

Response to Ofcom Consultation: Protecting children from harms online

Introduction

Samaritans is the UK and Ireland's largest suicide prevention charity. We answer a call for help every ten seconds and in 2023 Samaritans volunteers spent more than 900,000 hours responding to calls for help.

Over the last four years we have developed a hub of excellence in suicide prevention and the online environment with the aim of minimising access to harmful content and maximising opportunities for support. Our Online Excellence Programme includes industry guidelines for responding to self-harm and suicide content, an advisory service for sites and platforms offering advice on responding to self-harm and suicide content, a research programme exploring what makes self-harm and suicide content harmful and for whom, and a hub of resources helping people to stay safe online. We welcome the opportunity to respond to this consultation and commend Ofcom for moving rapidly following passage of the Online Safety Act.

We are concerned that Ofcom's proposals for protecting children from harms online are insufficient. While there are several strengths in these proposals, many of the concerns we raised in our submission to the illegal harms consultation earlier this year remain unaddressed. As we have emphasised in our discussions with Ofcom during the passage of the Online Safety Act, and in our responses to previous consultations, online suicide and self-harm content is a complex and nuanced issue. It requires a sensitive approach that reduces harm while preserving the valuable support and connection that people experiencing suicidality and self-harm find online. Unfortunately, this nuance was missing in the illegal harms consultation, and we are disappointed that it still has not been appropriately addressed. Ofcom needs to revise the guidance on illegal content to better reflect the necessary nuance and care in addressing suicide and self-harm content. Samaritans influenced every stage of the Online Safety Act to ensure that harmful suicide content was properly addressed in online safety regulation, but we cannot support the codes in their current form.

Our response to this consultation begins with an introductory section detailing our overall thoughts on the consultation, as well as a discussion of key themes. We have then presented our specific comments on each volume of the consultation documents below. We have serious concerns with Ofcom's approach to this consultation, and in turn the implementation of the Online Safety Act. In our Illegal Harms consultation response, we expressed our concerns regarding the capacity of civil society and lived experience to engage with the consultation. The extensive length and inaccessible structure of the consultation will have had an exclusionary impact on many individuals and organisations who would have otherwise submitted responses. We once again welcome the informal engagement with civil society, such as attending a meeting of the Online Safety Network, and through webinars. However, there will be a clear discrepancy in the content

which is considered in formal channels. This approach will have tipped the balance away from those who are most likely to have encountered real world harms and towards well-resourced organisations with the most capacity to respond. We strongly recommend that Ofcom supplement this consultation process with lived experience and re-considers its approach to future consultations.

Illegal harms consultation.

In our response to the illegal harms consultation, we were concerned that the proposals from Ofcom would produce a weak Online Safety environment. However, more concerningly, we felt that it would be difficult for the situation to be improved with subsequent edits or amendments. Ofcom have made clear that they are ‘currently carefully considering and analysing the responses’ received to the illegal harms consultation. Whilst we are glad that our contributions to the previous consultation are being considered, it introduces a challenge. Many of the comments that follow in this consultation response are mirrored in our response to the illegal harms consultation. It is disappointing that we do not know how extensive Ofcom’s revisions will be as a result of their analysis of the previous consultation. We accept the desire to introduce the legislation quickly, however this approach will weaken the evidence received by Ofcom. It would have been beneficial to know the extent to which the codes would change.

We expressed in our previous response significant concern regarding the language used by Ofcom to discuss suicide and self-harm. It is crucial to talk about suicide responsibly and sensitively. For example, the previous consultation used the phrase ‘commit suicide’ which dates back to when the act of suicide itself was a criminal offence. This phrase is rightly viewed by many people in the suicide prevention community as offensive and judgemental. We are pleased that Ofcom has improved its language use when discussing suicide.

Risks to all ages

The government has suggested that their stated aim for these codes is to deliver the ‘strongest protections for children.’¹ We do not agree that the proposals within the consultation documents would deliver this, and we have set out our reasons for this below.

Furthermore, we would like to re-iterate our disappointment that the Act does not require platforms to address other harmful suicide and self-harm content for adults. Whilst Samaritans supports the Government’s commitment to protecting children online, susceptibility to harm from suicide and self-harm content does not end when people reach the age of 18. Suicide and self-harm content affects people of all ages: between 2010-2020, 728 (8%) suicides of patients under the care of mental health services in the UK were related to internet use, an average of 73 deaths per year. 75% of these were people aged 25-64.¹

Samaritans consider the types of suicide and self-harm content that is legal but unequivocally harmful includes:

¹ *The National Confidential Inquiry into Suicide and Safety in Mental Health. Annual Report: UK patient and general population data, 2010-2020. 2023. University of Manchester.*

- Information, depictions, instructions, and advice on methods of self-harm and suicide
- Content that portrays self-harm and suicide as positive or desirable
- Graphic descriptions or depictions of self-harm and suicide

The Government has specified that it is instead up to individual adults to protect themselves from harmful suicide and self-harm content through new user empowerment duties. Yet, a study from the University of Bristol found that participants with severe suicidal thoughts actively used the internet to research an effective method, and often found clear suggestions.² Furthermore, the lack of protections for adults is out of step with what the public wants. Three-quarters (75%) agree that tech companies should be required by law to prevent harmful suicide and self-harm content being shown to users of all ages.³

Size of service as a risk factor

We have expressed during this consultation response that we are concerned with Ofcom's decision to consider the size of the service, rather than the level of risk.

During the passage of the Online Safety Act, there was a significant strength of feeling in Parliament that the harm caused online was not only occurring on large platforms but smaller ones too, and that the Bill needed to reflect this. As a result, Schedule 11 of the Act now allows the Secretary of the State to determine which providers should be in Category 1 based on functionality (and other 'characteristics') rather than requiring that they be of a certain size. This would allow a limited number of small but exceptionally dangerous forums to be regulated to the fullest extent possible. These include forums that explicitly share detailed or instructional information about methods of suicide or self-harm.

We were therefore deeply concerned to see that Ofcom, in the published advice to the previous Secretary of State on categorisation, explicitly recommended *not* using this power to address these extremely dangerous sites.⁴

It is unclear why this approach has been taken, given the evidence suggesting that some of the most harmful suicide and self-harm content is hosted on smaller platforms that would not be covered by large parts of this guidance.

In volume 4, Ofcom suggests a similar proposal to one found in the illegal harms consultation, relating to responsibilities for senior members of staff. These responsibilities include risk assessments, considerations on processing data and other important duties. However, similarly to the illegal harms consultation, this statement is only recommended for large or multi-risk platforms. These measures may come with costs, however Ofcom should be focusing on the risk to children rather than operating costs of industry. Although, the Government's own impact assessment makes reference to the fact that building safety design processes would reduce regulatory compliance costs for smaller platforms.⁴

² Biddle L, Derges J, Goldsmith C, Donovan JL, Gunnell D (2018) Using the internet for suicide related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital. *PLoS ONE* 13(5): e0197712. <https://doi.org/10.1371/journal.pone.0197712>

³<https://www.samaritans.org/news/government-is-failing-the-public-with-online-safety-bill-says-samaritans/>

⁴ Ofcom (2024) [Categorisation: Advice Submitted to the Secretary of State](#)

Safety by design

In our previous submission, we supported the Online Safety Act Network's argument that the Act would not deliver on its overall objective to deliver services which are 'safe by design.' This duty is set out in the children's safety codes in Section 12 (8). After parliamentary debate, the new 'clause 1' was added stating that it was government's intention that 'a main outcome of the legislation is that services must be safe by design. For example, providers must choose and design their functionalities so as to limit the risk of harm to users.'⁵ We were pleased that Ofcom confirmed their commitment to safety by design in their approach document published in November.⁶ However, we are disappointed that this intention is not reflected in the proposals put forward in this consultation. This response will assess the safety by design offer, however in particular, we are concerned with age gating, recommender systems and content-focused measures.

Joint response with Online Safety Act Network

We support the joint response from the Online Safety Act Network, calling for a rethink in approach to several areas to ensure that the online safety regime actually meets its objectives.⁷ In particular, we would emphasise

- The need to embed and incentivise a 'safety by design' approach to root out harmful content from the outset; and
- The need to take a democratic and accessible approach to consultation going forward that enables everyone to engage and respond on an equal footing, including individuals with lived experience of online harms.
- The need for duties in the children's code duties to apply equally to all services regardless of company size. Ofcom must make risk the primary aspect, rather than size.
- The need for additional measures relating to product safety testing and safety by design to be added to the draft codes, which would place the responsibility on services (of all sizes) to take measures that are proportionate to them to address the risk of harm that is identified in their risk assessment.
- The need for effective governance and risk assessment proposals. Ofcom should seek further advice on how to carry out risk assessments which are conscious of safe design and operation of products and services.

This response focuses on our specific expertise on suicide and self-harm content.

Comments on Volume 2

We are concerned that age assurance measures are not a safety by design measure. Ofcom's proposals lack a necessary complexity and nuance. It is right that services which are entirely inappropriate for children are required to ban children from accessing their site. Suicide and self-harm content online can be dangerous to children. However, Ofcom must consider the potential benefits the internet provides to some individuals accessing support. Online spaces provide individuals experiencing self-harm and suicidal feelings with an opportunity to connect with others who have similar experiences, share feelings without judgement and also to document feelings and journeys of recovery.⁸ It could be incredibly harmful if children are banned from platforms which take the approach of age gating certain material. Similarly, an obligation on age verification means that the quality of service providers' risk assessments are secondary. If they have prevented

children accessing the platform, they aren't required to keep monitoring risks. This is problematic as it fails to address the reality that some children will still access the content. Additionally, it then does not appropriately address the risk faced by adults who may encounter harmful content.

We are concerned that Ofcom has not introduced measures which would consider the risk of harm to different age groups of children. It is reasonable that content which is appropriate for a 17-year-old, might not be appropriate for an 8-year-old. And similarly, Ofcom must not restrict access to content that might be helpful for 17-year-olds, because it could be harmful to children under 10.

We are also worried that the age gating measures do not appropriately address the risk posed by small platforms. Volume 4, 12:50 'However, for the avoidance of doubt, we expect that any service with more than 1 million (or between 100,000 and 1 million) monthly UK child users would need a range of robust evidence to demonstrate that it does not in fact pose high (or medium) risk of harm to children in respect of a given kind of content.' This provision leaves open the possibility that small, but very harmful, platforms do not need to suitably prevent access by children.

Ofcom's use of 'highly effective' with regard to age assurance technologies is problematic. We agree with Ofcom's decision not to recommend certain tools. However, there needs to be an agreed method of measuring the effectiveness of any tool. Ofcom does provide criteria explaining different aspects of effectiveness; however, this is insufficient. Ofcom should specify a metric for each of their criteria that would measure the implementation and enforcement of the method. The stated reason for not specifying criteria, is to 'allow space for important innovation in the safety tech sector.' Yet, setting a benchmark for safety would not prevent innovation beyond that point.

Comments on Volume 3

Section 7 Draft Children's Register of risk

The consultation states that: 'the evidence suggests that online communities formed around experiences of mental health concerns present a significant risk of exacerbating these concerns. (7.2.26)' We would question this conclusion. Whilst it may be true in some communities due to poor moderation, there are many such communities ran by charity organisations with excellent content moderation and whose moderators are trained in managing this content and supporting vulnerable users. The distinction needs to be made here to ensure that Ofcom does not suggest these charity run communities can cause harm. Furthermore, this must not lead to communities on which vulnerable people rely, being taken away.

It is highlighted in the consultation (7.2.48) that the size of a platform's user base is a risk factor for harm. Whilst we agree with this point, it needs to be made clear that it is not just the number of registered users but the number of people who are able to access the content on the site, for example through search features. Considering only the number of registered users neglects a significant risk of harm to others.

We agree that user generated content searching can be a gateway to accessing harmful content. It is also important to acknowledge the role of autocomplete searches here which may be suggesting harmful searches to users. This should be introduced to the guidance in 7.2.99.

We appreciate the need for the category of non-designated content and the recognition that further research is needed to understand how 'depressive' or 'mental health' content should be defined. Whilst we understand the risks, there is also a huge risk that codes relating to this content will lead

to the majority of content being removed. If young people are not able to reach out for support, share their own experiences or connect with others with similar experiences it will lead to further stigmatisation and could impact on mental health outcomes if they are censored for experiencing poor mental health. Instead, the bigger concern is how online services present this content to its users and ensure this content is not actively promoted. It is essential that the codes prevent the content being promoted at scale, rather than being removed completely by the platform.

Whilst there is a wealth of research presented within this section, all of it explores harms and none of the benefits to users are presented. It is crucial this is detailed in the guidance. The Government must also consider the risk of preventing access to content which can be immensely supportive for vulnerable children. Self-harm and suicidal thinking can be difficult to talk about directly, and these topics continue to carry stigma. In moments of crisis, people find online engagement supportive, feeling a sense of community to seek support from.⁹ Almost three-quarters (73%) of Samaritans' lived experience panel respondents believe that online spaces can be helpful for individuals experiencing self-harm and suicidal feelings, and it is imperative that this support route is not taken away, but rather that it is made safe.¹⁰

Section 8 (Draft guidance on Content Harmful to Children

Section 8 discusses recovery content. Our research and work with people with lived experience also highlights that this content can cause harm if it is seen to promote 'toxic positivity'. This is the idea that everyone can recover and is in a better space. People with lived experience want authentic accounts of managing self-harm and suicide on a daily basis and the ups and downs of managing it, rather than unrealistic portrayals that can make them feel worse.

In table 8.3.2, 'Descriptions and examples of content which encourages, promotes or provides instructions for suicide,' it is highlighted that harmful content can include descriptions of novel methods of harm. It is unclear how platforms are expected to identify what constitutes novel or emerging. Some of the examples provided in the table are specific in their format e.g. 'a video recommending specific places where individuals have ended their life'. It may be helpful to make these examples broader to avoid portraying to platforms that it is the format of the content that causes the harm rather than the content itself. For example, a post in text form with the same level of description as a video could be equally harmful. Suicide pacts are also described as people encouraging other to take their lives. It is important to note that in many cases this is not malicious encouragement but people connecting with others who have made plans to end their life to avoid doing it on their own. We also feel that there is a section missing within the table relating to content that glamorises, sensationalises or normalises self-harm and suicide. This might include portrayal of suicide as positive way to end distress.

In table 8.3.3, which lists content not considered to be harmful, there is no mention of people sharing their daily lived experience of self-harm and suicide other than positive accounts and recovery. It is important to clarify whether it is being suggested that all other types of lived experience content are therefore classed as harmful. It is crucial that this is clarified. Academic articles are also mentioned here as not causing harm but we know some contain harmful details that can be used to inform a suicide attempt e.g. method information and lethal dosages. The nature of the content here must be the overriding factor in determining the level of harm rather than its format. This is also true for statistics around self-harm and suicide. It is stated that these are not

harmful but they could be if they are inaccurate or portray self-harm and suicide in a sensationalist way.

Summary

We are pleased that Volume 3 uses evidence to justify its approach to online harm. The guidance broadly covers the risks and types of suicide and self-harm content that children may interact with. However, we are concerned with the lack of nuance in the guidance, and the lack of consideration that is given to the benefits of the internet. Online communities can play a significant role in supporting vulnerable children who otherwise may not be able to access support. We are urging Ofcom to alter the guidance to avoid the risk of supportive platforms such as charity groups being removed.

Comments on Volume 4

Section 11 (Governance and Accountability)

The governance and risk assessment proposals are closely linked to the approach in the illegal harms consultation. We maintain our concern that Ofcom's proposals are balanced too far towards corporate governance literature, and too far away from adopting a safety by design approach. Focusing on corporate governance literature has meant that the consultation document is focused too heavily on what platforms and businesses are already doing. Yet, Ofcom has not made clear whether it believes these existing measures are effective or enough to deliver platforms duties under the OSA. The proposed Governance and Accountability measures in Volume 4 follow a familiar pattern of Ofcom not appropriately addressing the risk and harm presented by smaller platforms. Most of the proposed measures are only required by large services. This is not enough and does not deal with the reality that a large proportion of the most dangerous suicide and self-harm content is hosted on sites that would not be covered by these accountability measures.

Section 12 (Children's risk assessment guidance)

Effective risk assessments are crucial to delivering a safer Online environment, and the success of the OSA. Again, we are concerned that the proposals are based on 'best practice' which has been produced based on corporate governance and reputational risk literature. We are disappointed that Ofcom has not completed any qualitative assessment as to whether this is effective or sufficient.

The risk assessment guidance is presented as a tick box exercise. The list of things that platforms must consider are listed as tasks to complete, not outcomes to aim for or improvements for the services. There is no requirement for safety by design interventions. The guidance allows services to record that they have done something, but not what the outcome or result of that action is.

We are concerned that in 11.140, Ofcom confuses horizon scanning with capturing evidence of new/emerging harms after that have already happened. This approach would risk harm, and it is unacceptable for platforms to wait for a death by suicide to investigate potential risks. Risk assessments must be forward-looking and seek to identify new dangers facing children online. 11.147 also states that platforms must have a 'mechanism to notice new trends.' However, there is no related governance or measures in the codes about how a company might collect information through this mechanism. Furthermore, it is unclear why small, single-risk services are exempt from this tracking.

We are recommending that Ofcom reconsiders its approach to risk assessment. They should undergo an extensive review of literature and seek advice from the many experts available on how to deliver a risk assessment which addresses the risk to children, not just to the reputation of platforms.

Comments on Volume 5

Section 13 (Children's safety codes) + Section 14 (Developing the Children's safety codes)

Suicide and self-harm content is listed as Primary Priority Content (PPC) and will be subject to the strictest measures. This is described as 'content which encourages, promotes or provides instructions for suicide or encourages, promotes or provides instructions for an act of deliberate self-injury.' It is right that children are prevented from viewing the most harmful suicide and self-harm content. Of course, everyone, including adults, should be protected from this content.

We would welcome further guidance on suicide and self-harm content that recognises the complexity of the content. There is a risk that if suicide and self-harm content is not defined appropriately, platforms will categorise supportive and beneficial content as PPC. 73% of Samaritans' lived experience panel respondents believe that online spaces can be helpful for individuals experiencing self-harm and suicidal feelings, and that it is imperative that this support route is not taken away.¹¹ The internet provides access to a vital source of immediate support in times of crisis, which people have described as a lifeline.¹² It is vital that efforts to protect children from the most harmful online content do not cause harm by way of restricting safe support networks.

Section 16 (Content Moderation U2U)

It is important that content is moderated in an appropriate way which does not risk causing harm to users. We support the proposal that action may be needed to address gaps in moderators' understanding of specific harms. This applies to suicide and self-harm content, where providers should be ensuring that moderation staff understand the nuance around self-harm and suicide language, provide users with personalised responses, and quickly identify and react to emerging trends. We hope that this measure includes recognition of the sources of third-party expertise and support that providers can draw upon to support this area, including Samaritans' Online Excellence Programme.

People with lived experience of suicide and self-harm often use the internet as a way of sharing their stories of recovery and accessing a community. However, in many instances they have had content removed due to the presence of self-harm scars or other messaging that has been deemed harmful. This can have an intensely damaging impact on the user that posts the content, and it is vital that moderation is done sensitively. As part of Samaritans research, over half (52%) of respondents thought that refinement of the way posts are censored (ie, different policies for healed self-harm scars compared to graphic images) would have a positive impact on content. Participants stressed the need for a balance between censorship and allowing people to talk about their experiences.¹³

The guidance is not currently clear enough. Careful moderation is needed from people trained in the subject matter. Otherwise, there is a risk that moderation could cause additional harm to people with

lived experience of suicide and self-harm. Where platforms do need to remove illegal content that is posted by people with lived experience, it is crucial for providers to take a compassionate approach aimed at minimising any distress caused to the user. The tone of the communication explaining why the content has been removed should be sensitive and avoid negative language, explain how to re-post safely and where to find support.

Section 18 (User reporting and complaints)

We hear from people with lived experience of Online harms that they encounter significant barriers when trying to report or complain about content. For new regulatory frameworks to be effective, they need to be accompanied by easy to understand and robust processes for individuals to promptly report concerns about suicide and self-harm content, and mechanisms to raise wider issues that might require changes to practice across all platforms.

Section 20 (recommender systems)

We are pleased with some of the measures relating to recommender systems, however we are concerned that the focus is mostly on the content that passes through the system, rather than the development of the system. Furthermore, the consultation is narrow in its understanding of the relationship between recommender systems and the business of content creation.

Ofcom's guidance suggests that the role of the new online safety regime is not to restrict or prohibit the use of recommender systems, but to get services to put in place: 'safeguards which allow users to enjoy the benefits they bring, while managing the risks appropriately.' The guidance does not adequately explain what is meant by 'safeguards.' There is a risk that these safeguards would only apply, or be implemented, after the recommender system has presented an issue. If so, this would not meet the safety by design quality that the OSA aims to implement

The consultation document makes passing reference to businesses models in relation to risk assessments and risk profiles. However, this is insufficient. There is no consideration or introduction of measures to mitigate the commercial incentives for content creation. This does not address the significant issue posed by clickbait farms and harmful influencers. Content is being used by an increasing number of people as a way to make money, this leads to concerning practices in relation to recommender systems and monetisation.

In many instances, these recommender systems are pushing harmful suicide and self-harm content, 83% of participants reported that they had seen self-harm and suicide content on social media even though they had not searched for it (for example through recommended content on TikTok's 'for you' page and Instagram's 'explore' page.¹⁴ The majority of survey respondents (88%) wanted to have more control over the content they see, as self-harm content can still be easily found by people trying to avoid it.¹⁵ Ofcom must consider taking a tougher stance on the design of recommender systems, and asking platforms to consider the impacts of the business models of content creation.

Ofcom should also consider the potential benefits that a safety by design approach to recommender systems could have. These systems could be used by providers to promote and increase the likelihood of users finding helpful content. It should be evaluated whether codes of practice could do more to prompt providers to think through safety by design features. This would

aid in creating an online environment which benefits the individuals accessing content, rather than the business models of platforms and creators.