SAMARITANS

Managing self-harm and suicide content online

Guidelines for sites and platforms hosting user-generated content

Introduction

The internet can be an invaluable resource for individuals experiencing self-harm and suicidal feelings. It provides opportunities to access information, find options for support and provides a platform to speak openly about difficult feelings that can be hard to discuss face to face. However, it can also carry potential risks by presenting opportunities to access graphic content, details around methods of harm and content that glorifies or promotes self-harm and suicide. Access to such content can be distressing, triggering and may act to encourage, maintain or exacerbate self-harm and suicidal behaviours. It is, therefore, vital that sites and platforms hosting user-generated content take action to reduce the accessibility of potentially harmful content and put mechanisms in place to maximise support opportunities for vulnerable users.

These guidelines provide a framework of best practice principles to support platforms to manage self-harm and suicide content in a safe and sensitive way, allowing users to access the benefits from the online environment whilst minimising potential for harm.

Who are these guidelines for?

These guidelines are relevant for all sites and platforms hosting user-generated content, regardless of their size or function. This includes social media platforms, image and video sharing platforms, search engines and online forums and communities, all of whom can use these guidelines as appropriate for their functionality.

To promote excellence in suicide prevention in the online environment, sites and platforms should embed these principles in a way that is relevant, using innovative, well considered and evolving approaches. Sites wanting to find more information or get confidential support and advice on how best to embed the principles should contact the **Samaritans' Online Harms Service.**

These guidelines are also relevant for policy makers, moderators of websites, and individuals wanting to ensure sites and platforms are doing what they can to protect users and employees from potentially harmful content.

Understand the impact of self-harm and suicide content online

All sites and platforms must take proactive steps to understand the potential benefits and risks associated with self-harm and suicide content online and how it applies to their site.

User-generated suicide and self-harm content can take many forms, such as:

- lived experience accounts of self-harm and suicide

- depictions of self-harm and suicide, such as artwork, memes and TV stills

- images, videos and live streaming relating to self-harm and suicide

- feeds sharing information or discussing recent events

- memorial pages of people who have died by suicide

- online hoaxes, challenges and suicide pacts.

Understanding the impact of self-harm and suicide content is complex. Whilst there are some types of content that are obviously harmful, other types require more nuanced thinking and judgement on what is appropriate for the platform. What can be helpful for one user can be triggering to others. A user's experience of how harmful content is may also depend on factors such as their current level of distress and the volume of self-harm and suicide content they view.

Understanding the potential risks and benefits to users is critical for sites to:

- identify the potential impacts on their users of engaging with such content

- identify specific risks associated with site functionalities

- inform robust and sensitive approaches to managing the content in a safe way.

See our information sheet Understanding self-harm and suicide content to find out more.

2

Establish clear accountability

It is essential that all sites and platforms recognise self-harm and suicide as **serious harms** and ensure policies are in place to protect users.

Sites must understand and adhere to:

- Data protection and electronic communication laws

- Safeguarding responsibilities

- Responsibilities to report illegal content

- Responsibilities to protect users from online harm

An understanding of these requirements should inform clear and actionable policies for responding to self-harm and suicide content.

Companies should ensure that accountability for all policies relating to the protection of users is held at a **senior level. Clear roles and responsibilities should be assigned** to individual roles or teams ensuring that policies are well developed, implemented and reviewed.

For **larger companies** this may take the form of a dedicated team. For **smaller companies,** this may be an individual role or the platform manager who manages these responsibilities.

See our information sheet Establishing accountability and structure for further guidance.

3

Have a robust policy for addressing self-harm and suicide content

Sites and platforms should develop and implement clear and robust policies for addressing self-harm and suicide content, taking steps to proactively limit the volume of harmful content that is shared and having mechanisms in place to promptly address it if it does appear.

Policies should include:

I   Clear definitions of self-harm and suicide

II Content covered by the policy

III Mechanisms for responding to content covered by the policy

Priority should be given to address content that contains:

- **Promotion or encouragement of self-harm and suicide.** Content that intentionally encourages the suicide or attempted suicide of another person is considered illegal under the 1961 Suicide Act.

- **Graphic descriptions or depictions of acts of self-harm or suicide.**

- **Detailed methods or instructions for self-harm and suicide.**

- **Suicide pacts, challenges and hoaxes.**

- **Mockery or bullying** of people who have self-harmed or attempted, or died by suicide.

Sites and platforms accessed by vulnerable users, such as children and young people, will require stricter content policies.

Policies should be **developed in consultation with subject matter experts** and **regularly reviewed** to reflect the latest evidence, emerging trends and changes in regulation.

Policies should be translated into **accessible community guidelines** for users, explaining what can and can't be posted about self-harm and suicide and why.

See our information sheet on **Developing a content policy** for further guidance.

4

Put user friendly processes in place to report self-harm and suicide content

All sites and platforms should ensure that users accessing their site can easily report concerning content and the behaviour of other users that worries them. This includes:

Accessible reporting information

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.

Accessible reporting processes

For small sites this may be a dedicated email or reporting form. Larger sites may implement more sophisticated reporting functions, such as self-harm and suicide content specific reporting categories and trusted flagger functions, whereby credible organisations and users with a track record of making responsible and accurate reports are able to have their reports fast tracked.

Effective processes for reviewing and responding to user reports

In order to review reports effectively, sites should have trained content moderators and mechanisms in place to prioritise user reports based on risk.

Reports relating to users in need of urgent help should be prioritised for review.

See our information sheet **Implementing user friendly reporting for self-harm and suicide** for further guidance.

5

Effectively moderate all user-generated content, considering both human and AI approaches

All sites and platforms should moderate user-generated content, ensuring that self-harm and suicide content policies are successfully implemented and that users are protected from harm and directed to support.

Human moderation

Sites with low volumes of user-generated content may be able to rely on **human moderation** alone. This can be an effective way of detecting and responding to self-harm and suicide content as moderators can understand the nuance around self-harm and suicide language, provide users with personalised responses, and quickly identify and react to emerging trends. But consideration should be given to the speed at which content can be identified and times of day and night when it is most likely to be posted. All sites implementing human moderation should ensure moderators are provided with high quality training and support.

Artificial intelligence

Platforms hosting higher volumes of user-generated content should complement human moderation with **artificial intelligence** (AI) to prioritise user reports, flag potentially harmful content to be reviewed, and prevent harmful content from being uploaded. AI allows for the assessment and identification of harmful content at scale which can enable early detection and can prevent content from being widely shared.

See our information sheet Implementing effective content moderation for self-harm and suicide for further guidance.

Managing self-harm and suicide content online: guidelines for sites and platforms hosting user-generated content
© Samaritans 2020

6

Reduce access to self-harm and suicide content that could be harmful for users

Sites and platforms should take a proactive approach to reducing potentially harmful self-harm and suicide content on their platform. Example ways of reducing access to such content include:

- **Ensuring site algorithms don't push 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.

- **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.

- **Reviewing autocomplete searches** for terms and phrases relating to self-harm and suicide. Autocomplete searches should be turned off for harmful searches such as those relating to methods of harm and associated equipment.

- **Using age and sensitivity content warnings,** warning 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.

Removing content safely

Sites and platforms should remove content that breaks community guidelines using safe and empathetic approaches. Care should be taken to minimise any distress caused to the user, by ensuring the tone of the communication is sensitive and avoids negative language, and explains why the content has been removed, how to re-post safely and where to find support.

Pausing memberships

If a user repeatedly posts content that breaks community guidelines, companies may decide to pause their membership or close their account to protect other users. This should only be considered as a last resort. Companies should be mindful that this could withdraw a user's vital, and in some cases only, source of support. If pausing a membership, the user should be provided with an explanation of why their membership is being paused, signposts to support and information about how to appeal the decision.

> **See our information sheet Reducing access to harmful self-harm and suicide content online for further guidance.**

7

Take steps to support user wellbeing

Sites and platforms should take steps to support the wellbeing of their users. This includes providing signposting to support, implementing embedded support features and promoting positive content.

Signposting to support

All sites should provide signposts to users in distress, including:

- Emergency and non-emergency services.

- Support that can be accessed 24 hours a day, every day.

- Trusted helplines and services specialising in the provision of support for self-harm and suicide.

- Encouraging users to speak to family, friends or other people they trust.

Embedded support

Sites should consider embedding support features on their site to make it easier for users to access help when needed. Examples include:

- **Information resources**, empowering users to look after their wellbeing and stay safe online.

- **Self-care activities,** such as breathing exercises, mindfulness exercises and distraction techniques.

- **Instant support features,** such as the ability to talk to trained staff or volunteers who have specialist training in responding to self-harm and suicidal behaviour.

Promoting positive content

Sites should establish ways of promoting positive and supportive content, such as wellbeing information, messaging that encourages help-seeking, and stories of hope and recovery.

Sites with search functionality should develop ways of prioritising positive content in search results, which will also serve to bury content that could cause harm, making it more difficult to find.

Sites should also consider promoting additional supportive content in response to emerging trends or spikes in self-harm or suicide related content.

See our information sheet Supporting user wellbeing for further guidance.

8

Communicate sensitively with users in distress, taking a personalised approach where possible

When communicating with individuals who may be experiencing self-harm and suicidal feelings, thoughts and behaviours, sites must use safe and empathetic approaches, remembering that the user could be experiencing high levels of distress. When communicating with users, either through personalised or automated messages, sites should:

- **Use personalised approaches where possible,** to make communication more meaningful, such as using the user's name where appropriate.

- **Be understanding and non-judgemental,** avoid using negative language or making assumptions about the user.

- **Be concise and use clear and accessible language,** because it can be hard to process lots of information when experiencing distress. Language should be culturally appropriate so that it resonates with a diverse audience.

- **Encourage users to reach out for support** from people they trust, such as friends, family, health professionals or support services.

- **Be clear about support available,** both on the site and more widely. Ensure your signposts include services that provide support 24 hours a day.

See our information sheet Communicating with users in distress for further guidance and tips on developing personalised and automated messages.

Managing self-harm and suicide content online: guidelines for sites and platforms hosting user-generated content
© Samaritans 2020

9

Find ways to work collaboratively and demonstrate transparency in approaches to self-harm and suicide content

Collaboration within the sector

To promote online excellence in suicide prevention within the industry, companies should establish ways to share practices, insights and learnings around responding to self-harm and suicide content, and maximising support opportunities for vulnerable users. By working collaboratively, sites and platforms can promote consistency and keep more users safe.

Collaboration with subject matter experts

Site approaches to self-harm and suicide content should be informed by subject matter experts, ensuring they reflect the latest evidence and voices of people with lived experience. Consultation with academics, third sector organisations, health care professionals and individuals with personal experience of self-harm or suicide will help to ensure approaches are robust, evidence based and sensitive.

Transparency about site approaches

It is essential that companies are transparent with users and external audiences about their approaches to self-harm and suicide content, and the effectiveness of these. Whilst maintaining individual and global privacy laws, companies should be open to sharing key insights around the **prevalence** of self-harm and suicide content on the platform. For example, **average number of views** of content that breaks community guidelines, **mechanisms in place to detect and respond** to content, **action taken** to content that breaks community guidelines, and **resource allocated** to responding to self-harm and suicide content.

For larger companies this may take the form of an annual transparency report. Smaller companies may consider making this information available on their website or sharing it when requested by individuals with **legitimate interest,** such as researchers and government agencies.

> **See our information sheet Promoting excellence in online suicide prevention for further guidance.**

Managing self-harm and suicide content online: guidelines for sites and platforms hosting user-generated content

Establish processes to support the wellbeing of staff exposed to self-harm and suicide content

Exposure to self-harm and suicide content, particularly over an extended period, can negatively affect mental wellbeing.

Companies should take measures to protect and promote the wellbeing of all employees and volunteers working with self-harm and suicide content. This should not only include content moderators, but staff from across the organisation, such as trust and safety teams, policy teams, media and public relations. This should include **specialist training, psychological support and regular managerial check-ins as appropriate.**

Staff who need to access potentially harmful suicide and self-harm content should do so only when necessary, and steps should be taken to ensure others do not inadvertently view it. Harmful content should only be viewed when absolutely necessary, and for as short a time as possible.

© Chris O'Donovan Photography/Samaritans

See our information sheet Supporting staff wellbeing for further guidance.

Conclusion

The internet provides a vital space for individuals experiencing self-harm and suicidal behaviour, providing a wealth of opportunities to connect with others, express their feelings and to find support. By implementing these key principles, sites and platforms can create safe spaces, allowing users to access the benefits from the online environment whilst minimising the potential for harm. To promote online excellence in suicide prevention, it is essential that platforms endeavour to implement these key principles and constantly review and improve their practices.

Samaritans' Online Excellence Programme

Samaritans is developing 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 three-year programme includes:

- these 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

- a hub of user resources helping people to stay safe online

SAMARITANS

Contact Samaritans Online Harms Advisory Service

Samaritans online harms team is available to provide support to all sites and platforms to manage self-harm and suicide content online safely.

Email
onlineharms@samaritans.org

Web
samaritans.org/industryguidelines

Phone
020 3621 5096 or 07483 027 824