may be criminalised; artistic depictions of self-harm (art therapy is evidence-based and recommended for self-harm, carried out by, for example, the British Association of Art Therapists)

There is also a consideration relating to the services that are in scope: smaller platforms that otherwise would be deemed multi-risk for suicide and self-harm content would now not be in scope.

Do you agree with the risk factors proposed for the risk profiles for suicide and self-harm? Please provide the underlying arguments and evidence that supports your views.

Ofcom proposes adding the following risk factors to the existing risk profile for suicide and applying them to both suicide and self harm: 4b. Group messaging, 5b. Direct messaging, 5d. Commenting on content. This extends consideration of the functionalities related to risks from suicide content while also applying them to self-harm content, which is welcome.

However, Ofcom will need to consider which services are likely to have a significant risk of incidence of the suicide and self-harm material. Whilst some of these risk factors will be the same, some will not. For example, the most dangerous suicide material is instructional, and likely to be found in small forums. It is less clear what the most dangerous self-harm material involves, but it may be more likely to involve behaviours like bullying/abuse through one-to-one communications, or small group chats on larger platforms. If suicide and self-harm are considered as one issue, the significant differences between the sorts of behaviour that would lead to encouragement of either would need to be made explicit.

Do you agree with the risk factors proposed for the Risk Profiles for cyberflashing? Please provide the underlying arguments and evidence that supports your views.

Cyberflashing, Ofcom proposes the following risk factors and it is based on these factors that services will need to update their risk assessments accordingly: “1a. Social media services, 1b. Messaging services, 3a. User profiles, 3b. Anonymous user profiles or users without accounts, 4a. User connections, 5b. Direct messaging and 5e. Posting images or videos. It is recommended that Ofcom consider the other services and actions that could be added to this list, so that it is similar to the intimate image abuse and extreme pornography offences (which are listed in table 91, p 42-44 in the draft updated risk assessment guidance and risk profiles): eg discussion forums and chat rooms, reporting or forwarding content.

Do you have any views on our proposed changes to the suicide and self-harm section of the Register of Risks? Please provide the underlying arguments and evidence that supports your views.

Serious self-harm: It is welcome to see that Ofcom has acted on feedback from the Samaritans on the previous version, including examples of search terms or autocompletions that could exacerbate risk on services potentially draws attention to, or risks exposing users to, harmful or illegal content. (3.19) It is also welcome that the updates to the risk register include evidence relating to eating disorder content which could be a factor in encouraging self-harm (15.24-15.28) It is also helpful to see that Ofcom flags the risk of contagion, eg services that enable users to build online communities constitute a risk factor (coded language and account bios) and also the “competitive” aspect to communities focused on eating disorders, eg via user comments. The risk registers are also updated to observe the fact that anonymous profiles are a concern if adults are exploiting vulnerable children (NB link to terrorism); and that direct messaging and group messaging can be linked to Com Networks. Evidence is also cited from Molly Rose Foundation to support the risk of binge-watching behaviour through recommender systems, including autocomplete on search terms/recommended search terms.

We note the specific section dealing with chatbots (Risk Register, 15.89 – 15.91) but provide commentary elsewhere in our response on the incomplete approach to generative AI that Ofcom is

proposing.

Do you have any views on our proposed changes to the cyberflashing section of the Register of Risks? Please provide the underlying arguments and evidence that supports your views.

Cyberflashing: we recommended that Ofcom's Draft Updated Register of Risks (p61) is updated to include the most recent survey showing increasing prevalence of cyberflashing, particularly among young girls. The Risk Register document has been updated to include a 2024 survey, but there is a 2026 survey showing that 45% of 18-24 year old women have been cyberflashed. (See details here: bbc.co.uk/news/articles/cg7v75kv3dxo; https://cwasu.org/resource/non-contact-sexual-offences/)

Do you have any views on our proposed updates to the self-harm section of the ICJG? Please provide the underlying arguments and evidence that support your views.

No response provided.

Do you have any views on our proposed updates to the cyberflashing section of the ICJG? Please provide the underlying arguments and evidence that support your views.

We provide the following commentary from Prof Clare McGlynn of Durham University, which is also available on our website.

While Ofcom guidance says it is reasonable to infer that sending unsolicited genital images constitutes the criminal offence and therefore action should be taken to prevent this risk, it provides an exception for any service which claims that sending unsolicited dick pics is 'commonly accepted', such as dating sites for gay men. There is no such defence or exception in the cyberflashing criminal offence. Indeed, sending unsolicited dick pics is against the terms of service of most dating sites. It is not at all clear why Ofcom continues to provide this exemption to the law and excuse for services not to act.

State of Mind

1.93 It will be reasonable for service providers to infer the required intent or recklessness where a user sends content depicting genitalia, unless:

a) there is evidence of consent from the user(s) receiving the photograph or film or

b) it is posted on a service where it is a commonly accepted part of the culture to send and receive intimate images without prior agreement.

This is the same as the existing guidance. As a general statement, this is positive, as it says it is reasonable to assume that sending genitalia images contravenes the legislation and therefore action should be taken, unless other circumstances.

However, the suggestion that this presumption can be reversed, en masse, where the image is posted on a service where it is 'commonly accepted part of the culture' is deeply problematic. The target is dating sites, particularly those used by men seeking to have sex with men. It is commonly suggested that sending unsolicited dick pics is 'normal' and 'acceptable'.

However, this is problematic for two main reasons.

First, sending unsolicited dick pics is actually against the terms of service of most such dating sites. They prohibit it as it is non-consensual conduct of a sexual nature, as well as often constituting harassment. See Grindr for example.

Secondly, there are many who are on these sites who do not consider this ‘normal’ and ‘acceptable’, therefore challenging these assumptions and supposed norms.

This raises questions:

Why is Ofcom giving carte blanche to sites to decide for themselves if there is a ‘commonly accepted culture’ on their service of sending unsolicited dick pics, meaning they don’t have to take proactive action, when this is exactly the type of conduct that the law was introduced to prohibit?

Why is Ofcom allowing a situation to continue whereby Terms of Service prohibit this conduct, but this is not enforced?

Background to this issue being raised in debates on cyberflashing offence:

When the cyberflashing offence was being debated, the then Government Minister Chris Philp argued against a consent-based law on the basis that: ‘in an online environment where the exchange of those kind of images was generally considered fine, that might criminalise those kinds of exchanges.’ (Evidence to Women and Equalities Committee).

I responded in my evidence that, in fact, many dating websites prohibit this conduct, giving the example of Grindr whose guidelines state: ‘Ask for consent before sending nude photos or explicit messages, and don’t post these things publicly.’

Problems with Ofcom’s Usage Examples (page 22, after 10.1)

Ofcom repeats these assumptions that sending unsolicited dick pics can be acceptable in the examples it gives:

A photograph or video of a person’s genitals, sent over a messaging app or other communications service to all users in a ‘meeting’ or group chat where it would not be commonly expected to appear.

A photograph or video of a person’s genitals, posted to a public comments section on a service where this is not a commonly accepted part of the service’s culture.

This suggests it is acceptable to send an unsolicited genital image in a messaging app or similar if it could be said that this is ‘commonly expected’. Similarly, the second example.

It should be made clear that sending unsolicited genital images should be presumptively prohibited, and steps should be taken to prevent it. Only if there is clear consent, or lack of the requisite intent, should it then be allowed.

The focus should be on preventing the harms of cyberflashing.

Do you have any views on our proposed updates to the Record Keeping and Review Guidance? Please provide the underlying arguments and evidence that support your views.

No.

Do you have any comments on our proposed approach to updating the Codes, in light of the creation of the two new priority offences? Please provide the underlying arguments and evidence that supports your views.

Ofcom note that “As part of this consultation, we are not proposing any changes to the substance of the existing measures in the Codes. The steps we recommend providers take to mitigate the risk of illegal harms remain the same”. This is disappointing, particularly due to the problems we have

identified with the limited scope of the measures in the codes, the impact of the safe harbour and the fact that there is no requirement for services to address all the risks they may have identified in their risk assessments.

So, to underline the potentially narrow impact of these changes, the only current code of practice measures where illegal harm relating to suicide is specifically mentioned (and therefore will apply to this new combined offence) are ICU E1 (collection of safety metrics on platform testing of content recommender systems), ICU J1 (user blocking and muting) and ICU J2 (disabling comments). Other illegal harms code measures will also apply but - as we have repeatedly flagged in our commentary on Ofcom's implementation - that is the minimum with which services have to comply under the Act, even if their risk assessment might identify multiple other risks relating to these two offences from features and functionalities on their service.

With regard to cyberflashing, we welcome the fact that the measure on blocking and muting will now be applied to services at risk of cyberflashing, as this may bring in services otherwise not in scope of that measure. But this is an example of an ex-post measure: if a (female) user has already had an unsolicited dick pic sent to her, the harm is done; a more proactive measure would be to block the content first (ie not allowing the sending of dick pics unless the image was blurred, or prohibit it entirely to new users or recipients in DMs unless they accept. We recommend that Ofcom consider this as a matter of urgency in their next code of practice update.

Do you have any views on the proposed changes to the application of the Governance and Accountability measures or the impacts we have identified? Please provide the underlying arguments and evidence that supports your views.

It is unclear in Ofcom's proposals whether additional services come into scope of the governance and accountability measures. In the section on impact on users, Ofcom notes at para 3.45 "our proposals may bring a small number of additional services into scope of measures ICU A3/ICS A3, ICU A5/ICS A5, ICU A6/ICS A6 and ICU A7/ICS A7" (eg ICU A3 - set out a written statement of responsibilities for senior managers responsible for illegal-harm risk management; A4 - maintaining an internal monitoring and assurance function that reviews the effectiveness of illegal■harm controls and reports findings to a governance body or audit committee; A5-7 - tracking and monitoring illegal content and training staff) BUT, at 3.48 in the section on impact on services, Ofcom says "Although unlikely, it is possible that our proposals could bring a small number of new services in scope" - their reasoning here being that services already at risk of other forms of illegal content (eg harassment, coercion etc) will also be at risk of cyberflashing, and that there is an obvious link between suicide content and self-harm content. This is repeated in the sections on content moderation (para 3.59 and 3.62); reporting (3.70 and 3.72), recommender systems (3.84 and 3.86). It would be helpful if Ofcom could clarify which assessment is correct: will new services come into scope, or not?

Specifically in relation to cyberflashing, Ofcom's judgement in this section on governance is that: "For example, the risk factors associated with cyberflashing are similar to those associated with harassment, stalking and controlling or coercive behaviour. This overlap of risk factors means that it is unlikely that a previously single-risk service would now become multi-risk because it is medium or high risk for cyberflashing and/or self-harm." (para 3.50)

However, this is not entirely correct: the cyberflashing offence has two motive requirements. The first is intention to cause alarm, distress etc, which does fit with harassment, stalking, coercive

behaviour. But the second motive is sexual gratification (and recklessness as to causing distress). This is therefore different to harassment, stalking etc and is closer to the extreme pornography offences. Adding in “extreme pornography” as a cross-reference for the risk factors is recommended.

Do you have any views on the proposed changes to the application of the Content and Search Moderation measures or the impacts we have identified? Please provide the underlying arguments and evidence that supports your views.

No.

Do you have any views on the proposed changes to the application of the Reporting and Complaints measures or the impacts we have identified? Please provide the underlying arguments and evidence that supports your views.

No.

Do you have any views on the proposed changes to the application of the Recommender Systems measures or the impacts we have identified? Please provide the underlying arguments and evidence that supports your views.

No.

Do you have any views on the proposed changes to the application of the Search Design, Functionalities and User Controls measures or the impacts we have identified? Please provide the underlying arguments and evidence that supports your views.

No.

Do you have any views on the proposed changes to the application of the User Controls measures or the impacts we have identified? Please provide the underlying arguments and evidence that supports your views.

No.

Do you agree with our assessment of the additional impacts of our proposals in respect of the measures in the June 2025 Additional Safety? Please provide the underlying arguments and evidence that supports your views.

No response provided.

Do you agree with our assessment of the combined impacts of our proposals set out in the Combined Impact Assessment? Please provide the underlying arguments and evidence that supports your views.

No response provided.

Do you have any other feedback on our proposals? Please provide the underlying arguments and evidence that supports your views

Handling generative AI

Ofcom notes that they have updated the register of risks to reflect “the role generative AI can play in creating and sharing this kind of content. However, we have not at this stage included generative AI

as a risk factor in the Risk Profiles. This is because we consider a more comprehensive update to the Register of Risks and Risk Profiles, considering various kinds of priority illegal harm, would be needed to appropriately include generative AI as a risk factor. Such a comprehensive update would significantly expand the scope of this consultation and delay our ability to consult on the proposals in this consultation." This is a notable admission but leaves a big question mark as to when this "comprehensive update" will emerge to take account of the risks associated with generative AI - Ofcom only notes that, given the changes likely to flow from changes to the Crime and Policing Bill, they will need to do further updates and "when undertaking future updates, we will consider how to capture evidence of risks associated with generative AI in the regulatory products". However, if it is known now that there are extensive enough risks to require this update, then not doing so quickly - or providing interim guidance as to what steps services should be taking into account before that update - would appear to be an oversight. It would be good to understand whether the reluctance to act here is a result of a lack of resourcing (being unable to undertake this comprehensive update) or a lack of available evidence (if so, then clarification as to how Ofcom is going to collect this evidence and on what timescales would be welcome).

More welcome is the addition - described at 3.18 of the consultation document - that the risk register now "includes research on AI chatbots, which highlights potential risks from some generative AI systems that may produce harmful self-harm or suicide-related outputs. Such AI systems may also produce distressing responses for a minority of users seeking informal mental-health support. It also includes additional evidence on messaging functionalities, in particular direct and group messaging. This evidence highlights how these functionalities can be used to pressure or encourage vulnerable users into self-harm or suicide."

Scope

Ofcom notes that "we have strong evidence suggesting dating sites may have a higher risk of cyberflashing. We have not included dating sites in the Risk Profiles because they do not routinely appear in the evidence for other kinds of illegal harm and the Risk Profiles are intended to provide a simplified overview of the risk factors most frequently associated with risk of illegal harm." The decision to exclude dating services is deeply problematic. There is an assumption made (by Ofcom and the previous government when debating the cyberflashing offence) that unsolicited dick pics on dating services used by men seeking sex with men are 'acceptable' and normal. But in fact this contravenes those services' own terms of service. It is not at all clear, therefore, why Ofcom excludes dating services. (Or in the Illegal Contents Judgments Guidance gives them an effective exemption from the presumption that sending of dick pics is unlawful, unless evidence to the contrary. See further commentary provided by Prof Clare McGlynn elsewhere in this response.) Further, there is no exemption in the criminal law for sending an unsolicited dick pic on a dating website so for Ofcom to exclude it in their guidance runs counter to the legal position.

This - again - suggests an unnecessarily self-limiting approach from Ofcom and one which runs counter to Government intent when it brought cyberflashing into the priority offences in the Act: the Government press release stated: 'Dating apps and social media platforms now have to take proactive steps to prevent this vile content before users see it'. It did not say, as Ofcom does here, that it's fine for dating apps to carry on as usual and allow people to send unsolicited dick pics.

Will this make a difference?

It is disappointing to see that the regulator doesn't think so. The provisional conclusion says:

3.143 The changes proposed in this consultation will assist providers to comply with their duties under the Act and ensure users are better protected from cyberflashing and self-harm content. To the extent that our proposals result in additional costs to providers, we consider that such costs are minimal and outweighed by the benefits of our proposals. Overall, we do not consider these proposals would have any significant additional impacts on users or providers beyond those set out in the December 2024 Statement and the April 2025 Statement.

3.144 Having considered the benefits of the proposals and the impacts on service providers and on users, we provisionally consider our proposals to be proportionate.

Again, this runs counter to the Government announcement: 'Platforms will be required to take proactive steps to prevent this vile content from appearing in the first place, not just react after the harm is done. Tech firms will now face some of the strongest requirements under the Online Safety Act as ‘cyberflashing’ becomes a Priority Offence. Companies could tackle these images for example by using automated systems that pre-emptively detect and hide the image, implementing moderation tools or stricter content policies.'

It is also disappointing to see the lack of urgency. Having set out the process that needs to be followed (consultation, consideration of responses, statement in response to the consultation, publication of revised documents, (some of the) revised documents to be laid in Parliament, once approved they are in force), Ofcom says:

4.7 “Once the updated ICJG, Register of Risks, and Risk Assessment Guidance and Risk Profiles are published, we expect providers to update their illegal content risk assessments to assess the risks of the new kinds of illegal harm arising on their service. We recommend providers do this as soon as practical after the statement is published.

4.8 Regarding the measures that services should consider to mitigate the risk of harm from these new priority offences, providers should consider the updated Codes as soon as they complete the parliamentary process and come into force.

So, in summary, the proposals won’t make a difference but anyway service providers shouldn’t do anything about them until they are fully in force.

Process

While the introduction of each new priority offences clearly requires a process to consider the necessary changes to each of the relevant regulatory documents, there are a number of other such changes coming through (such as on deepfakes/NCII). With each new change, Ofcom will presumably run a consultation along the same lines, requiring engagement from industry as well as civil society groups representing the interests of those who are intended to be protected by the changes.

This consultation is complicated yet, as we note above, Ofcom is of the view it will have little impact on regulated services. Presumably, the communications that Ofcom will send out when the final versions of the risk assessment guidance or codes of practice are published will also say this.

When - as with this consultation - all the effort leads to very little minimal substantive change for services' compliance, the risks of non-engagement each time this process is undertaken significant. This is likely to be particularly the case for smaller services and community fora, leading to a wider risk in relation to their appetite for compliance more generally with the regime.

This is also a serious issue for the Government, given its tendency to trumpet the introduction of a priority offence as a significant move to protect users: if at the end of this resource-intensive process, its regulator is telling all interested parties that nothing has changed, then what is the point of the regulator?