

Response to Ofcom Consultation: Protecting people from illegal harms online

Introduction

Samaritans is the UK and Ireland's largest suicide prevention charity. We respond to a call for help every ten seconds and in 2022 Samaritans volunteers spent almost one million 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.

However, we have serious concerns about the way that Ofcom is recommending that suicide and self-harm content be dealt with. As you will know from discussions that we have had with you during the passage of the Online Safety Act, and our responses to previous consultations, this is a complex and nuanced area that requires a sensitive approach aimed at reducing harm while also preserving the valuable support and connection that people experiencing suicidality and self-harm find online. This nuance and – crucially – understanding of the perspective of people who are engaging with this content – is absent from the proposals. Ofcom needs to revise the wording of the Guidance on Judgement for Illegal Content to better reflect the nuance and care needed in addressing suicide and self-harm content. Samaritans influenced at 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 draft codes in their current form.

We also have serious concerns about the capacity of civil society organisations to meaningfully engage with this consultation given its unprecedented length and a structure that does not easily enable interested parties to ascertain which parts are most pertinent to focus upon. We welcome the ways that Ofcom have engaged informally with civil society, including attending a meeting of the Online Safety Act Network. However, there will be a clear imbalance in what is formally 'counted' through consultation responses with well-resourced industry organisations best able to engage in detail and respond. The impact of this is that the scales are tipped away from the people who have encountered real world harms as a result of shortfalls in online safety – both in terms of civil society being able to advocate on behalf of the people that they work with, but also in terms of Ofcom being able to hear directly from people with lived experience. We strongly recommend that Ofcom supplement this consultation process with lived experience engagement.

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. In particular, we would emphasise:

- the need to redress an undue emphasis on economic ‘burden’ on online businesses: the declared purpose of the Online Safety Act is to make the UK the safest place in the world to go online but this simply isn’t sufficiently reflected in the proposals;
- ensuring that the bar is not set so low as to not initiate any real change compared to current practice, particularly amongst smaller providers that host some of the most harmful suicide and self-harm content;
- 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.

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

Comments on Volume 2

While the evidence review of suicide and self-harm content in chapter 6D is broadly accurate, the language used throughout is jarring. It’s crucial to talk about suicide responsibly and sensitively: one example of this shortfall is inclusion of 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 judgmental.

While Ofcom is of course a communications regulator rather than a suicide prevention organisation, it is disappointing that the advice we have given during the genesis of Ofcom’s work on online safety about how to talk about suicide responsibly has not been taken into account. Greater care needs to be taken in any future communication on this topic to maintain faith from people with lived experience of online harms related to suicide and self-harm. We would be happy to discuss this further and there is also relevant guidance in our Media Guidelines.¹

It is important to delineate between the two separate offences of encouraging or assisting suicide, and encouraging or assisting serious self-harm as these are distinct offences in law. For example, it is inaccurate to state in the summary for chapter 6D that ‘this offence takes place when an individual intentionally encourages or assists a person to self-harm or end their life’ as this isn’t a single offence.

It is also important to note that people who are experiencing suicidal thoughts are not necessarily experiencing mental illness: two thirds of people who die by suicide are not in touch with mental health services in the year before they die. It would therefore be better to use broader language around people who may be posting content that are in vulnerable or difficult circumstances, rather than simply talking about mental health problems.

¹ [Media Guidelines FINAL.pdf \(samaritans.org\)](#)

The consultation document includes information in paragraph 6D.17 about the types of online content that Samaritans considers harmful. It is important to note that this covers a range of content, only some of which is illegal under either of the offences being discussed in this chapter. This full range of content may therefore be most useful for Ofcom to include in the forthcoming measures to protect children which straddle both illegal and legal harmful content.

We are pleased to see inclusion of the fact that discussion forums and chat room services can have positive benefits as well as potential areas where harmful content appears, given the importance of online support for people experiencing suicidality and self-harm. While we agree that features such as video streaming and recommender systems can increase the risk of harmful content, they could also be used by providers to promote and increase the likelihood of users finding helpful content. Ofcom should consider whether codes of practice could do more to prompt providers to think through ‘safety by design’ features such as this: a point covered more fully in the response from the Online Safety Act Network.

Comments on Volume 4

We do not agree with the rationale given for only applying additional measures only to smaller services that have at least two illegal harms present (11.44). No reason is given for the presumption that the extent of harm will be lower if only a single harm is present: if that single harm is harmful suicide content that is a high risk, priority harm as defined through the Online Safety Act. The consultation document also states that if there is a single risk that ‘risk is more likely to be well understood across the organisation’. No evidence is given for this statement and, crucially, there is no evidence that a risk being well understood translates into any action being taken to mitigate it. On the contrary, there is evidence on public record of harms being well known about but systematically ignored by providers.

We support the proposal that action may be needed to address gaps in moderation staff’s understanding of specific harms (13.158). 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 welcome the measure to include crisis prevention information on search services where users are making general searches related to suicide and queries seeking specific, practical or instructive information on suicide methods (22.83). It is important that prevention information on methods only appears when searches are suicide-specific given that some methods have other, common uses in everyday life: care must be taken to avoid raising awareness of suicide methods when this wasn’t the intention behind a search. As the consultation document notes, existing practices of search services indicate that this is already happening in practice, although it is important to note that this is not consistent across all search services.

However, the consultation document also states that this measure is already in place amongst both large general search services and several smaller search services (22.88). It therefore doesn’t make sense to apply this measure only to large search services: rather than focusing on applying this measure to smaller search services being a barrier to ‘entry and competition’ for new small search services, the emphasis should be on not disincentivising those small search services that already

have this measure in place from continuing with it. This is an example of an overly economic approach to proportionality, an overarching issue identified by the Online Safety Act Network, rather than the starting point of reducing harm. As the Online Safety Act Network response states, it is 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.

Comments on Annex 5

It is important that considerations of severity of risk of harm are not simply based on reach, as we know that the seriousness of harmful suicide and self-harm content means that there can be a devastating impact even if it only reaches a small number of people. The risk level table (Table 6) should therefore be amended to include severity of potential harm as a criteria for inclusion in 'high impact': at the moment this wording remains based largely on reach, despite the fact that risk profiles elsewhere in the consultation documents and referenced in this Annex suggest wider demographics and vulnerabilities of users should be taken into account.

Retrospective analysis of incidents of harm (table 10) should include users, especially people with lived experience of the harm involved in the incident. It is crucial to connect online safety to real world harms, and to position learning lessons and putting mitigations in place as a societal responsibility rather than simply an internal corporate process. Where appropriate and safe to do so, providers should also be required to share their learning more widely, including with civil society organisations.

Risk assessment reviews may be as involved and time consuming as original assessments so the wording in A5.123 around an expectation that reviews and updates 'should not be as burdensome as carrying out a new risk assessment' is unhelpful. The emphasis should be on providers ensuring that their risk assessment is reviewed at least every 12 months and is robust, irrespective of the 'burden': as the consultation document states, this is a duty that providers need to meet.

[X]

Comments on Annex 7

The requirements for providers to track evidence of new and increasing illegal content (3E) should include a requirement for the governance body to share this information with relevant trusted flaggers or other statutory bodies. In particular, reducing proliferation of less well known or new methods of suicide, and how to access them, is a key part of suicide prevention. Monitoring new and emerging intelligence is one of the commitments in the cross-government 2023 Suicide Prevention Strategy. It may be most appropriate for the accountable people within providers to share intelligence with the NCA as the most pertinent trusted flagger. However, care must be taken not to inadvertently publicise any information on these highly sensitive topics, given that the main aim is to stop wider proliferation of harmful content.

The recommendation for providers to create a code of conduct to protect users from illegal harm (3F) should include an expectation from Ofcom that this will include how to support users posting illegal content who may themselves be in vulnerable circumstances or distress. Providers should also be prompted to consider whether this is an opportunity to clarify and promote the support available to staff who are exposed to harmful content through their roles. There is further information and guidance about both of these areas in Samaritans guidelines for sites and platforms hosting user-generated content.²

The requirement for providers to ensure people working in content moderation receive training and materials (4F) that enable them to moderate content in accordance with illegal content policies should include recognition of the sources of third-party expertise and support that providers can draw upon to support this area, including Samaritans Online Excellence Programme.

It is unclear why the requirements for safety metrics for on-platform testing of recommender systems (8A) only apply to services at medium or high risk of at least two priority illegal harms. As noted above, this rationale does not seem to be explained in Volume 4 either. Our experience is that harmful suicide and self-harm content often appears on services dedicated to discussion of this topic alone, meaning that they would not be included in these safety metric requirements.

Comments on Annex 8

The requirement that platforms have a policy for establishing a dedicated reporting channel for trusted flaggers (3E) should include an approach to how intelligence of new and emerging harms arising from harmful suicide or self-harm content will be shared with relevant agencies. This may best sit with the NCA as the most pertinent trusted flagger, but care must be taken not to inadvertently publicise any information on these highly sensitive topics, given that the main concern is to stop wider proliferation of harmful content.

[3<]

Comments on Annex 10

The definitions in this draft code of illegal content that reaches the threshold for 'encouraging or assisting suicide' do not correspond to our understanding of what is covered by this offence. While there is limited case law in this area, we know that a guiding principle of the Online Safety Act is that content which is illegal offline should be illegal online. The proposals from Ofcom risk having the opposite effect in going beyond existing parameters of communications that meet the threshold for this offence. For example, we know that in many instances 'specific, practical information on suicide methods' is legal in the written word.

From our discussions with other organisations responding to this consultation, we understand that Ofcom may be relying on a civil law rather than a criminal law threshold but it remains unclear how and why Ofcom has drawn the line where it has. This is odd given how much detail and rationale is provided on many other areas covered in the consultation documents. While we would like to see as

² [Online Harms guidelines FINAL 1.pdf \(samaritans.org\)](#)

much harmful content addressed as possible – and it is always a choice open to providers as to how much they remove according to their own guidelines – we believe the bald interpretation of illegal content in the draft code risks opening Ofcom up to challenge on the faultline between illegal harmful content and legal but harmful content that was writ large during parliamentary passage of the Act.

The draft code states that ‘the nature of the post and context around the post’ should be considered to determine intent but without any further information or guidance about what this means in practice – and the difficulties of determining intent when content may be posted by people who are themselves in distress. It therefore feels likely that providers will simply opt to remove all suicide content. This would have a potentially devastating impact as we know that online supportive spaces can be a lifeline for many people and the widespread stigma around suicide and self-harm may mean that these are the only source of support that people rely upon.

This concern is compounded by the guidance in the draft code that method information posted in a space where people are discussing suicidal ideation ‘should be treated as illegal and taken down’. It is entirely possible that people talking about their own lived experience mention methods and this is not automatically illegal content. Nuance and careful moderation is needed from people trained in subject matter expertise. 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.

It is also important to consider the differences between the types of content that are listed here, for example someone seeking to organise a suicide pact online is different to someone encouraging others to take part in a challenge or game to harm themselves. Providers will need to make different judgements for these types of content rather than a blanket assumption that they are illegal.

We also do not support the interpretation of content that meets the threshold for ‘encouraging or assisting serious self-harm’. It is too simplistic to state that posting information about how a person may undertake serious self-harm is equivalent to encouraging or assisting anyone to do so. It is also too simplistic to state that talking positively about self-harm meets the threshold for illegal content. Many supportive spaces may include information about methods that seek to minimise harm, for example, using clean blades, including from charities that support people with self-harm. Self-harm is a complex behaviour that can be part of people’s coping mechanisms so they could also be talking about it positively from their own experiences. As discussed in detail during the parliamentary debate on this issue and in our previous responses to Ofcom, malicious intent is key to determining whether the offence of ‘encouraging or assisting serious self-harm’ has been committed. So malicious intention, context, and considered moderation must always be the heart of the careful judgements that providers make judging any content as illegal. This is not reflected in Ofcom’s proposals

Although some fictional accounts of suicide or self-harm may not meet the threshold of either of these offences, close moderation is needed to ensure that this is the case for relevant content. There may be instances where intent can be inferred.

While these are two separate offences and distinct types of content, it is striking how the suggested interpretation of the offence of encouraging or assisting serious self-harm includes recognition of the fact that someone may be posting about their own experiences but there is no reference to this in terms of the suicide offence. This links back to our wider concerns about an overall lack of compassion and understanding towards people with lived experience in all mentions of suicide throughout the consultation documents.

For these reasons, we do not support Ofcom's illegal contents guidance as currently worded for the offences related to suicide and self-harm and would like to see them substantially revised. There was considerable discussion of the meaning of these offences and where lines could be drawn during the parliamentary passage of the Online Safety Act. Ofcom should refer to this in order to help set the judgements for this content that better reflect the nuances and vulnerabilities involved, and that correspond to existing legal definitions and understanding of these offences.

It may also be helpful for Ofcom to apply its definitions to actual pieces of content to build an in-depth understanding of the complexities in this area and to share these with relevant stakeholders as part of the consultation.