

## **Context**

Understanding the processes by which an individual posts or engages with suicide and self-harm content, and that content is viewed, in public, online spaces requires many considerations. Firstly, the person posting may be struggling with distressing feelings or suicidal thoughts, and may struggle to express themselves in other ways. The Samaritans have a useful list of the reasons why someone may post such content:

*“Some of the reasons that users may post, search for, or engage with self-harm and suicide content include:*

- *Wanting to find out more about what they’re feeling and experiencing*
- *Finding a place to speak openly without fear of judgement*
- *Reading stories or connecting with others with similar experiences*
- *Seeking a supportive online community where they can get peer support*
- *Finding support for themselves or someone they know*
- *Finding help materials or support options*
- *Raising awareness of self-harm or suicide prevention*
- *Accessing immediate support– this is particularly important for users in crisis or waiting to access support from health professionals*

However:

*Users may also stumble upon content by accident or engage with it for more harmful reasons, such as to find information about methods of harm.”*

As this content may be visible to the wider public, and promoted or amplified by recommender systems, the need for moderation of the content is to prevent, potentially unintended impacts on the viewer. Elaborating upon what has been learnt in suicide prevention research, and embedded in their Media Guidelines, the Samaritans also note that content may:

- *Promote and glamourise suicide and self-harm*
- *Be unintentionally upsetting and triggering for other users*
- *Memes may present themes of suicide and self-harm humorously which can be upsetting*
- *Sharing methods of concealment – this includes advice on hiding historic injuries and scars which may be part of a user’s recovery. It also includes advice on hiding equipment or damage caused to the body relating to self-harm or suicide attempts. Helping users to hide harm from the people around them could delay them getting support.*
- *Memorial pages for people who have died by suicide – there are risks that the content could glamourise suicide or portray it as an effective way of ending distress.*

This does not take account of **the amount** of content that may be viewed, largely dictated by the recommender systems, which may be hard to avoid. This was powerfully described in the ySKILLS [“Young people experiencing internet-related mental health difficulties: the benefits and risks of digital skills”](#) Research Report:

*“Algorithms can act as a distorting mirror, magnifying problematic content and pushing young people with mental health vulnerabilities down a spiral of ever more overwhelming, upsetting or extreme content that they find hard to break away from.”*

Given that suicide, self-harm and eating disorder content will be engaging, especially to young people who are trying to understand and make sense of the complex issues in the world, there is a considerable risk that recommender systems promote and amplify extreme content that is ‘engaging’ (as evidenced in the Centre for Countering Digital Hate Report ‘[Deadly By Design](#)’) to those who do not want to see it, and are unprepared for what they see. Further, as stated above, some individuals may see greater amounts of that content, due to the operation of the recommender systems. They may then experience the further challenge of whom they can speak to about what they have seen, to process their experiences, and overcome what may have been traumatic. Of course, not everyone may see the same suicide, self-harm or eating disorder content, due to the way recommender systems operate, which may isolate further those who do see it, increasing feelings of guilt, and that they have brought this upon themselves.

The impact on a young person was vividly described in the same ySKILLS report:

*“You can read about symptoms, so that you can be aware of what kind of symptoms you should not show to your close family ... and to look at those pictures with self-harm content ... earlier I did not manage to harm myself, but by looking at the pictures, I have become used to blood and scars and such – so it has become easier to harm myself.”*

17-year-old girl with experience of self-harm and suicidal ideation

Overall, it is impossible to predict the impact of suicide/self-harm and eating disorder content on any individual, and anything explicit/graphic, and that is not clearly promoting recovery, should be moderated and taken down, **but with sympathetic support**. Industry needs to help those who have their content moderated understand why it was moderated, and what additional support they could access. Caution is required to avoid anything potentially punitive in the process of moderation, and Terms of Service/Community Guidelines need to be clear and helpful and not written in legal terms. Signposting is similarly a complex process, and should involve a few choices that someone in crisis could engage with, whilst they are also given some information about the service they are signposted to.

There are other aspects as to how moderation is perceived, such as bystander impacts in the wider community, if someone witnesses the incitement of a vulnerable person to self-harm, or if support from moderators isn’t visible.

### **Register of Risk chapter on suicide and self-harm content**

There are a few areas that may be worthy of further consideration:

1. It may be helpful to balance the emphasis on images with guidance on written posts related to suicide, self-harm or eating disorders, that may subtly normalise, idealise or glamourise unhelpful behaviours. Similarly, language that suggests painless death or

suicide is the answer to life's problems are all subtly promoting suicide. The Samaritan's media guidelines are very helpful on this, as are Beat's for Eating Disorders. Perhaps they could be in future examples

2. There is no reference to 'suicide hoaxes' in the guidance such as Momo or Blue Whale challenges. Given how contagious the sharing of information about these hoaxes can become, causing fear and panic in schools, families and young people, they warrant rapid moderation as well as media literacy campaigns on not sharing information on them, even if much sharing is altruistic, to warn other users.
3. Social reinforcement of self-harm behaviours may be quite powerful; this may be something that recommender systems need to address, as presumably they could promote some content?. The Brown et al research from Berlin, suggested:

*Results. Most commonly, pictures depicted wounds caused by cutting on arms or legs and were rated as mild or moderate injuries. Pictures with increasing wound grades and those depicting multiple methods of NSSI generated elevated amounts of comments. While most comments were neutral or empathic with some offering help, few comments were hostile. Pictures were mainly posted in the evening hours, with a small peak in the early morning. While there was a slight peak of pictures being posted on Sundays, postings were rather evenly spread across the week.*

*Conclusions. Pictures of NSSI are frequently posted on Instagram. Social reinforcement might play a role in the posting of more severe NSSI pictures. Social media platforms need to take appropriate measures for preventing online social contagion.*

4. Recommender systems do not just push content, they also shape the behaviours of influencers, and because mental health and self-harm content is engaging, recommender systems can nudge influencers to become more extreme. This account of [the lived-experience of an influencer](#) evidences how maintaining a successful channel can shape behaviour and content.
5. The Business Model...as this area of content is engaging, and promoted and amplified by recommender systems, the business model is fundamentally supportive of this content being amplified.
6. Grey area content – this is an area where the corruption of beliefs and values can be subtle and hard to detect, and those working in moderation need space, time and training, to promote reflective practice, to not only protect them from the potentially traumatic impact of some content, or if a suicide occurs, but to help them consider what, on balance might be the best next steps in terms of moderation. Striking the balance between individual needs, different vulnerable groups and the wider community is challenging, and a degree of compassion for those that moderate is required; KPIs emphasising the rapid take down of content, that do not allow for deeper considerations of complex content should be questioned.

**Illegal Content Judgements Guidance (ICJG) chapter on illegal suicide and self-harm content (Annex 10)**

Assessing intent, when it relates to 'Assistance', is likely to be a complex process, but it may be possible to derive some insights from those assessing intent when bullying is occurring? For example, services may usefully consider how frequent or repetitive the assistance is? Whether it is occurring as a group activity or an individual activity? The level of detail shared, in terms of describing methods of suicide etc may also be a helpful pointer to the degree of intent. Cognitive development and developmental stage are also important when assessing intent, as well as health knowledge, as some young people think it is okay to give information on methods because they are supporting another person with what they are requesting, and may not anticipate the consequences.

Overall, the summary provided in Volume 5 is very helpful:

*We are consulting on our view that where specific, practical or instructive information on how to end one's life is posted to a forum or within a chat in which suicidal ideation is discussed, it may be reasonable to infer that intent to assist (attempted) suicide exists by virtue of information having been posted. Where an encouragement to end one's life is posted in response to what appears to be a credible threat by another user that are about to take their own life, it may also be reasonable to infer intent.*

Dr Richard Graham FRCPsych  
Consultant Child & Adolescent Psychiatrist  
Clinical Director, stem4