Your response

Question 1: To assist us in categorising responses, please provide a description of your organisation, service or interest in protection of children online.
Is this a confidential response? (select as appropriate) No
Samaritans is the UK and Ireland's largest suicide prevention charity. We respond to a call for help every ten seconds and, in 2021, Samaritans volunteers spent over one million hours supporting people who called us for help. Over the last three 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.

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?
Is this a confidential response? (select as appropriate) Yes

Question 3: What information do services have about the age of users on different platforms (including children)?
Is this a confidential response? (select as appropriate) No
No comment from Samaritans, this is a direct question for service providers

Question 4: How can services ensure that children cannot access a service, or a part of it?

Is this a confidential response? (select as appropriate)

No

Services need to improve their abilities to accurately verify the age of child users as a foundation to being able to ensure they are not promoting harmful content to young users nor allowing them to search for/encounter it. Our desk-based research of age verification processes showed that they are easily bypassed leading to large numbers of children accessing material of an adult nature. There is a range of other evidence on this point, for example a 2019 report by non-profit Thorn found that 45% of under 13s surveyed were using Facebook daily.

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

Is this a confidential response? (select as appropriate)

No

No comment from Samaritans, this is a direct question for service providers/industry

Question 6: Can you provide any evidence relating to the presence of content that is harmful to children on user-to-user and search services?

Is this a confidential response? (select as appropriate)

No

We are particularly concerned about content online that encourages or assists the suicide of another person. We provided more details in our previous consultation response last September

Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?

Is this a confidential response? (select as appropriate)

No

We have carried out research with the University of Swansea which highlighted the impact of viewing harmful content online. One of the findings of the study was that more than three quarters of people in the survey saw self-harm content online for the first time at age 14 or younger. Individuals with a history of self-harm were more likely to report being 10 years old or younger when they first viewed it, whereas those with no history of self-harm were more likely to have been 25 and over at the time of first viewing it. From this we can see that there is a correlation between viewing harmful content at a young age and future harmful behaviour.

Question 8: How do services currently assess the risk of harm to children in the UK from content that is harmful to them?

Is this a confidential response? (select as appropriate)

No

No comment from Samaritans, this is a direct question for service providers/industry

Question 9: What are the exacerbating risk factors services do or should consider which may have an impact on the risk of harm to children in the UK?

Is this a confidential response? (select as appropriate)

No

Question 9: What are the exacerbating risk factors services do or should consider which may have an impact on the risk of harm to children in the UK?

No comment from Samaritans, this is a direct question for service providers/industry

Question 10: What are the governance, accountability and decision-making structures for child user and platform safety?

Is this a confidential response? (select as appropriate)

No

Samaritans Industry Guidelines recommend that:

- Companies should ensure that accountability for all policies relating to the protection and safety of users is in place at a senior level.
- Clear roles and responsibilities should be assigned to individual roles or teams to ensure that policies are well developed, implemented and reviewed.
- For larger companies this may be dedicated data protection, safeguarding or policy teams.
- For smaller sites and platforms, this may be an individual role or the platform manager who manages these responsibilities.

Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

Is this a confidential response? (select as appropriate)

No

Our industry guidelines recommendations around accessibility suggest that users should be provided with clear and accessible community guidelines about what content is allowed on the site. They should also be given step-by-step information including how to make a report and what action may be taken. This information should be clearly displayed to new users, and existing users should be regularly reminded, empowering them to report any content that concerns them.

Our recent research with people with lived experience of self-harm and suicidal thoughts indicates that whilst there it a good understanding of the purpose of community

Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

guidelines "to keep users safe" very few online users have seen or read them. Many participants said that they would only check out the community guidelines in response to having their own content removed by the platform. Overall, there were very low levels of awareness that community guidelines specifically relating to suicide and self-harm content existed. It is therefore of vital importance that platforms take additional steps to o make community guidelines more visible and accessible for users, so it is clear what content is and is not allowed on their site.

In our research, what resonated most with online users were messages that:

- adopted a human and friendly tone, that avoids authoritative or triggering language
- used simple and directive language without too much text and jargon -guidelines should be easy to navigate and for all users to understand.
- included specific examples of things that are allowed and prohibited. This could be a list of things that users can and can't do
- included clear guidelines on appeals and contact information for a person to speak to about the appeal, rather than an automated help system.
- included clear and straightforward information about what happens if users breach the guidelines.

The best way to ensure that terms of service are clear and keep children and young people safe is to ensure that they meet a universally high standard for people of all ages. Children often bypass age verification checks when using the internet (for example, a report by the non-profit Thorn in 2019 found that 40% of under 13s surveyed were using Instagram) meaning that they will be navigating the exact same online environment as adults. Therefore, the most practical way to ensure to ensure that policies and terms of service are clear and accessible for children is to ensure that they are clear and accessible for users of all ages.

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[1]

Is this a confidential response? (select as appropriate)

No

[1] See A1.2 to A1.3 of the call for evidence for more information on the indicative list of harms to children.

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[1]

Whilst this is a direct question for service providers/industry, Samaritans informal desk based research shows that the classification/categorisation/references to harmful content in existing terms of service, community guidelines and reporting mechanisms are very diverse and the definition of an agreed approach that could be applied across all platforms would benefit users.

Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?

Is this a confidential response? (select as appropriate)

No

Providers should ensure that they are co-designing safety features with young people and sector experts, as well as methods to educate on those features as young users sign up and to reiterate/remind frequently of those features. In addition, they should ensure that access to safety features is as obvious and instant as possible.
Samaritans experience of co-designing has yielded excellent results, for example an Instagram online safety campaign aimed at 18-24 year olds that ran to coincide with Internet Safety day was seen by over 2.6 million 18-24 year olds, over half that demographic in the UK.

Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?

Is this a confidential response? (select as appropriate)

No

Samaritans Industry Guidelines recommend that as a minimum, all sites and platforms hosting user generated content should have a dedicated email address or reporting form that users can access to flag concerns about self-harm and suicide content or user behaviour.
Larger sites and platforms should consider having more sophisticated reporting functions, such as Trusted flagger functions, whereby credible organisations and users with a track record of making responsible and accurate reports can have their reports fast tracked.

They should also consider

Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?

- Prioritised reporting. Where possible, the filtering and prioritisation of reports should be automated and based on urgency to ensure human moderators focus on the most harmful content and users at greatest risk
- Trained content moderators. Moderators should be trained and supported to effectively review and respond to reports in line with the community guidelines
- Processes for identifying key themes and accuracy of reports. Identifying trends in
- reporting will make false or inaccurate reports easier to identify and manage.

We also recommend sites and platforms must provide the following to all users who have reported content about self-harm or suicide:

- Acknowledgement, that their report has been submitted successfully for review.
- Information about what happens next, such as the action that may be taken, eg, reviewed by moderators, removal of content or provision of support to the reported user.
- Action taken. Where appropriate, information should be given to users to assure them harmful content has been removed or addressed.

If a report indicates a user is at risk of imminent harm, the following should be provided to the person making the report:

- Information about contacting emergency services including contact details for the country where the user is based (eg, 999 in the UK).
- Signposting to support for themselves. Reminding them of the importance of looking after their own wellbeing when supporting others online and signposting to relevant information.

Question 15: What actions do or should services take in response to reports or complaints about online content harmful to children (including complaints from children)?

Is this a confidential response? (select as appropriate)

No

See question 14

Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.

Is this a confidential response? (select as appropriate)

No

No comment from Samaritans, this is a direct question for service providers/industry

Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?

Is this a confidential response? (select as appropriate)

No

Our Industry Guidelines make several recommendations of steps that services can take to support the safety and wellbeing of users, including children. The best way to ensure that the safety and wellbeing of children is protected is to ensure that all users, regardless of ages, are protected from harmful content. Steps that sites can take include:

- Ensuring site algorithms don't push harmful self-harm and suicide content towards users. For example, platforms that make suggestions based on previous browsing should disable this functionality for self-harm and suicide content.
- Ensuring that censored content does not appear as suggested content for users
- Blocking harmful site searches, such as those relating to methods of suicide, online suicide challenges and hoaxes, or searches for websites that are known to host harmful content.
- Autocomplete searches turned off for harmful searches such as those relating to methods of harm and associated equipment.
- Using age and sensitivity content warnings, to warn users that content may be distressing as it mentions self-harm or suicide.
- Embedding safety functions, allowing users to have more control over the content that they see. For example, by having more functions to block content by muting words, phrases and hashtags.

Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?

As there is the likelihood that children will be able to bypass age restrictions where they are in place on website, the best way to ensure that these functionalities and features keep children safe is for them to be in place for users of all ages.

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?

Is this a confidential response? (select as appropriate)

No

See previous question

Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?

Is this a confidential response? (select as appropriate)

No

Whilst this is a direct question for service providers/industry we do believe that services must make efforts to stay abreast of ever evolving search terms, tags and keywords that children may use to bypass content restrictions and should collaborate/share knowledge of emerging terms amongst their industry/sector peers.

Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what?

Is this a confidential response? (select as appropriate)

No

Some platforms utilise messaging intercepts whereby direct user to user communication is filtered for harmful keywords before the message is sent, this could be extended to other appropriate platforms.
Instant access to support for children could also be explored, effectively providing direct communication with trained moderators/support within the platform, bypassing slower response reporting mechanisms to flag harmful content and access signposting help.

Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?

Is this a confidential response? (select as appropriate)

No

No comment from Samaritans, this is a direct question for service providers/industry

Question 22: How are human moderators used to identify and assess content that is harmful to children?

Is this a confidential response? (select as appropriate)

No

No comment from Samaritans, this is a direct question for service providers/industry

Question 23: What training and support is or should be provided to moderators?

Is this a confidential response? (select as appropriate)

No

Samaritans industry guidelines recommends
• Moderator shadowing whereby new starters should be paired with more experienced moderators during their induction before moderating independently.
• Refresher training so moderators are kept updated on emerging online trends and trained on how to respond to these.
• Self-care training with guidance on how to look after their wellbeing when interacting with users in distress and viewing harmful content.
• Suicide and self-harm awareness training to increase knowledge and understanding of the area and confidence in responding to users in distress.

Supporting moderators, ideally all sites and platforms should put in place measures to ensure moderators feel fully supported, such as:
• Regular breaks or shorter shifts for moderators viewing high volumes of self-harm or suicide content, particularly if working alone or during unsocial hours.
• Access to managerial support, including when working remotely or out of hours, for support when responding to content or with issues relating to safeguarding.
• Access to a safe, quiet space if distressed by content.
• Opportunities for debriefs at the end of shifts to discuss any difficulties faced or feelings experienced.
• Access to external support, including out of hours, such as an Employee Assistance Programme.
• Reflective practice sessions with a trained counsellor (individual, group or both).
• Peer support training so moderators build the skills to support each other during shifts.
• Provision for secondment opportunities away from harmful content in larger companies with multiple departments.
• Additional days of annual leave

Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?

Is this a confidential response? (select as appropriate)

No

No comment from Samaritans, this is a direct question for service providers/industry

Question 25: In what instances is content that is harmful to children, that is in contravention of terms and conditions, removed from a service or the part of a service that children can access?
Is this a confidential response? (select as appropriate) No
No comment from Samaritans, this is a direct question for service providers/industry

Question 26: What other mitigations do services currently have to protect children from harmful content?
Is this a confidential response? (select as appropriate) No
No comment from Samaritans, this is a direct question for service providers/industry

Question 27: Where children attempt to circumvent mitigations in place on a service, what further systems and processes can a service put in place to protect children?
Is this a confidential response? (select as appropriate) No
The best way to protect children from harmful suicide and self-harm content online is to put the same levels of protection in place for people of all ages. The wording of this question confirms that there are legitimate concerns about children being able to bypass mitigations, which is an inevitable consequence of an approach that seeks to make platforms safer for only some users.

Question 27: Where children attempt to circumvent mitigations in place on a service, what further systems and processes can a service put in place to protect children?

Question 28: Other than those covered above in this document (the call for evidence), are you aware of other measures available for mitigating the risk, and impact of, harm from content that is harmful to children?

Is this a confidential response? (select as appropriate)

No

Samaritans has no direct research in this area but an observation from our work is that the major platforms have and are developing different approaches to child safety features in isolation, rather than pooling resources and collaborating effectively. Child safety is a shared responsibility, not something that should be considered as a competitive advantage. Greater sharing of what works well across education initiatives, parental controls, duration limits, time scheduling, access to instant support would benefit the whole sector and crucially, increase safety for everyone online.