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

Following the tragic loss of 14-year-old Molly Russell in November 2017, the Russell family and their friends set-up a charitable foundation in her memory. The Molly Rose Foundation [MRF] is a charity with the aim of suicide prevention for young people under the age of 25. MRF aims to raise awareness about suicide and its causes and to connect young people at risk of suicide to the help, support and practical advice they need.

Three of MRF's priorities are:

1) Education to promote better understanding.
By working primarily in schools, we provide training in Mental Health First Aid and encourage wider conversations about wellbeing and good mental health in the school community.

2) Guidance to promote better support.
We aim to connect young people who are struggling with suicidal thoughts to the support they need and to guide them away from the unhelpful and harmful content that is available online.

3) Policy to promote safeguarding.

We work to improve the online safety of young people by: i) Raising the general public's awareness of online harms and their impact on health, ii) Engaging with governments, regulators and other charities to help develop legislation, regulation and policies to reduce the effect of online harms, iii) Encouraging governments, companies and individuals to prioritise the online safety of young people who use digital technology.

MRF's responses to this call for evidence draw from what we have learnt from the inquest into Molly Russell's death; the many connections we have made with other charities and organisations who work in in this sector; and other families sadly bereaved because of the harms that can be found 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)

No

There should be a general presumption that services will be accessed by children unless they can demonstrate otherwise. Any service that states a minimum age for its use as 17, or under, by definition, will be likely to attract child users.

Although it is a common practice for services to state a lower age limit of 13, some services are expressly designed to be used by younger children, for example, YouTube Kids, which offers, “Nursery School” Mode which is designed for children under four. https://www.youtube.com/intl/ALL_uk/kids/parent-resources/

Other services have indicated they are keen to engage with a younger user-base by designing a version of their service for children, for example, Instagram Kids aimed at 10-12-year-olds. Development of this service has been ‘paused’ by the platform to allow more time for Instagram to “work with parents, experts and policymakers” after public concern was raised about its safety.
https://about.instagram.com/blog/announcements/pausing-instagram-kids

User-to-user, social media services with a large global reach, are naturally more likely to attract a greater number of child users. The pattern of this use changes as new services are launched, with young people reporting, over time, popularity has tended to shift from Facebook, to Instagram, to Snapchat, to TikTok, to BeReal, for example. In our experience, child users more readily adopt new services and they are more likely to be influenced to join services used by their peers, than adult users.

When assessing safety, services should not be considered in isolation as one service is often used by children, even if inadvertently, to cross promote another service. For example, children frequently use WhatsApp to share TikTok posts. The combined impact of all services should be considered when assessing or monitoring the spread of any online harm. Consideration should also be given to the likelihood that this cross promotion may also encourage children to set-up accounts on additional services. Peer pressure among children should never be underestimated.

In the event of an online harm being cross promoted by more than one service, as described above, the regulator will need robust policies to determine which service is the ‘risk owner’. In the above example, would the ‘risk owner’ be the service used to share the content, WhatsApp; or the service hosting the content, TikTok; or a combination of the two?

We have found that widely used services and their communities may help connect a child to other, less well-known services. Evidence submitted to the coroners presiding over the inquests into the deaths of Molly Russell and Frankie Thomas respectively revealed

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

smaller-scale services such as TalkLife and Wattpad have been used by children who later ended their own life. Although it is often difficult to obtain data to assess the impact any service may have on a child, all services likely to be accesses by children should prioritise their safety and make data available to the regulator or a coroner (and ensure its preservation) in a timely and accessible manner whenever officially requested.

The methods utilised by services to promote the engagement of their users are highly developed and greatly affect how likely a service is to attract children. Engagement Based Ranking [EBR] algorithms, directly affect what is seen in a user's 'feed' for each service and therefore in turn EBR affects the user's experience of that service.

In Molly Russell's case, services used algorithmically produced, in-app prompts to encourage her engagement. One platform, Pinterest, also sent her emails and push notifications encouraging her to view further harmful content. For example, one email sent to Molly after her death was titled, "18 Sad depression quotes Pins you might like" another, "Stay strong, Depression problems, and more Pins trending on Pinterest" – both emails also contained images connected with self-harm (some graphic); depression; and suicide (including method).

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

It is evident that, especially since the introduction of engagement-based algorithms, platforms gather a rich seam of data about all their users, including children, to promote their engagement with the service. This serves their business models.

Historically, the information obtained by services about the age of their users has varied and will probably continue to do so as technology and industry practices evolve.

Molly Russell's inquest showed that Instagram and Pinterest did not know Molly's age (other than she claimed to be over 13 when she set-up her accounts), as they did not ask their users for a date-of-birth until 2018. Meta's witness, head of Health & Well Being, was unable to tell the Court how many children were on Instagram.

Services commonly adopt a lower age limit of 13, predominantly because of US federal law. The Children's Online Privacy and Protection Act 1998, which effectively means anyone younger than 13 cannot open an account without verifiable parental consent. This lower age-limit is in place, not as a safety measure but simply in order to comply with this

Question 3: What information do services have about the age of users on different platforms (including children)?

legislation. This was confirmed by the Meta witness and we believe many parents are not aware that this ‘age limit’ is not a safety measure.

To prevent harm to children, services must more quickly adopt any new proven age protections as soon as they become available.

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

For the Online Safety Bill, as currently drafted, to meet the requirements of the ‘Triple Shield’, effective use of up-to-date Age Assurance & Age Verification technologies is vital in terms of online safety for children.

However, it is unlikely that children will ever be wholly prevented from gaining access to a service or part of a service. The effectiveness of such measures will remain in a state of flux; a game of ‘cat and mouse’ between the services’ Age Assurance and Age Verification methods and the ‘hacks’ young users collectively develop to circumvent the protections.

It should be remembered that just as children naturally encourage each other to join different services, they are particularly adept at sharing methods that allow them to circumvent the various protections put in place to protect them.

Many children gain access to services by using their parent’s log-in, either with or without parental permission. As well as providing a service access to the child’s data, this common practice may put children at risk. For example, Duncan McCann’s (5Rights) complaint alleging the Age Appropriate Design Code has been broken by YouTube who he accuses of harvesting Children’s data:
https://www.theguardian.com/technology/2023/mar/01/father-reports-youtube-to-watchdog-over-harvesting-uk-childrens-data

In all cases, it is essential for the service to regularly review the effectiveness of any Age Assurance or Age Verification measures they employ and to openly publish their findings.

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

MRF is aware that historically, popular user-to-user and search services are slow to adopt current Age Assurance and Age Verification practices, leaving children at risk (see response to Question 3 above).

Established Age Assurance techniques include: Age Gates, Self-Declaration, Third-Party Verification Services, Parental Consent, and Identity Verification. Generally, the more effective the method, the more impact it will have on the user's experience and the greater the likely cost.

As safety concerns about children's access to age-appropriate content have grown, many services have gradually introduced improved measures to better assure the age of a user. Currently, this is often simple user-declared Age Assurance, undertaken when an account is set-up. Sometimes additional precautionary measures are employed as a deterrent, such as delaying date-of-birth re-entry after an unsuccessful under-age submission. We are not aware of any published evidence supporting the effectiveness of such additional measures.

Linked sign-ups, using a user's previously established social media accounts (such as Facebook or Google), can assist in determining age, always assuming the originating service has itself correctly verified a user's age. However, this process does have data privacy and security implications which are often overlooked by a young user, who is likely to favour the convenience of the linked sign-up process. For these reasons, we think there are better solutions.

Although services seem to have been slow to adopt state-of-the-art Age Verification techniques, they are now beginning to be introduced, for example Instagram's recent announcement:
https://about.instagram.com/blog/announcements/new-ways-to-verify-age-on-instagram

We have reservations about the practice of some of these measures. Social Vouching, asking mutual followers over 18 years of age to confirm the new user's age, may prove ineffective, or even unsafe. It encourages young people to connect with older mutual followers and so may be detrimental to a child's online safety, by providing a route to under-age verification rather than preventing it, and increasing the likelihood of a child being introduced to unknown adults.

Whereas partnerships between services and established third-party Age Verification companies (such as Yoti) should be encouraged and best practice for the use of such Age Verification, should be shared across the industry.

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?

It should be noted that in May 2022, Yubo adopted Age Verification for all its users in a partnership with Yoti, a partnership that began in 2019. Although a comparatively small service, this nonetheless demonstrates Age Verification can be achieved at scale and demonstrates that other large global platforms have been slower to adopt up-to-date Age Verification technologies.
https://www.yubo.live/blog/yubos-new-age-verification-feature-helps-keep-you-safe

This should be compared to the recent data filed with Ofcom which shows Snap removed just 700 suspected underage accounts from Snapchat in the UK between April 2021 and April 2022, representing just 0.4% of some 180,000 accounts removed by TikTok over the same period. The setting of industry standards for the prevention and removal of under-age accounts should be considered by the regulator.

Whatever Age Assurance and Age Verification techniques are employed by a service, full risk assessments of such measures should be regularly undertaken by the services, and these should be available to both the regulator and the public.

There is always more that can be done. Some cybersecurity experts suggest a secure, anonymised, 'double-blind', third-party, key-encrypted register of online users ages would be the gold standard for Age Verification.

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

MRF was founded after the death of 14-year-old Molly Russell in November 2017. The inquest into her death took place in September 2022. The nearly five-year process revealed evidence of how user-to-user and search services were, as the coroner put it, 'in a more than minimal way' connected to Molly's death.

During the coroner's investigation, evidence was obtained showing the harmful content that Molly had engaged with online. We are limited to sharing details only of content referred to in Court during the inquest (much to our frustration).

At the direction of the coroner, in 2019, the police were able to recover data from Molly's electronic devices and thereafter exercised his powers to seek records from tech companies about Molly's experience on them in the last sixth months of her life. This revealed the extent of harmful depressive, self-harm and suicide related content Molly

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?

had been exposed to when online. Some of this content was immediately shocking being very graphic in nature. Other, often algorithmically recommended, posts had a longer-term cumulative effect on Molly (and others who viewed them) and are considered by many to be the most harmful.

In December 2019, the coroner issued five Schedule V notices to WhatsApp, Twitter, Snapchat, Pinterest and Meta, to request further data in connection with Molly's accounts. The responses of the platforms varied:

WhatsApp: Provided a signed witness statement saying Molly's account had been deleted on 23rd March 2018, most probably after 6 months inactivity. When this happened Molly's group contacts received a message stating, "Molly Russell left." Some WhatsApp data was available from Molly's devices and her contacts' accounts – none through WhatsApp/Meta.

SnapChat: Provided minimal information about Molly's account. Nothing about her activity, but we know from her phone log she last accessed the app at 00.42 on the morning of her death. The content of that chat remains unknown, Snap stated that a US court order would be required for them to engage further.

Twitter: Allowed Molly's account to be relinked to her father's email which enabled a password reset and a self-access download from the platform. This didn't show us what Molly had seen on Twitter – such as the tweets promoted to her, or accounts recommended.

Pinterest: Employed a team of 17 staff for weeks to provide over 9,354 pages of documents in an accessible format including pins Molly liked, saved, close-upped on or scrolled over and paused (pin impressions).

Meta: After delay, initially provided material by way of an inaccessible and vast spreadsheet that was heavily redacted and contained 6 months of Molly's Likes, Saves and Shares only. At the time it was first provided only material from public accounts was accessible to the Russell family and their lawyers, however the Court later ordered Meta to provide more. Meta was either unable or unwilling to share what material they promoted to Molly, the time she spent on the platform, or what searches she made. Some of the content from private accounts was provided in early 2022, but it was less than a month before the inquest began that further content was provided, each round of disclosure revealing more and more harmful content, and binges of, for example, horrific videos (none of which Meta had deemed necessary to remove from Instagram in 2020 when they provided the data).

Among data not disclosed by Meta were: two accounts that blocked Molly; 23 accounts that Molly blocked; 846 accounts that Molly followed; and 272 accounts following Molly.

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?

The Russell family's legal team estimate that the evidence submitted by Meta amounted to only 5-10% of the material Molly saw on Instagram in the last six months of her life.

In this six-month period, there were approximately 16,300 images Molly engaged with on Instagram submitted as evidence – and about 2,100 of them are depression/self-harm/suicide-related. There are only 12 days when Molly doesn't engage with some kind of suicide/self-harm content on Instagram. Suicidal material appears to have been viewed by Molly on 84 days out of 183; and self-harm material on 51 days out of 183.

In total, the thousands of pages of evidence submitted to Molly Russell's inquest, demonstrates the range of harmful content that was widely accessible prior to Molly's death in 2017. At the inquest it was also shown that, in this respect, too little had changed by 2022, the same content and similar harmful content to that which Molly had seen was found still to be easily available, nearly five years after Molly's death. Indeed, similar videos were shown to the House of Lords and others in January 2023 by the Russell family's lawyers.

For Molly's inquest both Meta and Pinterest became 'Interested Person's' and sent witnesses to provide further oral evidence under oath in open Court, during the September 2022 inquest. Both these services admitted Molly had seen content on their platforms that had contravened their guidelines that were in place in 2017, when Molly died.

The evidence provided to Molly Russell's inquest showed that some of the harmful content, especially video content, utilised sophisticated production techniques. Sophisticated in terms of generating engagement with young service users, but also sophisticated in terms of the techniques employed to overcome service moderation processes. The self-harm and suicide related content showed a similarity of production to other harmful content, particularly content connected to Child Sexual Abuse.

For example, harmful written content was often distributed across the video timeline so that any single freeze-frame or screenshot would not convey the totality of the harmful written message contained in the video. Or the inclusion of a helpline phone number in the video, which even led Meta's witness at Molly Russell's inquest to suggest one particularly graphic video was 'safe' to view solely as, among the harmful content it contained, a source of support was provided to the user.

Ofcom should investigate genre-crossing patterns of harmful content and more independent research is urgently needed to uncover widespread pattens of abuse. For example, we have heard there is a growing base of evidence linking extreme misogynistic content to pro-suicide forums. So, for the safety of children online, it is vital to better understand any connection there may be between Incel forums and content that encourages young females to self-harm or consider suicide.

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

A sample of the harmful content obtained as evidence for the inquest into the death of Molly Russell was shown in open Court at her inquest. This included photographic and text-based posts and videos.

Sometimes a single post or video proved immediately shocking or disturbing to the viewer. When the combined effect of the algorithmically amplified stream of harmful content Molly had engaged with in the last six months of her life was taken into account, the overall scale of its harmful effect was generally considered to have the potential to cause great harm, especially to children.

In his Prevention of Future Deaths report the coroner said, "It is likely that the above material viewed by Molly, already suffering with a depressive illness and vulnerable due to her age, affected her mental health in a negative way and contributed to her death in a more than minimal way."
https://www.judiciary.uk/prevention-of-future-death-reports/molly-russell-prevention-of-future-deaths-report/

In Court, prior to the small selection of the many harmful videos seen by Molly were shown, the coroner said, "What was suggested at some point, that the footage being so uncomfortable to view, simply in some way to edit it, but Molly had no such choice. So, we would in effect be editing the footage for adult viewers when it was available in an unedited way at the time. So, my view, Mr Sanders [the Russell family's KC], is that the video footage should be played as a standalone piece of evidence but I must say is that I say this with the greatest of warnings, the footage appears to glamorise harm caused to young people. It is of the most distressing nature, and it is almost impossible to watch. So, I say to anyone in Court if you are likely to be affected by such images, please do not stay. Particularly, I say this to members of Molly's family. There is no need for any of you to stay to see anything in this Court that you might find distressing. You do not need to ask to step outside. In my view, this sequence of video footage ought to be seen."

At Molly Russell's inquest the witness from Pinterest expressed regret for the platform sending harmful content to her. When asked if he would show the content seen by Molly to his children, he answered, "no."

When asked, when Molly was on the platform, whether the content she was looking at wasn't safe, the witness from Meta replied, "Molly viewed some content that violated our policies, and we regret that."

When asked if it is safe for mentally unwell children to be in this online environment, Meta's witness responded, "Instagram is a safe place and we take many measures and

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

efforts to safeguard our users. I cannot speak to every factor affecting somebody with a clinical situation."

In evidence provided by an expert Child and Adolescent Psychiatrist to Molly Russell's inquest, the Court heard that, 'Depressive disorder in adolescence is common worldwide but often unrecognised. The incidents notably in girls rises sharply after puberty and by the end of adolescence the one-year prevalent rate for adolescent girls exceeds 4%.' It follows that on services with millions of young users, there will be a large number (one in 25) of children with a depressive disorder (either diagnosed or undiagnosed) who may be more vulnerable to any harmful content they experience.

This calls into question the judgement of platform representatives, when deciding what content may or may not be harmful.

The evidence submitted to Molly Russell's inquest had a profound impact on the adults who saw it. The Russell family's solicitor has also spoken publicly of requiring professional assistance to assist her and her team with the negative impact of reviewing the content. Merry Varney described herself as a resilient adult yet still felt the harmful effects of this content. Dr Navin Chandra Kunigal Venugopal, the child psychiatrist who gave expert evidence under oath said, he'd slept poorly for weeks, having seen some of the content seen by Molly. Other adults who have seen a sample of the harmful content evidence since the inquest (including MPs, Peers and professionals working on the Online Safety Bill) have been moved to tears, or left the room because of its harmful nature.

Research carried out for Samaritans by Swansea University, published in November 2022, shows over three-quarters of young people have experienced harm online by the age of 14, furthermore 83% of survey respondents reported that they had seen self-harm and suicide content social media even though they had not searched for it.
https://media.samaritans.org/documents/Samaritans_How_social_media_users_experience_self-harm_and_suicide_content_WEB_v3.pdf

Exposing the majority of UK children to profoundly disturbing harmful content at what is acknowledged to be vulnerable stage of their development is likely to produce negative effects and be detrimental to their health. Further independent research is needed and access to anonymised data from the platforms would greatly aid our understanding of the effects of digital technology on the health of children who use it.

This is underlined by recent research (published in the Journal of Child Psychology and Psychiatry on 21st March 2023 by Karima Susi, Francesca Glover-Ford, Anne Stewart, Rebecca Knowles Bevis, and Keith Hawton) which concluded, "Viewing self-harm images online may have both harmful and protective effects, but harmful effects predominated in the studies. Clinically, it is important to assess individual's access to images relating to self-harm and suicide, and the associated impacts, alongside pre-existing vulnerabilities and contextual factors. Higher quality longitudinal research with less reliance on

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

retrospective self-report is needed, as well as studies that test potential mechanisms. We have developed a conceptual model of the impact of viewing self-harm images online to inform future research."
https://acamh.onlinelibrary.wiley.com/doi/full/10.1111/jcpp.13754

At the end of the inquest into the death of Molly Russell, HM Senior Coroner Andrew Walker concluded, "Molly Rose Russell died from an act of self-harm whilst suffering from depression and the negative effects of on-line content."

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

Having considered both the coroner's conclusion and the evidence provided for Molly Russell's inquest, MRF believe services do not do enough (if indeed anything) to assess the risk of harm to children in the UK from content that is harmful to them.

The Court heard that, in his 7th February 2019 Press Release entitled, 'Finding the Right Balance,' Adam Mosseri, Head of Instagram said, "We have allowed content that shows contemplation or admission of self-harm because experts have told us it can help people get the support they need. But we need to do more to consider the effect of these images on other people who might see them." Yet still, we are not aware of the results of Instagram's considerations into finding the right balance nor any significant change of policy resulting from this initiative.

The platforms represented at the inquest were unable or unwilling to provide to the Court, any expert advice and recommendations they received relating to the risks to children using their platforms in 2017, or that led to the development of their guidelines. At best, only a limited number of short summaries of their third-party advice was submitted.

When asked why third-party evidence was not available, Meta's witness said, "There are some written records and notes, of course, of different meetings, certainly. But some of the conversations we might have with a variety of experts might be confidential; or might happen in a more informal context. We consider different factors when we make the decisions."

Meta's witness was also asked if she had any statistics or data about how many times an act of self-harm or suicide may have been prevented because of a post on Instagram, she

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

answered, “I do not know the answer to that question. I don’t know if that sort of data exists.”

To view more than this incomplete picture, it will be essential for the regulator to gain full access to any third-party advice given to services. Ofcom will then be able to assess the integrity of platform guidelines.

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

Platforms often resort to increasing parental controls when there are calls to increase safety of children online. For example, following the publication of the CCDH report:
https://counterhate.com/research/deadly-by-design/
TikTok introduced new ‘Family Pairing’ safety features for families and children including teen screen controls and mute notifications for caregivers:
https://newsroom.tiktok.com/en-us/new-features-for-teens-and-families-on-tiktok-us

Increased parental supervision is, of course, welcome, but it may have the unintended consequence of increasing barriers between children and their caregivers. An often cited, an important aspect of maintaining the safety of children online, is to encourage greater dialogue between generations, to make it more likely young people will seek support and advice from their parents and carers when they encounter harmful content.

The regulator should take steps to ensure they are able to gain evidence obtained from a service, whenever requested, in a timely manner. For example, agreed procedures for data provision from services should be established to avoid delay arising from discussions regarding privacy rights or restrictions to disclosure.

At the request of the regulator, the preservation of a child’s data held by a service should also become an accepted norm, especially as accessing data from the locked device of a child often proves to be time consuming. Again, this should become an established process, as should the privacy protection practices employed when supplying data. For example, services may be required to redact usernames or other identifying information.

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

Experience (and evidence from Molly Russell's inquest) has shown that governance, accountability and decision-making structures for child user and platform safety tend to be complex and opaque, even when evidence is directly requested. For example, debate and delays about how much of Molly's data could be supplied by Meta to her inquest resulted in around a two-year delay to the proceedings.

For the proposed regulation to be effective it will be imperative for Ofcom to insist on greater transparency and for platforms to comply with all official requests fully and in a timely, co-operative manner.

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

Complex guidelines and conditions can be particularly difficult for children to comprehend. This has led to, for example, the current draft of the OSB to be simplified to a three-page document by SWGfL (part of UKSIC) and Schillings LLP: https://swgfl.org.uk/magazine/swgfl-partners-with-schillings-on-new-online-safety-bill-guide/

The UKSIC also provides a good example of the use of accessible communication techniques to aid understanding, for example, this video explaining their 'Report Harmful Content' service: https://saferinternet.org.uk/report-harmful-content

Online services should develop their own accessible content and tools designed to allow young people to exercise their rights, and to enhance the clarity and accessibility of their terms and service and public policy statements. Sharing best practice in communicating complex ideas to young people should be encouraged by the regulator.

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)

[Please select]

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

(See 14 below)

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

The UKSIC ‘Report Harmful Content’ service demonstrates how an easy-to-use reporting service can be effectively implemented.
https://saferinternet.org.uk/report-harmful-content

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

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

When some of the harmful online content that Molly engaged with was reported to platforms, after her death, in 2017/8, they responded by claiming it did not breech their Community Guidelines and so would not be taken down. This continued to leave many vulnerable young people exposed to the same harmful content and provided an opportunity for others to encounter it.

It is imperative content is removed swiftly when found to be harmful, to protect young and vulnerable people online.

Two platforms (Meta and Pinterest) became ‘Interested Persons’ in the inquest into Molly Russell’s death. In some respects, their approaches to co-operating with the coroner’s investigation revealed marked differences. This ‘Tale of Two Platforms’ is reflected in or response to question 6 above.

Whenever official requests are made to services by coroners, law-enforcement officials or the regulator, the services should be encouraged to respond more like Pinterest, and less like Meta, in the case of Molly Russell. Establishing what is considered best practice when official requests are made would save time and could save lives.

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

In 2019 Instagram introduced ‘Sensitivity Screens’ to blur out content that might be immediately harmful when viewed. These do not prevent the user from accessing the content, but they now must click to show the post.
https://www.theguardian.com/technology/2019/feb/04/instagram-to-launch-sensitivity-screens-after-molly-russell-death

Although this technique has obvious advantages, we are not aware of research into its effectiveness. Whereas children have reported, they are rarely deterred from viewing the

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.

content, with some saying it can increase their curiosity to view the post. It should be expected that independent research is commissioned, and its results published openly, so the effectiveness of such safety measures can be properly assessed.

There are services, developed by third parties, that can be used to improve the safety of children when online. Services should be encouraged to adopt effective safety measures and to aid their development.

One example of such a safety product, developed in the UK, is R;pple suicide prevention. https://www.ripplesuicideprevention.com

This is an interceptive online tool which presents a visual prompt to sources of support when a person's search includes harmful keywords or phrases relating to self-harm or suicide. While developing the plug-in, extraordinary determination was required from R;pple's CEO and founder, Alice Hendy, in order to overcome resistance from tech platforms and their expert advisors. Instead, the development of new safety features should be encouraged. Alice, a person bereaved by the suicide of her brother, with comparatively limited resources, provides an example to big tech platforms showing how improved online safety can be achieved.

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

The concept of 'Safety by Design' should be embraced by all services likely to be accessed by young people. Functionalities to measure risk should be developed alongside the platform function. No service should launch a new product likely to be used by children, without first providing a full Risk Assessment of its use in practice that demonstrates it will be safe for children to use.

As emerging technologies are developed, not only should safety be considered an intrinsic part of their design process, but the safety of older technologies should also not become overlooked. Proportionate safety measures for all levels of tech, likely to be accessed by children, must always be maintained.

Safety features are often adopted by services, reactively rather than pro-actively. For example, two years after Molly Russell's death, and within a fortnight of her story breaking, on the 4th February 2019, Adam Mosseri, Head of Instagram wrote, "We rely

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?

heavily on our community to report this content, and remove it as soon as it is found. The bottom line is we do not yet find enough of these images before they're seen by other people."
https://www.telegraph.co.uk/news/2019/02/04/changing-instagram-support-people-tormented-suicidal-thoughts/

At Molly Russell's inquest in September 2022, the Court heard that Instagram moved from a chronological feed to one based on an engagement based ranking algorithm in 2016. Yet there was no evidence provided of a connected change to the Terms of Use; no connected change to the Community Guidelines; and no connected change to the Privacy Policy the service had in place.

Such delays between the introduction of new operating processes and the introduction of new safety measures may result in a period when a service is significantly less safe for young users and therefore these delays should be avoided. Every new development should be properly Risk Assessed, and effective safety measures should be in place, prior to its introduction.

At Molly's inquest, when asked if the Instagram focus, is on the intentions and the benefit to the poster, the Meta witness replied, "Not now, no. We consider the very delicate nuance of those viewing the content; and those posting the content." When questioned about the balance between the poster and the viewer in 2016/7 she responded, "That was more where we, our policy line was, based on expert guidance to really consider the potential harm and safety of those posting the content."

It is important for services to have established policies, that are accessible to both the public and the regulator, that make clear how the platform balances the needs of the admissive poster against the potential risk to the number of viewers the post is likely to reach. This is particularly important for any service that employs a content ranking feed as the danger of harmful content will be algorithmically amplified by such systems.

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

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

When developing products, services should adopt a ‘Safety by Design’ approach, i.e. consider safety as an intrinsic part of their design process; from product conception, through its launch, and beyond when providing ongoing support and development.

Service providers should publish Risk Assessments of their service, clearly identifying the steps they have taken to ensure their product will be safe for children to use. They should also commission independent research to monitor the way their platform is actually used in practice and proactively introduce new safety features without delay whenever new, emerging harms are discovered.

Safety procedures should be transparent to both the regulator and the public and any emerging harms and services should share the best way to combat the new harms without undue delay.

Better signposting services should be employed so when a user is identified as being at risk, they are connected to the help and support they need. An example of this is the ‘Find a Helpline’ service, which can be found on the MRF website: https://mollyrosefoundation.org/fah/

Search engines should develop and utilise existing industry best practices to signpost to support. Initiatives such as the ‘R;pple’ browser extension (see our response to Question 16 above) provides an example of an online safety measure that should receive greater industry support.

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

For the regulator to assess the effectiveness of algorithms designed to support the function of a service, it is essential algorithmic transparency is provided by the platforms.

Nearly all services utilise many interconnected algorithms to create the way their platform works. Gaining a clear understanding of how these algorithms collectively affect the safety of young people who use the service will be a constant challenge for the regulator.

In fact, during her oral evidence given under oath during the inquest into Molly Russell’s death, the Meta witness said, “We don’t know what content was recommended to Molly; and we also don’t know what content may have surface resources.” It is concerning,

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)?

especially given the circumstances, that Meta's Head of Health and Wellbeing Policy, was unaware the type of content recommended algorithmically and whether or not any support was provided as a result.

Emerging technologies will mean this is a constantly and fast-changing aspect of regulation. For example, we understand, LinkedIn are already using Generative AI to produce posts which promote user engagement of their service. The interplay between these automated posts and the service's existing algorithmic amplification may quickly create radically different forms of content that is digitally recommended to new users in unprecedented ways across multiple services.

The evidence obtained by the coroner for his investigation into the death of Molly Russell is widely mentioned elsewhere in this response. In terms of algorithmic risk, the findings can perhaps simply be summarised by saying, in Molly's case, far more harmful than any single piece of content, was the algorithmic amplification of thousands of pieces of content relentlessly sent to Molly.

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

Services should not deliver innovative features for their users without fully considering their safety first, especially for services likely to be accessed by children.

Platforms often point to advances in emerging technologies, such as Artificial Intelligence, as being required to improve content moderation. It is important to be aware of the limitations of such automated processes. For example, Bill Ready, CEO of Pinterest has recently said, "as much as there's great advancement in AI, can AI catch all of it? The answer to that is no, at least not today."

It is therefore imperative that services devote sufficient funds to the development of AI (or other technological) based safety features. The regulator should be able to assess the percentage of turnover/profit a service devotes to maintaining and developing suitable technologies to support its content moderation.

In the meantime, it remains imperative for services to invest sufficiently in human moderation resources, if they are to be effective at removing harmful content. Similarly, to above, a service's spend on human moderation should reported to the regulator.

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

(See above question 20)

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

Despite the advances in AI technology, human moderation will always remain an essential part of any service's obligation to remove harmful content. This especially applies when dealing with any appeals.

The rules by which human moderators make their decisions, and the effectiveness of those decisions, need to be monitored by the service and shared with the regulator if Ofcom is to ensure a service is investing sufficiently in moderation to keep children safe when online.

Most importantly, it is vital that human moderators are trained to err on the side of caution when making decisions about protecting children from harmful online content. The regulator should ensure the service has clear human moderation policies to protect children's safety rather than preserve harmful content.

All services should clearly state their policies in connection to human moderation in their guidelines and be able to demonstrate these practices are strictly adhered to.

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

Is this a confidential response? (select as appropriate)

No

We understand that in some cases, human moderators are only given 17”, for each piece of reported content, to make their decision about its safety.

Services should provide the regulator with details about the working practices of their human moderators, so an assessment can be made as to the viability of the service’s human moderation process.

In a similar way people who are required to look at disturbing content for legal or other established reasons, moderators will need regular support and phycological evaluation.

Ofcom should have a supervisory role to ensure services take sufficient measures to safeguard their human moderators.

Above all, it should never be forgotten, the harmful online content easily available to children detrimentally affects everyone who sees it. The Russell family’s legal team, when working on Molly’s inquest, sought professional help to safely review the evidence. The child psychiatrist who supplied expert evidence under oath said, he’d slept poorly for weeks, having seen the evidence. And police officers admitted they’d been moved to tears, having seen the content – content we know had been seen by a fourteen-year-old, and without doubt, by countless other children.

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)

[Please select]

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

It is evident that content that is harmful to children and in contravention of terms and conditions is too infrequently removed from a service.

Evidence obtained for the inquest into the death of Molly Russell showed in court that:

i) After Molly's death, harmful content encouraging self-harm or suicide, reported to services in 2017/8, was not removed as it was judged by the platform moderation process not to have broken the service's Community Guidelines. We understand this seems to be an all too frequent occurrence in other cases.

ii) Content Molly engaged with, that contravened service guidelines, was found to be harmful to her, affecting her mental health and contributing to her death in a more than minimal way.

iii) In some cases, the harmful content Molly engaged with in 2017 was still readily available in August of 2022, despite the service's terms and conditions becoming more stringent in that time.

In fact, rather than removing contravening harmful content, the engagement-based ranking algorithms employed by services ensure that it is likely most children will have seen harmful content by the time they are 14, and often much younger (see Samaritans report mentioned in our response to Question 7 above).

The three-year UK-wide study, by C Rodway *et al,* of all young people aged 10-19 who died by suicide, based on national mortality data, states suicide-related online experience was reported in 24% (n=128/544) of suicide deaths in young people between 2014 and 2016. This is equivalent to 43 deaths per year and was more common in girls than boys and those identifying as LGBT. The study also concludes this is likely to be an underestimation and for public health, wider action is required on internet regulation and support for children and their families.
https://pubmed.ncbi.nlm.nih.gov/35587034/

Question 26: What other mitigations do services currently have to protect children from harmful content?

Is this a confidential response? (select as appropriate)

No

Question 26: What other mitigations do services currently have to protect children from harmful content?
Services should provide assets for improved digital literacy education so that children (and their care providers) are better informed about the dangers to be found online and better equipped to deal with them appropriately.

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) *[Please select]*

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
A transcript of the inquest touching the death of Molly Russell has been approved by HM Senior Coroner Andrew Walker. MRF recommends that Ofcom keep a copy of this transcript on file as we believe it will be a useful reference both during the 'Roadmap to Regulation' and when Ofcom takes on its new role as safety regulator. We will provide this transcript as an Annex to this response. If MRF can provide further information or clarity, or assist in any other way to support Ofcom with their preparations to become the UK online regulator, we would be only too pleased to do so.

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?

--