

Your response – Additional terms of service duties

Questions 1 – 5: Terms of service and policy statements

For all respondents

Question 1: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements?

Please submit evidence about what features make terms or policies clear and accessible.

It is essential that provider policies and communications around self-harm and suicide content should be informed by subject matter experts, including people with lived experience. Our work with people with lived experience tells us that providers of online services can enhance the clarity and accessibility of their terms and statements by:

- Ensuring that clear and accessible language is used throughout.
- Ensuring that, as far as possible, harsh or punitive policy language is eradicated or kept to a minimum in relation to self-harm and suicide content to avoid the risk of sounding judgemental or stigmatising. For example, stating that someone is ‘violating content policies’ could increase distress. Instead, a tone of care and warmth should be implemented.
- Make terms of service and policy statements easier to find. In our research with MEL exploring user views of platform messaging around self-harm and suicide, most people did not know where to access platform content policies so relied on finding them through search engines.
- Provide the content in multiple formats to increase accessibility e.g. text and video. Some people we spoke to highlighted the need for graphics and icons to aid information processing.
- Present the policies regularly to ensure that users are reminded about them, rather than just presenting them once at the point of sign up.
- Make the terms of service and policies compulsory to read at sign up.
- Any updates and amendments should be effectively communicated to users in a timely way.
- It is best practice that the clarity and accessibility of terms of service and policy statements should be user tested rather than relying on the provider’s view. Those relating to self-harm and suicide should be tested by people with lived experience to ensure that they are sensitive and accessible, especially to individuals who may be reading the content in high levels of distress. This is a service that we offer through our Samaritans’ Online Excellence Programme.

Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Question 2: How do you think service providers can help users to understand whether action taken by the provider against content (including taking it down or restricting access to it) or action taken to ban or suspend a user would be justified under the terms of service?

In your response to this question please consider and provide any evidence related to the level of detail provided in the terms of service themselves, whether services should provide user support materials to help users understand the terms of service and, if so, what kinds of user support materials they can or should provide.

Service providers should:

- Include clear links to the content policy and how the action taken on the content is covered by the policy
- Include links and information about the appeals process
- Where possible, providers could provide education on how users could post the content again in a safer way to avoid breaking the policy e.g. avoid naming mentions of methods of harm
- Removing someone's content relating to self-harm or suicide can be extremely distressing and potentially cut off avenues for vital online support. In these cases, providers should take a sensitive and empathetic approach and ensure that the user is signposted to support.

For more guidance on removing self-harm and suicide content safely, visit our Industry Guidelines: <https://www.samaritans.org/about-samaritans/research-policy/internet-suicide/guidelines-tech-industry/reducing-access-harmful-content/>

Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Questions 6 – 8: Reporting and complaints processes

For all respondents.

Question 6: What can providers of online services do to enhance the transparency, accessibility, ease of use and users' awareness of their reporting and complaints mechanisms?

In your response to this question, please provide evidence about what features make user reporting and complaints systems effective.

In your response to this question, please provide information relating to (a) – (h) where relevant.

Response:

(a) reporting or complaints routes for registered users, non-registered users and potential complainants (being affected persons who are not users of the service)

Service providers should provide:

• Easy to access and navigate reporting functions. • Clear guidance on reporting content, what happens when you make a report, what will happen and the timescales and the avenues for appeal.
(b) how to ensure that reporting and complaints mechanisms are not misused
(c) the key choices and factors involved in designing these mechanisms
It is vital that providers are considering the specific sensitivities around suicide and self-harm content in designing these mechanisms, including: • User testing the functionality and language with subject matter experts and people lived experience of self-harm and suicide. • Consider using language that is not punitive for suicide and self-harm reporting. E.g. 'Flagging' 'Raising a concern' or 'I'm worried' • Include self-harm and suicide content specific reporting categories • Allow for a conversation to take place between user and moderator – so that reports do not just disappear with little feedback on consequences of the report. • Provide users with alternative mechanisms of feedback for content they would rather not see, but that does not necessarily violate community guidelines e.g. "I'd rather not see content like this" • Use a trusted flagger system, where credible organisations and users with a track record of making responsible and accurate reports can have their reports fast tracked especially where there is content that may provide a risk to public safety – for example, containing information about methods of harm.
(d) how users can or should be supported to report/complain about specific concerns (e.g., other users, certain types of content or, appeal content takedowns or account bans)
• Users should be regularly reminded about the reporting functions and how these are dealt with. Often people don't report content relating to self-harm/suicide because they fear that they will get the person in to trouble. Providers should educate users on how they deal with these reports and how they support users.
(e) how to ensure they are user-friendly and accessible to all users (e.g., disabled users, children)
(f) whether users are informed that their reports are anonymous (e.g., other users will not be informed about who has reported their content or account);
• Users should be able to stay anonymous to the rest of the community when making a report, and providers should make clear that this is an option for them in their communication about reporting mechanisms.

(g) any user support materials that explain how to use the reporting and complaints process and what will happen when users engage with these systems
When a report is being made, users should be given step by step instructions on how to make a report and the process and timescales involved. They should also be able to see how their report is being progressed and any outcomes.
(h) any other information.
For further best practice principles on developing user friendly reporting processes for self-harm and suicide related content, see our Industry Guidelines: https://media.samaritans.org/documents/Implementing_using_user-friendly_report_processes_for_self-harm_and_suicide_co_ybN6MW6.pdf
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Questions 9 – 15: Moderation

For all respondents

Question 9: Could improvements be made to content moderation to deliver more consistent enforcement of terms of service, without unduly restricting user activity? If so, what improvements could be made?

In your response to this question, please provide information relating to (a) –(c) where relevant.

It is helpful to see that there are questions on training and wellbeing support for moderators included in the questions for providers in this consultation. The nuance of suicide and self-harm content means that training and human moderation is needed to make contextual judgements, including about the intent and wellbeing of the person posting the content.

Wellbeing support is essential for those providers whose moderators are likely to come across suicide and self-harm content but our experience is that this content can appear in seemingly completely unrelated or innocuous online spaces so Ofcom could encourage a universal approach to supporting moderator wellbeing. It is unclear why the consultation asks providers about the costs of training and wellbeing support but cost should not be a factor that impedes anyone working in this sector from receiving this vital support.

(a) improvements in terms of user safety and user rights (e.g., freedom of expression), as well as any relevant considerations around potential costs or cost drivers;

• Users should be provided with a range of safety features and mechanisms of feedback for content they wish to avoid but that does not necessarily violate community guidelines e.g. “I’d rather not see content like this” • AI approaches should be adopted if they are cost-effective and offer proportionate solutions to increase the speed and efficiency of moderation. However, companies should never rely solely on AI approaches. Instead, they should be used to prioritise content for human moderation.
(b) evidence of the effectiveness of existing moderation systems including any relevant examples of the accuracy, bias and or effectiveness of specific moderation processes;
(c) any other information.
• All service providers should ensure that moderators are provided thorough training and support options for their wellbeing if they are likely to encounter self-harm and suicide related content. • See our Industry Guidance on supporting staff wellbeing: https://www.samaritans.org/about-samaritans/research-policy/internet-suicide/guidelines-tech-industry/supporting-staff-wellbeing/
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Your response – User empowerment duties

Question 26: Impact of relevant content

For all respondents

Question 26: Can you provide any evidence on whether the impact of relevant content differs between adults and children on user-to-user services?

We are interested in particular in briefings, investigations, transparency reports, media investigations and research papers that provide more evidence.

As noted during the passage of the Online Safety Act, anyone can be at risk from harmful suicide and self-harm content – it is not possible or appropriate to assign higher levels of vulnerability to particular parts of the population, including any presumption that adults are less vulnerable than children. This should be a guiding principle of Ofcom’s approach to implementing and overseeing the user empowerment duties.

Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Questions 29 and 30: Features employed to enable greater control over content

For all respondents

Question 29: What features exist to enable adult users to have greater control over the type of content they encounter?

In your response to this request, please provide information relating to (a) – (d) where relevant.

(a) features offered to users to reduce the likelihood of them encountering content they do not wish to see

Our Industry guidance and work with people with lived experience highlights the need for service providers to provide users with:

- Feedback mechanisms that allow users to indicate whether they do/do not want to see more of this content.
- The ability to set sensitivity levels for different types of content and the ability to set filters, or prioritise content based on characteristics or accounts.
- The ability to set time limits on how much time they spend on the service

(b) features offered to users to alert them to the presence of certain categories of content

Service providers should consider providing:

- Content warning and filters – that can be customised by the user.
- Ability for user creating content to mark it as ‘sensitive’ or ‘triggering’.
- Interstitials used to divert users from content or proceed to view.

(c) features offered to users to enable them to control their interactions with different types of users (e.g., non-verified)

Service providers should consider providing:

- Ability to block or mute users or keywords
- The ability to control the volume of certain types of content that you view e.g. favouriting to make content from that user more prominent
- Control over the level of contact you have from other users e.g. only allow comments or messages from accounts that you like/follow.

(d) whether certain features are particularly valued or of use to users with protected characteristics, or by users likely to be affected by encountering relevant content

The following features are useful for suicide and self-harm content:
--

- | |
|---|
| <ul style="list-style-type: none">• Interrupting harmful browsing of content using reminders or recommending alternative content.• Users should have the ability to reset or turn off recommender systems.• Users should be in control of what data and interactions is being used to recommend content and when e.g. a user who has previously felt suicidal and looked at self-harm and suicide content, may not want to see this content when they are feeling better.• User controls should be prominent and easy to use |
|---|

Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Your response – Access to information about a deceased child's use of a service

Is this response confidential? (if yes, please specify which part(s) are confidential)

Response: yes [CONFIDENTIAL.]
